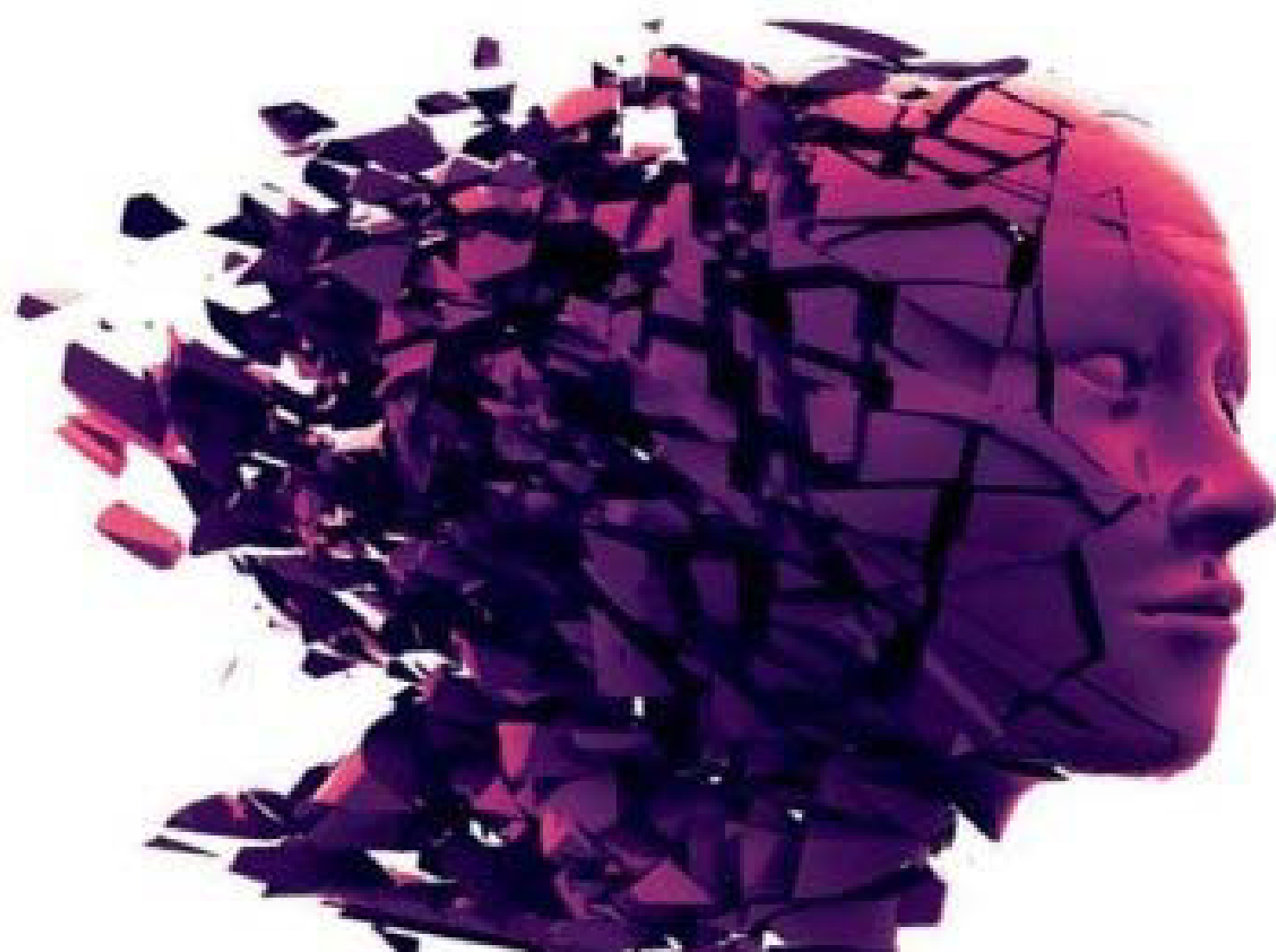


OXFORD

Experimental Design **and** Analysis **for** Psychology

Hervé Abdi, Betty Edelman, Dominique Valentin, & W. Jay Dowling

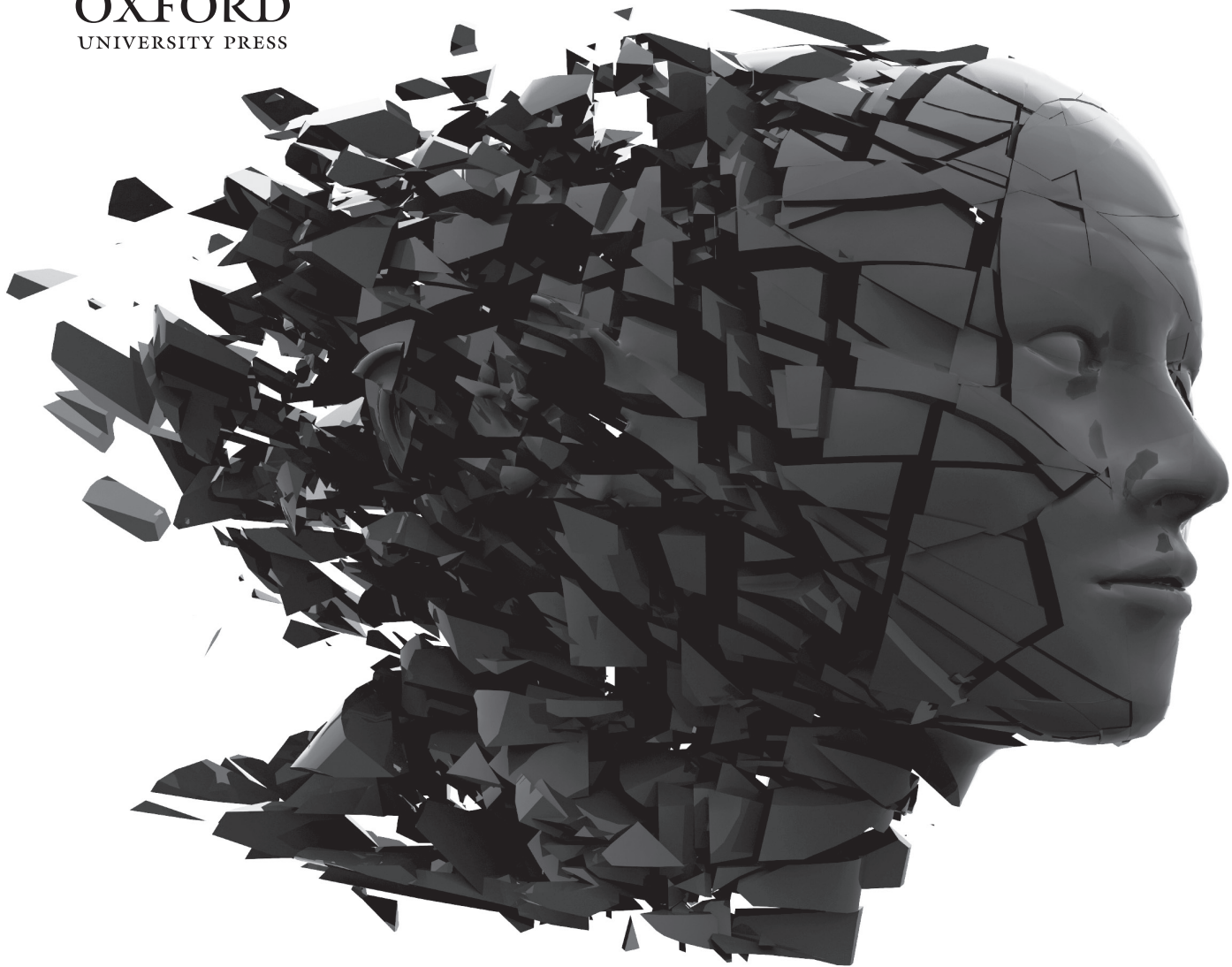


Experimental Design and Analysis for Psychology

Experimental Design and Analysis for Psychology

Hervé Abdi, Betty Edelman, Dominique Valentin, & W. Jay Dowling

OXFORD
UNIVERSITY PRESS



OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi

Kuala Lumpur Madrid Melbourne Mexico City Nairobi

New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece

Guatemala Hungary Italy Japan Poland Portugal Singapore

South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press, Inc., New York

© Hervé Abdi 2009

The moral rights of the authors have been asserted
Database right Oxford University Press (maker)

First published 2009

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
without the prior permission in writing of Oxford University Press,
or as expressly permitted by law, or under terms agreed with the appropriate
reprographics rights organization. Enquiries concerning reproduction
outside the scope of the above should be sent to the Rights Department.
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover
and you must impose the same condition on any acquirer

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Data available

Typeset by CEPHA Imaging Pvt. Ltd., Bengaluru, India
Printed in Great Britain
by Ashford Colour Press Ltd.

ISBN 978-0-19-929988-1

1 3 5 7 9 10 8 6 4 2

Preface

The impetus for this work, like so many of its genre, came from our students and from the lack of an appropriate text for them. After more than one hundred combined years of teaching statistics, experimental design, and quantitative methodology we wanted to present the important current techniques for analyzing data in the psychological and neuro-sciences. Over the past fifteen years, we have written, developed, integrated, and rewritten our respective class notes that formed the basis of the current volume. The general idea behind this text is to present the modern tools of experimental design and analysis in an approachable and understandable manner and show their applications to complex experimental designs. A lot of the recent developments in the psychological and neuro-sciences came from the advancement of brain imaging techniques because they present many challenges for traditional methods of hypothesis testing. This volume hopes to give students and researchers alike the basic tools necessary to analyze the complex designs that have become prevalent in psychological experiments.

In this book we decided to take a non-traditional approach to presenting statistical concepts and each method is linked to real-world applications and real (albeit simplified) examples of the technique. Because we want students to comprehend statistical techniques, rather than becoming computers, we have often ‘massaged’ the data in order to create examples with ‘nice’ numbers. With these nice numbers, computations become easy (e.g., mean, standard deviations, mean squares become simple integers) and therefore students can concentrate their attention on the meaning of statistical concepts rather than wasting their time with obtuse computational formulas.

Because statistical practices are changing, we decided to present new and essential notions of data analysis that are rarely presented in standard statistical texts, such as contrast analysis, mixed effect models, and cross-validation techniques (e.g., Monte-Carlo methods, permutation tests, and Bootstrap).

Also, in order to keep the cost of the book as low as possible, we have decided to include several chapters as online resources only. These include advanced topics such as matrix algebra and the general linear model as well as a workbook and SAS, R, and SPSS companions.

Acknowledgements

As always, there are many individuals who need to be thanked for either directly or indirectly assisting with this work. First we would like to thank our students, both in Dallas, in Dijon, and elsewhere who have taken our courses, who have read and tried to understand the material presented and given us suggestions to improve and clarify the information presented.

We wish to thank our colleagues for their comments, their conversations, their encouragement and their support. Although no doubt missing several, we wish to specifically thank Lucille Chanquoy, Joseph Dunlop, Anjali Krishnan, Nils Pénard, Mette Posamentier, Dana Roark, and John Vokey. A special mention is due for the most devoted reader who was indubitably Dr Lynne Williams who detected a frighteningly large number of typos, computational errors, and multiple examples of plain incoherence. A special thanks to the editorial staff at OUP, specifically Jonathan Crowe and Helen Tyas. Their patience, constant support, good humor, and help were invaluable. We also wish to thank Robin Watson, our copy-editor. Finding such an individual with absolute dedication to the proper use of the semicolon is rare, indeed, and was very much appreciated.

Despite all the outstanding help that we have received, it is still likely that we have left some marks of our collective linguistic and mathematical creativity (we do not call these things errors anymore!). And, in a collective work such as this, each of the authors takes comfort in the fact that any trace of such creativity can always be attributed to one of the other coauthors.

Contents

1	Introduction to experimental design	1
1.1	Introduction	1
1.2	Independent and dependent variables	2
1.3	Independent variables	4
1.3.1	Independent variables manipulated by the experimenter	4
1.3.2	Randomization	5
1.3.3	Confounded independent variables	5
1.3.4	'Classificatory' (or 'tag') and 'controlled' independent variables	8
1.3.5	Internal versus external validity	11
1.4	Dependent variables	12
1.4.1	Good dependent variables	13
1.5	Choice of subjects and representative design of experiments	14
1.6	Key notions of the chapter	15
2	Correlation	16
2.1	Introduction	16
2.2	Correlation: overview and example	16
2.3	Rationale and computation of the coefficient of correlation	18
2.3.1	Centering	19
2.3.2	The four quadrants	20
2.3.3	The rectangles and their sum	22
2.3.4	Sum of the cross-products	23
2.3.5	Covariance	24
2.3.6	Correlation: the rectangles and the squares	26
2.3.6.1	For experts: going from one formula to the other one	26
2.3.7	Some properties of the coefficient of correlation	27
2.3.8	For experts: why the correlation takes values between -1 and $+1$	27
2.4	Interpreting correlation and scatterplots	28
2.5	The importance of scatterplots	28
2.5.1	Linear and non-linear relationship	28
2.5.2	<i>Vive la différence?</i> The danger of outliers	28
2.6	Correlation and similarity of distributions	30
2.6.1	The other side of the mirror: negative correlation	30
2.7	Correlation and Z-scores	32
2.7.1	Computing with Z-scores: an example	34
2.7.2	Z-scores and perfect correlation	34
2.8	Correlation and causality	36
2.9	Squared correlation as common variance	37
2.10	Key notions of the chapter	37
2.11	Key formulas of the chapter	38
2.12	Key questions of the chapter	38
3	Statistical test: the <i>F</i> test	39
3.1	Introduction	39

3.2	Statistical test	40
3.2.1	The null hypothesis and the alternative hypothesis	40
3.2.2	A decision rule: reject H_0 when it is 'unlikely'	41
3.2.3	The significance level: specifying the 'improbable'	42
3.2.4	Type I and Type II errors: α and β	42
3.2.5	Sampling distributions: Fisher's F	44
3.2.5.1	Empirical (Monte-Carlo) approach	45
3.2.5.2	Theoretical (traditional) approach	47
3.2.6	Region of rejection, region of suspension of judgment, and critical value	50
3.2.7	Using the table of critical values of Fisher's F	52
3.2.8	Summary of the procedure for a statistical test	53
3.2.9	Permutation tests: how likely are the results?	53
3.3	Not zero is not enough!	55
3.3.1	Shrunken and adjusted r values	56
3.3.2	Confidence interval	57
3.3.2.1	Fisher's Z transform	58
3.3.3	How to transform r to Z : an example	58
3.3.4	Confidence intervals with the bootstrap	59
3.4	Key notions of the chapter	61
3.5	New notations	61
3.6	Key formulas of the chapter	61
3.7	Key questions of the chapter	62
4	Simple linear regression	63
4.1	Introduction	63
4.2	Generalities	63
4.2.1	The equation of a line	63
4.2.2	Example of a perfect line	64
4.2.3	An example: reaction time and memory set	65
4.3	The regression line is the 'best-fit' line	67
4.4	Example: reaction time and memory set	68
4.5	How to evaluate the quality of prediction	70
4.6	Partitioning the total sum of squares	72
4.6.1	Generalities	72
4.6.2	Partitioning the total sum of squares	73
4.6.3	Degrees of freedom	75
4.6.4	Variance of regression and variance of residual	75
4.6.5	Another way of computing F	75
4.6.6	Back to the numerical example	76
4.6.6.1	Index F	77
4.7	Mathematical digressions	77
4.7.1	Digression 1: finding the values of a and b	77
4.7.2	Digression 2: the mean of $\hat{Y}$ is equal to the mean of Y	80
4.7.3	Digression 3: the residuals ($Y - \hat{Y}$) and the predicted values $\hat{Y}$ are uncorrelated	80
4.7.4	Digression 4: $r_{\hat{Y}, Y} = r_{X, Y}$	82
4.8	Key notions of the chapter	84
4.9	New notations	84
4.10	Key formulas of the chapter	84
4.11	Key questions of the chapter	85

5	Orthogonal multiple regression	86
5.1	Introduction	86
5.2	Generalities	87
5.2.1	The equation of a plane	87
5.2.2	Example of a perfect plane	88
5.2.3	An example: retroactive interference	90
5.3	The regression plane is the 'best-fit' plane	92
5.4	Back to the example: retroactive interference	93
5.5	How to evaluate the quality of the prediction	96
5.5.1	How to evaluate the importance of each independent variable in the prediction	98
5.5.2	How to evaluate the importance of each independent variable for the dependent variable	99
5.5.3	From the $r_{\hat{Y}}$ coefficients to the r_Y coefficients	100
5.6	F tests for the simple coefficients of correlation	100
5.7	Partitioning the sums of squares	101
5.7.1	What is a score made of?	102
5.7.2	The score model	102
5.7.3	Life is simple when X and T are orthogonal: partitioning the sum of squares regression	103
5.7.4	Degrees of freedom	104
5.7.5	Mean squares	104
5.7.6	The return of F	104
5.7.7	Back to the example	105
5.8	Mathematical digressions	107
5.8.1	Digression 1: finding the values of a , b , and c	107
5.9	Key notions of the chapter	110
5.10	New notations	110
5.11	Key formulas of the chapter	110
5.12	Key questions of the chapter	111
6	Non-orthogonal multiple regression	112
6.1	Introduction	112
6.2	Example: age, speech rate and memory span	112
6.3	Computation of the regression plane	113
6.4	How to evaluate the quality of the prediction	116
6.4.1	How to evaluate the importance of each independent variable in the prediction	117
6.4.2	The specific contribution of each independent variable: the semi-partial contribution	118
6.5	Semi-partial correlation as increment in explanation	121
6.5.1	Alternative formulas for the semi-partial correlation coefficients	123
6.6	F tests for the semi-partial correlation coefficients	124
6.7	What to do with more than two independent variables	125
6.7.1	Computing semi-partial correlation with more than two independent variables	125
6.7.2	Multicollinearity: a specific problem with non-orthogonal independent variables	126
6.8	Bonus: partial correlation	127
6.9	Key notions of the chapter	128
6.10	New notations	128

6.11	Key formulas of the chapter	128
6.12	Key questions of the chapter	129
7	ANOVA one factor: intuitive approach and computation of F	130
7.1	Introduction	130
7.2	Intuitive approach	130
7.2.1	An example: mental imagery	130
7.2.2	An index of effect of the independent variable	131
7.3	Computation of the F ratio	133
7.3.1	Notation, etc.	133
7.3.2	Distances from the mean	135
7.3.3	A variance refresher	136
7.3.4	Back to the analysis of variance	137
7.3.5	Partition of the total sum of squares	138
7.3.5.1	Proof of the additivity of the sum of squares	139
7.3.5.2	Back to the sum of squares	140
7.3.6	Degrees of freedom	140
7.3.6.1	Between-group degrees of freedom	141
7.3.6.2	Within-group degrees of freedom	141
7.3.6.3	Total number of degrees of freedom	142
7.3.7	Index F	142
7.4	A bit of computation: mental imagery	142
7.5	Key notions of the chapter	144
7.6	New notations	144
7.7	Key formulas of the chapter	145
7.8	Key questions of the chapter	146
8	ANOVA, one factor: test, computation, and effect size	147
8.1	Introduction	147
8.2	Statistical test: a refresher	147
8.2.1	General considerations	147
8.2.2	The null hypothesis and the alternative hypothesis	148
8.2.3	A decision rule: reject H_0 when it is 'unlikely'	148
8.2.4	Sampling distributions: the distributions of Fisher's F	148
8.2.5	Region of rejection, region of suspension of judgment, critical value	149
8.3	Example: back to mental imagery	149
8.4	Another more general notation: $\mathcal{A}$ and $S(\mathcal{A})$	152
8.5	Presentation of the ANOVA results	153
8.5.1	Writing the results in an article	154
8.6	ANOVA with two groups: F and t	154
8.6.1	For experts: Fisher and Student ...Proof	155
8.6.2	Another digression: F is an average	156
8.7	Another example: Romeo and Juliet	157
8.8	How to estimate the effect size	161
8.8.1	Motivation	161
8.8.2	R and η	162
8.8.2.1	Digression: $R^2_{Y \cdot \mathcal{A}}$ is a coefficient of correlation	163
8.8.2.2	F and $R^2_{Y \cdot \mathcal{A}}$	164
8.8.3	How many subjects? Quick and dirty power analysis	165
8.8.4	How much explained variance? More quick and dirty power analysis	166

8.9	Computational formulas	167
8.9.1	Back to Romeo and Juliet	167
8.9.2	The ‘numbers in the squares’	168
8.9.2.1	Principles of construction	168
8.9.2.2	‘Numbers in the squares’ and the Universe ...	169
8.10	Key notions of the chapter	170
8.11	New notations	170
8.12	Key formulas of the chapter	170
8.13	Key questions of the chapter	171
9	ANOVA, one factor: regression point of view	172
9.1	Introduction	172
9.2	Example 1: memory and imagery	173
9.3	Analysis of variance for Example 1	173
9.4	Regression approach for Example 1: mental imagery	176
9.5	Equivalence between regression and analysis of variance	180
9.6	Example 2: Romeo and Juliet	182
9.7	If regression and analysis of variance are one thing, why keep two different techniques?	185
9.8	Digression: when predicting Y from $M_{a\cdot}$, $b = 1$, and $a_{\text{regression}} = 0$	185
9.8.1	Remember ...	185
9.8.2	Rewriting SS_X	186
9.8.3	Rewriting SCP_{YX}	186
9.8.4	$a_{\text{regression}} = 0$	187
9.8.5	Recap	187
9.9	Multiple regression and analysis of variance	187
9.10	Key notions of the chapter	189
9.11	Key formulas of the chapter	189
9.12	Key questions of the chapter	189
10	ANOVA, one factor: score model	191
10.1	Introduction	191
10.1.1	Motivation: why do we need the score model?	191
10.1.2	Decomposition of a basic score	191
10.1.3	Fixed effect model	192
10.1.4	Some comments on the notation	194
10.1.5	Numerical example	194
10.1.6	Score model and sum of squares	195
10.1.7	Digression: why ϑ_a^2 rather than σ_a^2 ?	196
10.2	ANOVA with one random factor (Model II)	197
10.2.1	Fixed and random factors	197
10.2.2	Example: $S(\mathcal{A})$ design with $\mathcal{A}$ random	198
10.3	The score model: Model II	198
10.4	$F < 1$ or the strawberry basket	199
10.4.1	The strawberry basket	200
10.4.2	A hidden factor augmenting error	200
10.5	Size effect coefficients derived from the score model: ω^2 and ρ^2	201
10.5.1	Estimation of $\omega_{\mathcal{A}\cdot Y}^2$	202

10.5.2	Estimating $\rho_{A,Y}^2$	203
10.5.3	Negative values for ω and ρ	203
10.5.4	Test for the effect size	203
10.5.5	Effect size: which one to choose?	203
10.5.6	Interpreting the size of an effect	204
10.6	Three exercises	204
10.6.1	Images ...	204
10.6.2	The fat man and not so very nice numbers ...	205
10.6.3	How to choose between fixed and random-taking off with Elizabeth Loftus ...	206
10.7	Key notions of the chapter	208
10.8	New notations	208
10.9	Key formulas of the chapter	209
10.10	Key questions of the chapter	210
11	Assumptions of analysis of variance	211
11.1	Introduction	211
11.2	Validity assumptions	211
11.3	Testing the homogeneity of variance assumption	213
11.3.1	Motivation and method	213
11.4	Example	214
11.4.1	One is a bun ...	214
11.5	Testing normality: Lilliefors	217
11.6	Notation	218
11.7	Numerical example	219
11.8	Numerical approximation	220
11.9	Transforming scores	220
11.9.1	Ranks	221
11.9.2	The log transform	221
11.9.3	Arcsine transform	221
11.10	Key notions of the chapter	222
11.11	New notations	222
11.12	Key formulas of the chapter	222
11.13	Key questions of the chapter	222
12	Analysis of variance, one factor: planned orthogonal comparisons	224
12.1	Introduction	224
12.2	What is a contrast?	225
12.2.1	How to express a research hypothesis as a contrast	227
12.2.1.1	Example: rank order	228
12.2.1.2	A bit harder	228
12.3	The different meanings of alpha	228
12.3.1	Probability in the family	229
12.3.2	A Monte-Carlo illustration	230
12.3.3	The problem with replications of a meaningless experiment: 'alpha and the captain's age'	231
12.3.4	How to correct for multiple comparisons: Šidák and Bonferroni, Boole, Dunn	232
12.4	An example: context and memory	234
12.4.1	Contrasted groups	235

12.5	Checking the independence of two contrasts	236
12.5.1	For experts: orthogonality of contrasts and correlation	236
12.6	Computing the sum of squares for a contrast	237
12.7	Another view: contrast analysis as regression	238
12.7.1	Digression: rewriting the formula of $R_{Y \cdot \psi}^2$	238
12.8	Critical values for the statistical index	240
12.9	Back to the context ...	241
12.10	Significance of the omnibus F vs significance of specific contrasts	244
12.11	How to present the results of orthogonal comparisons	245
12.12	The omnibus F is a mean!	246
12.13	Sum of orthogonal contrasts: sub-design analysis	246
12.13.1	Sub-design analysis: an example	247
12.14	Trend analysis	248
12.15	Key notions of the chapter	251
12.16	New notations	251
12.17	Key formulas of the chapter	251
12.18	Key questions of the chapter	252
13	ANOVA, one factor: planned non-orthogonal comparisons	253
13.1	Introduction	253
13.2	The classical approach	254
13.2.1	Šidák and Bonferonni, Boole, Dunn tests	254
13.2.2	Splitting up $\alpha[PF]$ with unequal slices	255
13.2.3	Bonferonni <i>et al.</i> : an example	256
13.2.4	Comparing all experimental groups with the same control group: Dunnett's test	257
13.3	Multiple regression: the return!	258
13.3.1	Multiple regression: orthogonal contrasts for Romeo and Juliet	258
13.3.2	Multiple regression vs classical approach: non-orthogonal contrasts	262
13.3.3	Multiple regression: non-orthogonal contrasts for Romeo and Juliet	263
13.4	Key notions of the chapter	266
13.5	New notations	266
13.6	Key formulas of the chapter	267
13.7	Key questions of the chapter	267
14	ANOVA, one factor: <i>post hoc</i> or <i>a posteriori</i> analyses	268
14.1	Introduction	268
14.2	Scheffé's test: all possible contrasts	270
14.2.1	Justification and general idea	270
14.2.2	An example: Scheffé test for Romeo and Juliet	272
14.3	Pairwise comparisons	273
14.3.1	Tukey test	273
14.3.1.1	Digression: What is F_{range} ?	274
14.3.1.2	An example: Tukey test for Romeo and Juliet	275
14.3.2	The Newman-Keuls test	276
14.3.3	An example: taking off ...	277
14.3.4	Duncan test	279

14.4	Key notions of the chapter	280
14.5	New notations	280
14.6	Key questions of the chapter	281
15	More on experimental design: multi-factorial designs	282
15.1	Introduction	282
15.2	Notation of experimental designs	283
15.2.1	Nested factors	284
15.2.2	Crossed factors	285
15.3	Writing down experimental designs	285
15.3.1	Some examples	285
15.4	Basic experimental designs	286
15.5	Control factors and factors of interest	287
15.6	Key notions of the chapter	289
15.7	Key questions of the chapter	289
16	ANOVA, two factors: $\mathcal{A} \times \mathcal{B}$ or $S(\mathcal{A} \times \mathcal{B})$	290
16.1	Introduction	290
16.2	Organization of a two-factor design: $\mathcal{A} \times \mathcal{B}$	292
16.2.1	Notations	293
16.3	Main effects and interaction	294
16.3.1	Main effects	294
16.3.2	Interaction	295
16.3.3	Example without interaction	295
16.3.4	Example with interaction	296
16.3.5	More about the interaction	297
16.4	Partitioning the experimental sum of squares	297
16.4.1	Plotting the pure interaction	299
16.5	Degrees of freedom and mean squares	299
16.6	The score model (Model I) and the sums of squares	301
16.7	Example: cute cued recall	303
16.8	Score model II: $\mathcal{A}$ and $\mathcal{B}$ random factors	306
16.8.1	Introduction and review	306
16.8.2	Calculating F when $\mathcal{A}$ and $\mathcal{B}$ are random factors	307
16.8.3	Score model when $\mathcal{A}$ and $\mathcal{B}$ are random	307
16.8.4	$\mathcal{A}$ and $\mathcal{B}$ random: an example	308
16.9	ANOVA $\mathcal{A} \times \mathcal{B}$ (Model III): one factor fixed, one factor random	310
16.9.1	Score model for $\mathcal{A} \times \mathcal{B}$ (Model III)	310
16.10	Index of effect size	311
16.10.1	Index R^2 'global'	311
16.10.2	The regression point of view	312
16.10.2.1	Digression: the sum equals zero	313
16.10.2.2	Back from the digression	314
16.10.3	F ratios and coefficients of correlation	316
16.10.3.1	Digression: two equivalent ways of computing the F ratio	317
16.10.4	Index R^2 'partial'	317
16.10.5	Partitioning the experimental effect	319
16.11	Statistical assumptions and conditions of validity	319
16.12	Computational formulas	320

16.13	Relationships between the names of the sources of variability, df and SS	321
16.14	Key notions of the chapter	322
16.15	New notations	322
16.16	Key formulas of the chapter	323
16.17	Key questions of the chapter	323
17	Factorial designs and contrasts	324
17.1	Introduction	324
17.2	Vocabulary	324
17.3	Fine-grained partition of the standard decomposition	325
17.3.1	An example: back to cute cued recall	325
17.3.1.1	The same old story: computing the sum of squares for a contrast	326
17.3.1.2	Main effect contrasts	326
17.3.1.3	Interaction contrasts	327
17.3.1.4	Adding contrasts: sub-design analysis	327
17.4	Contrast analysis in lieu of the standard decomposition	328
17.5	What error term should be used?	330
17.5.1	The easy case: fixed factors	330
17.5.2	The harder case: one or two random factors	330
17.6	Example: partitioning the standard decomposition	330
17.6.1	Testing the contrasts	331
17.7	Example: a contrast non-orthogonal to the standard decomposition	332
17.8	<i>A posteriori</i> comparisons	333
17.9	Key notions of the chapter	334
17.10	Key questions of the chapter	334
18	ANOVA, one-factor repeated measures design: $S \times A$	335
18.1	Introduction	335
18.2	Examination of the F ratio	337
18.3	Partition of the within-group variability: $S(A) = S + AS$	338
18.4	Computing F in an $S \times A$ design	340
18.5	Numerical example: $S \times A$ design	340
18.5.1	An alternate way of partitioning the total sum of squares	342
18.6	Score model: Models I and II for repeated measures designs	343
18.7	Effect size: R , R , and R	343
18.8	Problems with repeated measures	344
18.8.1	Carry-over effects	345
18.8.2	Pre-test and Post-test	345
18.8.3	Statistical regression, or regression toward the mean	346
18.9	Score model (Model I) $S \times A$ design: A fixed	346
18.10	Score model (Model II) $S \times A$ design: A random	347
18.11	A new assumption: sphericity (circularity)	348
18.11.1	Sphericity: intuitive approach	348
18.11.2	Box's index of sphericity: ϵ	349
18.11.3	Greenhouse–Geisser correction	350
18.11.4	Extreme Greenhouse–Geisser correction	351

18.11.5	Huynh–Feldt correction	351
18.11.6	Stepwise strategy for sphericity	351
18.12	An example with computational formulas	352
18.13	Another example: proactive interference	353
18.14	Key notions of the chapter	354
18.15	New notations	355
18.16	Key formulas of the chapter	355
18.17	Key questions of the chapter	357
19	ANOVA, two-factor completely repeated measures: $S \times \mathcal{A} \times \mathcal{B}$	358
19.1	Introduction	358
19.2	Example: plungin’!	358
19.3	Sum of squares, means squares and F ratios	359
19.4	Score model (Model I), $S \times \mathcal{A} \times \mathcal{B}$ design: $\mathcal{A}$ and $\mathcal{B}$ fixed	361
19.5	Results of the experiment: plungin’	362
19.6	Score model (Model II): $S \times \mathcal{A} \times \mathcal{B}$ design, $\mathcal{A}$ and $\mathcal{B}$ random	367
19.7	Score model (Model III): $S \times \mathcal{A} \times \mathcal{B}$ design, $\mathcal{A}$ fixed, $\mathcal{B}$ random	368
19.8	Quasi- $F : F'$	369
19.9	A cousin F''	370
19.9.1	Digression: what to choose?	371
19.10	Validity assumptions, measures of intensity, key notions, etc.	371
19.11	New notations	371
19.12	Key formulas of the chapter	372
20	ANOVA, two factor partially repeated measures: $S(\mathcal{A}) \times \mathcal{B}$	373
20.1	Introduction	373
20.2	Example: bat and hat	375
20.3	Sums of squares, mean squares, and F ratios	377
20.4	The comprehension formula routine	378
20.5	The 13-point computational routine	380
20.6	Score model (Model I), $S(\mathcal{A}) \times \mathcal{B}$ design: $\mathcal{A}$ and $\mathcal{B}$ fixed	381
20.7	Score model (Model II), $S(\mathcal{A}) \times \mathcal{B}$ design: $\mathcal{A}$ and $\mathcal{B}$ random	382
20.8	Score for Model III, $S(\mathcal{A}) \times \mathcal{B}$ design: $\mathcal{A}$ fixed and $\mathcal{B}$ random	383
20.9	Coefficients of intensity	384
20.10	Validity of $S(\mathcal{A}) \times \mathcal{B}$ designs	384
20.11	Prescription	385
20.12	Key notions of the chapter	385
20.13	Key formulas of the chapter	385
20.14	Key questions of the chapter	386
21	ANOVA, nested factorial design: $S \times \mathcal{A}(\mathcal{B})$	387
21.1	Introduction	387
21.2	Example: faces in space	388
21.2.1	A word of caution: it is very hard to be random	388

21.3	How to analyze an $S \times A(B)$ design	389
21.3.1	Sums of squares	389
21.3.2	Degrees of freedom and mean squares	390
21.3.3	F and quasi- F ratios	391
21.4	Back to the example: faces in space	392
21.5	What to do with A fixed and B fixed	393
21.6	When A and B are random factors	394
21.7	When A is fixed and B is random	394
21.8	New notations	394
21.9	Key formulas of the chapter	394
21.10	Key questions of the chapter	395
22	How to derive expected values for any design	396
22.1	Introduction	396
22.2	Crossing and nesting refresher	397
22.2.1	Crossing	397
22.2.2	Nesting	397
22.2.2.1	A notational digression	397
22.2.2.2	Back to nesting	397
22.3	Finding the sources of variation	398
22.3.1	Sources of variation. Step 1: write down the formula	398
22.3.2	Sources of variation. Step 2: elementary factors	398
22.3.3	Sources of variation. Step 3: interaction terms	399
22.4	Writing the score model	400
22.5	Degrees of freedom and sums of squares	401
22.5.1	Degrees of freedom	401
22.5.2	Sums of squares	401
22.5.2.1	Comprehension formulas	402
22.5.2.2	Computational formulas	402
22.5.2.3	Computing in a square	402
22.6	Example	403
22.7	Expected values	404
22.7.1	Expected value: Step 1	404
22.7.2	Expected value: Step 2	405
22.7.3	Expected value: Step 3	405
22.7.4	Expected value: Step 4	406
22.8	Two additional exercises	406
22.8.1	$S(A \times B(C))$: A and B fixed, C and S random	407
22.8.2	$S(A \times B(C))$: A fixed, B , C , and S random	407
Appendices		409
A	Descriptive statistics	411
A.1	Introduction	411
A.2	Some formal notation	414
A.2.1	Notations for a score	414
A.2.2	Subjects can be assigned to different groups	414
A.2.3	The value of a score for subjects in multiple groups is $Y_{a,s}$	414
A.2.4	The summation sign is $\sum$	414

A.3	Measures of central tendency	415
A.3.1	Mean	416
A.3.2	Median	416
A.3.3	Mode	417
A.3.4	Measures of central tendency recap	417
A.4	Measures of dispersion	417
A.4.1	Range	417
A.4.2	Sum of squares	417
A.4.3	Variance	419
A.4.4	Standard deviation	421
A.5	Standardized scores alias Z-scores	422
A.5.1	Z-scores have a mean of 0, and a variance of 1	422
A.5.2	Back to the Z-scores	423
B	The sum sign: Σ	424
B.1	Introduction	424
C	Elementary probability: a refresher	426
C.1	A rough definition	426
C.2	Some preliminary definitions	427
C.2.1	Experiment, event and sample space	427
C.2.2	More on events	427
C.2.3	Or, and, union and intersection	430
C.3	Probability: a definition	430
C.4	Conditional probability	432
C.4.1	Bayes' theorem	434
C.4.2	Digression: proof of Bayes' theorem	435
C.5	Independent events	436
C.6	Two practical counting rules	438
C.6.1	The product rule	438
C.6.2	Addition rule	439
C.7	Key notions of the chapter	440
C.8	New notations	441
C.9	Key formulas of the chapter	441
C.10	Key questions of the chapter	442
D	Probability distributions	443
D.1	Random variable	443
D.2	Probability distributions	444
D.3	Expected value and mean	446
D.4	Variance and standard deviation	447
D.5	Standardized random variable: Z-scores	449
D.6	Probability associated with an event	451
D.7	The binomial distribution	453
D.8	Computational shortcuts	456
D.8.1	Permutation, factorial, combinations and binomial coefficient	456
D.9	The 'normal approximation'	460
D.9.1	Digression: the equation of the normal distribution	460
D.10	How to use the normal distribution	462
D.11	Computers and Monte-Carlo	465

D.12	Key notions of the chapter	467
D.13	New notations	467
D.14	Key formulas of the chapter	468
D.15	Key questions of the chapter	468
E	The binomial test	470
E.1	Measurement and variability in psychology	470
E.1.1	<i>Kiwi</i> and <i>Koowoo</i> : a binomial problem	471
E.1.2	Statistical test	472
E.2	Coda: the formal steps of a test	475
E.3	More on decision making	477
E.3.1	Explanation of 'tricky' wording	479
E.4	Monte-Carlo binomial test	479
E.4.1	A test for large N : normal distribution	480
E.5	Key notions of the chapter	483
E.6	New notations	483
E.7	Key questions of the chapter	484
F	Expected values	485
F.1	What is this chapter about?	485
F.2	A refresher:	485
F.3	Expected values: the works for an $S(A)$ design	488
F.3.1	A refresher	488
F.3.2	Another refresher: score model	489
F.3.3	Back to the expected values	490
F.3.4	Evaluating A	490
F.3.5	Evaluating AS	492
F.3.6	Evaluating 1	493
F.3.7	Expected value of the sums of squares	494
F.3.8	Expected value of the mean squares	494
	Statistical tables	495
Table 1	The standardized normal distribution	497
Table 2	Critical values of Fisher's F	499
Table 3	Fisher's Z transform	502
Table 4	Lilliefors test of normality	505
Table 5	Šidàk's test	506
Table 6	Bonferroni's test	509
Table 7	Trend Analysis: orthogonal polynomials	512
Table 8	Dunnett's test	513
Table 9	F_{range} distribution	514
Table 10	Duncan's test	515
	References	518
	Index	531

1

Introduction to experimental design

1.1 Introduction

There are many ways of practicing psychology, but *experimentation* is the best and the easiest way to discover cause and effect relationships. The general principle is simple even if the implementation is not always that easy. In an experiment, groups of subjects are treated in exactly the same way in every respect except one: *the experimental* treatment itself. The observed differences in the subjects' behavior reflect, as a consequence, the effect of the experimental treatment. The statistical analysis that follows the experiment should reveal if there are differences between the experimental conditions. Actually, the *rôle* of statistics can be reduced to two functions: on the one hand, to *describe* correctly the results of research (this is the function of *descriptive statistics*); and on the other hand, to *evaluate how consistent* are the effects found (this is the function of *inferential statistics*). Thus we might say that descriptive statistics is a way of discovering (literally, to remove the cover from something) the results of a study or an experiment in order to show what is essential. From this point of view, inferential statistics has the task of attesting to the 'solidity' or 'reliability' of conclusions arising from descriptive analysis.

Statistics thus appears as an auxiliary of the normal process of science—as an aid to scientific reasoning. It is neither the bogeyman nor the panacea that some see in it. But this mixture of superstitious fear and magical respect can be an obstacle to learning of statistics and can easily generate misinterpretations. Among the uses and abuses of statistics, a frequent error, and not only among students, is to confuse *significance* and *relevance*.¹ The problem is that for the difference between the means of two groups to be statistically significant it suffices that the two groups be quite large (we will see why later). Nevertheless, such a significant difference can be entirely without interest. For example, the time to read one word is, let us say, 500 ms (cf. Anderson, 1980). Suppose we have two very large groups with means of 498 ms and 502 ms. If the groups are sufficiently large, 'the difference will be significant', as we would write in an

¹ If you do not know or have forgotten what 'significance' means, you can assume that it means something like 'It is very unlikely that the results I have observed are due to chance'.

experimental report in a journal, but it would be judged irrelevant by researchers in the field because of its small size (difference is only about 2 minutes per 300 pages, or a couple of minutes per book). This problem is detailed further in Chapter 3 (see the stimulating book by McCloskey and Ziliak, 2008, if you want to read more on this theme).

The rest of this chapter provides the vocabulary necessary to understand the construction and treatment of *experimental designs*. First, we introduce the key concepts of *independent variables* and *dependent variables* whose details we shall provide later. Next, we discuss some classic problems involving flawed experimental designs and the selection of subjects.

1.2 Independent and dependent variables

Let us start with an example: for several decades some psychologists have passionately and intensely pursued the study of verbal learning (Underwood, 1983). One of their favorite techniques has involved asking subjects to memorize word pairs such as ‘cabbage–bicycle’; ‘kindness–carrot’; ‘durable–rabbit’. Then the subjects are asked to give the second word in response to the first. A psychologist observing this could think:

This kind of task resembles rote-learning. The subjects should be able to improve their performance if they would give a ‘meaning’ to each word-association, as they would do, for example, by trying to form an image linking the two words of a pair.

In fact, this psychologist has just formulated a research hypothesis. In addition to asking a question, the psychologist expects a certain response: namely that memorization with images leads to better performance than memorization without images.

To answer this question, the psychologist could design an experiment. First, the initial question should be translated into a testable proposition (that is, the psychologist needs to *operationalize* the original idea): if subjects memorize word pairs using images, then their performance will be superior to that of subjects who memorize them ‘by rote’. (Note that the psychologist now contrasts ‘learning with images’ with ‘learning by rote’). Hence, there are two experimental conditions: (1) ‘memorization with images’ and (2) ‘memorization by rote’. We shall say that these two experimental conditions define two *levels* of the *independent variable* ‘encoding strategy’.

Independent variable is abbreviated I.V., and is used almost interchangeably with the term ‘*experimental factor*’, or just ‘*factor*’ for short. Traditionally, we use ‘independent variable’ in a methodological context and ‘factor’ in a statistical context, but this is not a hard and fast rule.

Note: *One* independent variable has *several* levels corresponding to several experimental conditions (in the word ‘variable’ we find the root ‘to vary’).

When experimenters vary the independent variable, they expect to observe a change in some aspect of the subjects’ behavior. This aspect of behavior that the researcher observes (or measures) represents the *dependent variable* (dependent, because the researcher believes that it depends on the independent variable). If we remember that one of the goals of experimentation is to expose cause-and-effect relationships, then we can say that:

- the independent variable represents the presumed cause,
- the dependent variable reveals the supposed effect.

In other words, the dependent variable (abbreviated D.V.) expresses the effect of the independent variable on the subjects' behavior. Here are a few examples:

1. Research hypothesis 1

Suppose we present subjects with a different verbal message in each ear and ask them to 'shadow' the message coming in one ear (that is, to recite it as they hear it). This task will be more difficult when the two messages are in the same voice than when they are read by quite different voices.

- *Independent variable:* Similarity of voices reading the messages.

Note that the independent variable can involve more than two conditions. Here, for example, we might decide to use the following conditions: (1) two identical voices, (2) two different men's voices, (3) a man's voice and a woman's voice. We could imagine other conditions (try to come up with some!).

- *Dependent variable:* For example, we could count the number of errors the subject makes repeating the shadowed message (incidentally, you might think about how we should define 'error' in this context).

Note that here we could translate the notion of 'task difficulty' in more than one way. For example, we could ask the subjects themselves to evaluate the difficulty on a five-point (or two-point, or ten-point) scale. We could measure heart rate, or GSR responses (GSR refers to the galvanic skin response that indicates hand palm sweating). We could imagine—and this is a necessity for an experimenter—many other dependent variables. Try coming up with some. Then ask yourself, how should we evaluate the quality and relevance of a dependent variable? How should we decide, for example, which of two possible dependent variables is better, and in what sense? This may take a little time, but it is worth the trouble.

2. Research hypothesis 2

What we learn now interferes with what we will learn later (Underwood, 1983).

- *Independent variable:* The number of word lists learned before learning the 'test list'.

Note that the independent variable can be quantitative. How many levels of this variable would you use?

- *Dependent variable:* The number of words (or percentage of words) correctly recalled from the test list. (Does it matter whether we use number or percentage?)

3. Research hypothesis 3

The more one studies, the more one retains (Ebbinghaus, 1985).

- *Independent variable:* The number of practice trials devoted to a list of 16 nonsense syllables. Ebbinghaus used the following numbers of trials: 0, 8, 16, 24, 32, 42, 53, 64.

Note that one could choose other independent variables here, such as the time spent studying. By the way, do those two translations of the independent variable appear equivalent to you?

- *Dependent variable:* Number of trials (or the time) required to re-learn the list 24 hours after the first learning session.

4. Research hypothesis 4

Chess masters are familiar with a large number of chess games. Because of their familiarity with patterns of play, they can easily retain the placement of pieces on the board, even after just one glance at the position, when the pattern is part of an ongoing game (DeGroot, 1965). This is something that average players cannot do.

- *Independent variable:* Here we have two independent variables. The first involves the level of the player, with possible levels of: novice, moderately good player, highly experienced player, and master (the Chess Federation keeps won-lost records—behavioral data—for the best players so that the top levels of skill are easily operationalized). The second independent variable contrasts board positions drawn from actual games with random patterns of the same set of chess pieces. By ‘crossing’ these two independent variables we obtain eight experimental conditions. (Specify them.)
- *Dependent variable:* One possibility would be the number of pieces correctly placed on the board 5 minutes after the subject was shown the board pattern.

As a general rule the dependent variable is presented as a *function* of the independent variable. For example, the title of an article might read: ‘Processing time [dependent variable] as influenced by the number of elements in a visual display [independent variable]’ (Atkinson *et al.*, 1969); or ‘Short-term memory for word sequences [dependent variable] as a function of acoustic, semantic, and formal similarity [independent variables]’ (Baddeley, 1966); or ‘The effects of familiarity and practice [independent variables] on naming pictures of objects [dependent variable]’ (Bartram, 1973); or ‘Time of day [independent variable] effects on performance in a range of tasks [dependent variables]’ (Blake, 1967). For practice you might try taking one of your psychology textbooks and look through the bibliography, trying to find the independent variables and dependent variables specified in titles. (Of course, not all titles will contain independent variables and dependent variables explicitly.)

1.3 Independent variables

1.3.1 Independent variables manipulated by the experimenter

As we noted above, one or more independent variables represent the cause(s) entering into the cause-effect relation explored in the experiment. The aim of the researcher is to demonstrate the effect of the independent variable on the dependent variable *without ambiguity*. The most efficient way to do that is to have experimental groups that are *equivalent in all respects* except for the differences introduced by the levels of the independent variable; that is, to control all the potential independent variables—all the independent variables that might possibly have an influence on the dependent variable. When the experimenter is able to manipulate the independent variable, and only the independent variable, we say that the independent variable is *under the control* of the experimenter. This is the ideal case. In fact, when experimenters manipulate the independent variable and nothing else, they are assured that the effect they observed comes only from the action of the independent variable.

But how can we be certain that we are controlling ‘all the potential independent variables?’ As a rule we cannot know that *a priori* (if we knew all that, we would not need to

design an experiment). In practice we try to control all the variables that we know (or suppose) to affect the dependent variable.

The choice of what variable to control depends also upon the goal of the experiment. In general, an experiment is designed in order to defend a theory against one or more competitor theories. And so, instead of controlling all the potential independent variables, we try to control only the independent variables that bear on the theoretical issues at stake (see below the section on confounded independent variables).

1.3.2 Randomization

When the independent variable is manipulated by the experimenter (independent variables concerning attributes of the subjects, such as age or sex, obviously, cannot be manipulated²), we want that to be done in such a way that the pairing of subjects and level of the independent variable is arbitrary. That is, we cannot know which subject will be in which group *before* the treatment condition is assigned to a given experimental group. In particular, the assignment of subjects to groups should be random. If, for example, the experimenter wishes to divide the subjects into two groups, he or she can write their names on slips of paper and put the slips into a hat, shake vigorously (the hat), and ask an innocent person to draw the slips from the hat. The first set drawn will constitute the first group, etc. One could as well number the subjects (to protect the innocent) and consult a table of random permutation of N numbers, putting the first $N/2$ subject numbers in one group, and the second $N/2$ numbers in the other. (Pick a permutation from the table in some arbitrary and unbiased way, for example by using another source of random numbers, or by asking a friend to stick a pen with a thin strip of paper dangling from it in one of the rows of numbers while blindfolded.) By assigning the subjects randomly to groups, we avoid the systematic pairing of experimental group and some characteristic of the subjects. For example, in an experiment on memory where we were varying number of learning trials on the list of words, we could ask the subjects to come to the laboratory when they wanted to do the experiment. Suppose we decide to assign the first batch of subjects who arrive at the lab to the first experimental condition (for example, 8 trials). Maybe these subjects are more motivated, or more anxious, than subjects who arrive later (and who are assigned to the 64-trial condition). This characteristic may lead them to behave in a systematically different way from the later arrivals, and that difference may interfere with the effect of the independent variable. Of course, it is always possible that the two groups might differ from each other even when we assign subjects randomly—it is just that in that case systematic differences are highly unlikely. In addition, random assignment to groups is consistent with the statistical procedures used to evaluate the effect of the independent variable on the dependent variable. In fact, those statistical procedures are based explicitly on the provision that the assignment is random.

1.3.3 Confounded independent variables

If we wish to conclude that the independent variable affects the dependent variable, we must avoid confounding our independent variable with other independent variables. We say that two independent variables are *confounded* if *the levels of one independent variable*

² In a statistical sense, that is!

are associated systematically with the levels of the other. In the preceding example, the order of arrival was confounded with the independent variable being investigated. When two independent variables are confounded it is simply impossible to interpret the results. In fact, a positive result might arise from one independent variable or the other, or from an interaction³ between the two (that is, the effect occurs only for specific combinations of two variables). A negative result (in which there is no difference between the experimental conditions) might arise from compensation in which the effects of the variables cancel each other out, even though each taken alone would affect the dependent variable.

An example: suppose you wish to explore the effects of three methods of teaching arithmetic. You choose three teachers from among the volunteers. (How will you choose them? Are volunteers equivalent to non-volunteers in such a case?) Each teacher learns one of the methods, and then is put in charge of a class. At the end of the year the students are tested on arithmetic. The dependent variable is the result of the test. The independent variable that the experimenter wants to evaluate is teaching method (with three levels). But do the results of the experiment permit us to evaluate teaching methods? (You should think how you would answer that).

You have—we hope—come to the conclusion that nothing is less sure. In fact, the teachers, aside from the methods that they used, differ in a large number of personal characteristics, which could themselves affect the outcome, working either for or against the effects of the independent variable. (What other variables could be confounded in this case with teaching method?) In fact, this experiment does not provide us with the means of separating the effects of the independent variable ‘teaching method’ from the effects of the independent variable ‘personal characteristics of teachers’. *Those two variables are confounded.* The observed results might come from one or the other of the independent variables, as well as an interaction between them. As always, the argument appears trivial in an example, but in practice it is not always easy to untangle confounded independent variables. Underwood and Shaughnessy (1983), who are experts in this area, provide the following recipe: the ability to discern confounded variables comes with practice, and then still more practice Incidentally, you must have noticed that many heated disputes in psychology arise when some researchers accuse other researchers of being misled by confounded and perverse independent variables. In other words, the accusation is that the independent variable the researchers thought they were manipulating was not the one that produced the effect, or that the effect appears only when that independent variable is used in conjunction with some other variable.

Apart from practice and careful thought, one way to track down confounded independent variables lies in the scientific approach of replicating important experiments. If a replication does not lead to the same results as the original, we can suspect the presence of confounded independent variables. Difficulties of replication can arise from the effects of another independent variable interacting with the first. A classic example comes from studies of motivation: a rat learns to negotiate a maze faster when there is a greater reward (in the form of food pellets) at the end. However, this only works with hungry rats. In other words, the independent variable ‘intensity of reward’ has an effect only under certain conditions (see the chapters on experimental designs with two factors for more details on this topic).

The problem of confounded independent variables is particularly clear in the area of language and memory (‘semantic memory’—cf. Shoben, 1982). For example, suppose we

³ The notion of interaction will be detailed later concerning 2-factor designs (see, e.g., Chapter 16).

want to test the hypothesis that the response time (dependent variable) to verify sentences of the form: 'An object is a member of a certain category' (such as 'A bird is an animal') depends on the size of the semantic category, in terms of number of elements (independent variable). Clearly the category 'living thing' is larger than the category 'animal'. These levels of the independent variable 'category size' lead to a verification of the hypothesis (Landauer and Freedman, 1968). However, Collins and Quillian (1969) later suggested that it was not the independent variable 'category size' that was having the effect, but rather the confounded independent variable 'inclusion relation between categories' (that is, while all animals are living things, all living things are not animals). Following that, Smith *et al.* (1974) showed that, in fact, the effect of the two preceding independent variables was to be explained in terms of another confounded independent variable: 'semantic resemblance between categories'. At this point McCloskey (1980) appeared on the scene, demonstrating that the really important independent variable was 'category familiarity'.... Little did they realize at the time, but the next episode in this story brought even more confounded factors.

Serious difficulties in separating independent variables from each other arise when we must use language in our experiments. To convince yourself of this, try to formulate a test of the hypothesis: 'The more concrete the word, the easier it is to memorize'. To test this hypothesis we need to manipulate this independent variable *while excluding concomitant changes in every other* independent variable (for example, the frequency of words in the language, their length, their power to evoke imagery, etc.). We wish you luck (cf. among others, Paivio, 1971; Yuille and Marschark, 1983). The preceding discussion bears on the theme of control of '*parasite*' independent variables. It would be a good idea to review at this point the paragraph on independent variables manipulated by the experimenter.

Here is another example. In a classic study on stress, Brady (1958) put several pairs of rhesus monkeys in seats and kept them there for several hours. The two monkeys in a pair both received an electric shock every 20 seconds or so. One monkey in the pair (whom Brady called the 'executive monkey') could avoid both of them getting shocked by pressing a button before the shock occurred. To determine which monkey should be the 'executive' Brady used a pretest. After assigning monkeys to pairs randomly, he chose as the 'executive' the monkey that learned most quickly to press the button to avoid the shock—in a sense the most 'competent' monkey. Brady wished to test the following hypothesis: 'The executive monkey, having more responsibility, will be more subject to stress'. (What are the independent variable and the dependent variable?) Stress was measured by a number of variables including the incidence of gastric ulcers. There were significantly more ulcers among the executive monkeys than among their passive companions.

Has Brady produced a convincing experiment? *No!* The independent variable 'intensity of stress' is confounded with another independent variable: 'speed of learning' (cf. Weiss, 1972). Because of this the results of Brady's experiment cannot be interpreted unambiguously.

Quite often the confounded independent variable involves the simple fact of participating in an experiment. This is referred to as the 'Hawthorne effect', after the place where the original study was done. (For a history of the Hawthorne effect see Parsons, 1974, 1978). Sometimes the confound involves the subjects' perception of the experimenter's expectations (called the 'Pygmalion' effect after the Greek legend—see Rosenthal, 1977, 1978). We can attempt to minimize these effects by leaving both the experimenter and subjects ignorant of the expected results, or even of the experimental condition they are running or participating in. This is called the 'double-blind' procedure to differentiate it from the usual 'single-blind' procedure in which only the subjects are ignorant of the experimental condition. But even

the double-blind procedure does not prevent the subjects from having some idea of the results expected of them. It is even harder to hide the actual experimental condition from experimenters. For example, if we design an experiment comparing performance of subjects under the influence of neuroleptic drugs with those of normal controls (who took a placebo), the secondary effects of the drugs (trembling, dry mouth, etc.) will reveal the state of the subjects to the experimenter. Note that if we do use a double-blind procedure, we need to debrief the subjects and experimenters afterwards concerning the purpose and results of the experiment and which condition they served in (for ethical reasons which should be apparent).

In the preceding paragraphs we have emphasized the negative aspects of confounded independent variables, and you may surmise that an important attribute of a good experiment is to avoid confounding independent variables. But in some cases we might confound variables on purpose, essentially when we want to demonstrate the possibility of an effect on the dependent variable. Suppose, for example, we wish to show that it is possible to improve the test scores of school children in a disadvantaged area. In that case, we can employ the '*steam roller*' method: choose the very best teachers, change the textbooks, change the teaching methods and ... change everything that could possibly help improve performance (supposing we have the means at our disposal to do all this). Clearly if we get an effect we cannot attribute it to any particular independent variable. However, we will have demonstrated the *possibility* of improving performance. Later we can design particular experiments to analyze the effect by assessing the role of each independent variable and combination of independent variables. 'Steam roller' studies aim essentially at revealing the possibility of modifying behavior (in education, therapy, rehabilitation, etc.), not at its experimental analysis.

1.3.4 'Classificatory' (or 'tag') and 'controlled' independent variables

'Classificatory' (or 'tag') independent variables are to be contrasted with independent variables 'manipulated by the experimenter'. We say that an independent variable is a 'classificatory' or a 'tag' variable when it involves a *natural characteristic* by which the experimenter defines an experimental condition. In particular, descriptive characteristics of subjects can serve as tags and define classificatory independent variables; for example: age, motivation level, intelligence, ability to produce images, socio-economic status, sex, etc. The important point to note about classificatory independent variables is the *impossibility of random assignment* to groups. The subjects are decked out in their 'tags' before the experiment begins. For example, the experimenter will not be able to assign subjects randomly to the 'male' or the 'female' group, or to the groups of subjects over or under the age of 60. The experimenter simply uses the levels of the tag independent variable to divide the subjects into different groups. The classification has been done by Mother Nature, and not by the experimenter. Consequently, the subjects assigned to different groups might well differ systematically in other ways than the one designated by the initial classification, and those differences will be confounded with the original independent variable. Besides, quite often the classificatory independent variable is not exactly the one the researcher wanted to use, but merely represents the independent variable that the researcher *thinks* is affecting behavior. For example, if we compare the performance of boys and girls on tests of visual-spatial ability, our classificatory independent variable is 'biological sex', even though the independent variable we believe causes the differences between groups is actually 'psychosocial sex', which is tied to 'biological sex',

though different. The same for age: we generally use ‘chronological age’ when we wish to evaluate the effects of age in terms of maturation and development.

The statistical procedures used with independent variables manipulated by the experimenter are used as well with classificatory independent variables. However, all the conclusions we can draw from the former cannot be drawn from the latter. In particular, manipulated independent variables can be used to justify ‘cause and effect’ explanations, while this is not the case with classificatory independent variables. Why? Refer to the preceding section on confounded variables and note that classificatory independent variables are inevitably accompanied by a multitude of confounds with parasite independent variables, without the possibility of untangling them.

Unfortunately there are many problems in psychology that are posed in terms of classificatory variables, including a large number of problems of practical importance. We might be doubtful about calling studies involving these classificatory independent variables ‘experimental’, since the experimenter does not actually administer different treatments to (arbitrarily assigned) subjects, but merely *observes* the effects of preexisting conditions on their behavior. To emphasize this point, and to avoid the ensuing multiplication of independent variables, certain authors propose the term ‘quasi-experimental’ to describe studies using classificatory independent variables. Note that the term quasi-experimental covers a wider range than indicated here. Nevertheless this approximate definition captures the essentials. We owe the term quasi-experimental as well as a thoughtful treatment of the problems of applying the experimental approach in natural settings to Donald Campbell (Campbell and Stanley, 1966; Cook and Campbell, 1979). For a useful introduction to these problems, see Cherulnik (1983).

Classificatory independent variables are often, but not exclusively, tied to characteristics of subjects (which is why Underwood, 1975, treats the problems we have been discussing under the heading of ‘subject variables’). However, we will encounter them every time we use preexisting characteristics of subjects as levels of an independent variable (for example, different schools, different regions, different time periods).

Sometimes the very same variable can be used as a manipulated or a classificatory independent variable. For example, if we wish to study the effects of anxiety (independent variable) on response time (dependent variable), we could contrast the results obtained with a group of anxious subjects with those of a group of non-anxious subjects. To obtain these two groups we could select anxious and non-anxious subjects based on a test of anxiety (such as the Manifest Anxiety Scale of the MMPI, Taylor, 1953, with items such as ‘I often sweat more than my friends’ and ‘I truly believe the bogeyman is trying to get me’⁴). This would give us a classificatory independent variable. Alternatively, we could make some subjects anxious in various ways (adrenaline injections, prolonged exposure to a disturbing environment, threats of pop quizzes, scaring them in elevators, etc.). This will correspond to a manipulated independent variable. We could then wonder whether these two versions—these two ‘operational definitions’—of the same variable really represent the same thing: Are we studying the same kind of ‘anxiety’ in both cases?

We can include an example here of a case where the researchers allowed Nature to manipulate the independent variable for them, merely positioning the experimenter correctly to catch the results of the manipulation. It will be clear from the researchers’ concerns that we

⁴ But as you know, he’s only really trying to get you when you don’t do your homework!

are still dealing with a classificatory independent variable. We shall see in the development of the study what they did about that.

Dutton and Aron (1974, see also Wiseman, 2007) wished to study the effects of heightened anxiety (independent variable) on sexual attraction (dependent variable). They found two bridges in the same area of Vancouver, one a suspension bridge swaying high above a deep canyon that was rather scary to cross, and the other a solid wooden bridge across a shallow ravine that fed into the main canyon. The 'subjects' for their study were people who just happened to cross these bridges. Dutton and Aron had an attractive female experimenter accost men as they came off the ends of the bridges. (In a partially double-blind design, the experimenters were unaware of the experimental hypothesis.) The experimenter asked some innocuous questions and had the subjects respond to a Thematic Apperception Test (TAT) with no obvious sexual content (subjects had to tell what was happening in a picture of a young woman covering her face with one hand while reaching with the other). Dutton and Aron reasoned that men coming off the scary bridge would be more anxious than those coming off the not-so-scary bridge, and that that anxious arousal would get translated cognitively into a feeling of attraction for the female experimenter (since the men may not have wanted to admit to themselves that they were scared by the bridge). The experimenter contrived to give most of the subjects her phone number, telling them they could call her for the results of the experiment. The dependent variable was how many men called her the following week. Out of 16 who took her phone number after crossing the control bridge, only 2 called the experimenter; while 9 out of 18 called who had crossed the scary bridge. Such a pattern of results cannot be attributed to chance and so we can conclude that the men from the scary bridge were more likely to call the experimenter.

How to interpret this result? Dutton and Aron concluded that the men's heightened arousal (arising from the anxiety engendered by the scary bridge) led to greater sexual attraction for an attractive female encountered while they were in that emotional state. However, because they were dealing with an independent variable that, strictly speaking, is a classificatory one (though very close to a manipulated one), they had to be sure to include control conditions to rule out plausible alternative explanations. They had to be concerned, for example, that the scary bridge was more of a tourist attraction than the non-scary bridge, but they could argue that tourists probably would not be around to call the next week, and so that would militate against their result. They wanted a convergent check on sexual attraction as a mediating variable, so they scored the subjects' TAT responses for sexual imagery and found a difference (more sexual imagery following the scary bridge). As a further check on whether sexual arousal was involved, Dutton and Aron replicated the study with a male experimenter. Very few subjects in either condition called the male experimenter the next week. Dutton and Aron thought maybe people who cross scary bridges are 'thrill seekers' (a possible confounding independent variable), as opposed to 'timid types'. Maybe thrill seekers call attractive females more often in general than do timid types, quite apart from scary bridges—leading to our results. Therefore Dutton and Aron replicated the experiment using as a control subjects who had crossed the bridge (that is, thrill seekers), but who had crossed it 10 minutes before (and who thus had time to calm down). They got the same result as before: 13 of 20 who had just crossed the bridge called, versus 7 of 23 who had crossed it earlier. Once again, this pattern cannot be attributed to chance.

Dutton and Aron, still concerned about depending for their conclusions on a classificatory independent variable, brought their study into the laboratory, where they could manipulate

the anxiety of subjects who were assigned randomly to conditions (producing a manipulated independent variable). They led male subjects to believe that they (along with an attractive female confederate of the experimenters playing the role of a subject) would be shocked either mildly or severely (two levels of the independent variable) during the experiment. Dutton and Aron then had subjects rate (dependent variable) the attractiveness of their fellow subject (whom they expected would be shocked as well). The subjects reported that the confederate was more attractive when they expected to be shocked more severely—this pattern of results supports Dutton and Aron's hypothesis (there were no effects of the shock level the subjects expected the confederate to receive, leading to a rejection of the 'attractiveness of a damsel in distress' hypothesis).

This is an example of *converging operation* being brought to bear on a single hypothesis. We find the laboratory study more believable because it can be replicated qualitatively in a natural setting (albeit with a classificatory independent variable); and we find the field study more convincing because it can be replicated when its naturalistic classificatory independent variable is replaced with a manipulated independent variable in the laboratory.

1.3.5 Internal versus external validity

In a number of examples up to this point the researchers used only a single independent variable. We might be tempted to conclude that this represents an ideal. Nothing of the sort. Studies involving a single independent variable (a 'single factor', as we say) have the indisputable advantage of simplicity of execution, but may at times seem far removed from any natural setting and from everyday life. To make experiments more realistic we might choose to increase the number of independent variables. By increasing the number of independent variables we approach more natural conditions, and hence increase the realism of the experiment. We say that we are increasing the '*external validity*' or '*ecological validity*' (Brunswick, 1956) of the experiment. We hope in doing that to obtain experimental results that are valid beyond the limits of the laboratory (cf. Lachman *et al.*, 1979, p. 120; Baddeley, 1976, p. 150).

It is useful to specify the meaning of the terms *ecological* or *external validity* in contrast to the term *internal validity* (cf. Campbell and Stanley, 1966; Conrad and Maul, 1981). Internal validity depends on the precision of the experiment, and is used in an experimental context. Essentially, internal validity varies inversely with the size of experimental error. We mean by experimental error the amount of the total variability of the dependent variable to be attributed to causes other than the experimental factors (the independent variables that were controlled). For example, it will be almost inevitable that the subjects in a given condition do not behave in *exactly* the same way, and those differences form part of the experimental error. In other words, internal validity is greater to the extent that the effects observed on the dependent variable can be more confidently attributed to the operations of the independent variable(s). Controlling potential independent variables, assigning subjects randomly to groups, eliminating confounding independent variables—all those have the goal of increasing internal validity (recall the above discussions of those techniques). High internal validity means good experimental designs.

From a certain point of view we could place internal and external validity in opposition. One insists on the strength or clearness of the evidence for an effect; the other on its relevance outside the laboratory. Thus the clarity of an effect contrasts with its generality; just as the

simplicity (generally linked to internal validity) of designs employing one experimental factor contrasts with the complexity (generally linked to external validity) of designs employing several factors.

1.4 Dependent variables

Psychologists observe the effect of the independent variable on the behavior of subjects. But we can observe only a very limited part of the behavior. The specific aspect of the behavior that we will pay attention to and record is called the *dependent variable*. In general, the dependent variable measures some kind of performance: number of words recalled, time used to read a word, intensity of the galvanic skin response (GSR, see p. 3), etc.

It could be possible to argue—as a caricature of empiricism—that an experiment takes the values of the dependent variable literally, and records them only for what they are. We could have, thus, a Psychology of the number of meaningless syllables remembered after a 15 minute interval, of the time to decide if the proposition ‘a canary is a fish’ is true⁵ or not, or even of the number of eye blinks when the syllable ‘ga’ is uttered.⁶

Actually, the dependent variable is chosen in order to *represent* an aspect of the behavior we are interested in. The number of meaningless syllables remembered is of no interest in itself. It represents or expresses the effect of learning or memory which is what we want to know about. Similarly, what interests a researcher is not the number of electric shocks a hungry squirrel will endure in order to find a nut, but the fact that this number of shocks reflects motivation.

Another way of saying this is that the number of syllables remembered is an *operational definition* of learning. Actually, learning *per se* is not something that we can observe directly, it is a *theoretical* concept. However, its effect is expressed in the number of syllables remembered (i.e. the greater the amount of learning the larger the number of remembered syllables). The big problem is when the dependent variable does not actually measure what the researcher wants to measure. A dependent variable which measures what it should measure is said to be *valid*. For example, GSR (see page 3) is sometimes used as a dependent variable (cf. Underwood and Saughnessy, 1983), but what it measures *precisely* is not clear at all. This is illustrated, for example, by the wide controversy about its use in ‘lie-detectors’.

A variation on that theme happens when several different dependent variables are supposedly measuring the same thing but do not give the same result. For example, suppose we want to measure *memory* for a list of words using different dependent variables. The first one is the number of words recognized, the second is the number of trials needed to re-learn the list two days later, the third one is the number of words recalled, and the fourth one is priming (i.e. how much faster one reads the word the second time, compared to the first time it was presented). All these dependent variables give the impression of measuring the same thing. In some cases, however, the conclusions reached using one dependent variable will differ from the conclusions using another dependent variable. That is to say some experimental manipulations will show an *improvement* in memory for a given dependent variable and a *deterioration*

⁵ It is not, by the way!

⁶ All these examples are real. Indeed, if you look at some psychological journals, you’ll find some dependent variables even more exotic than these.

for another dependent variable (cf. Bransford, 1979; Baddeley, 1994; Searleman and Herrman, 1994, and others). This is the case, for example, when memory is measured by an explicit test (e.g. recall or recognition) or an implicit test (e.g. priming, speed of reading). Paying attention to the material to be learned improves the explicit measurements but not the implicit ones.

The technical term for this effect is a *dissociation* (i.e. the independent variable creates a *dissociation* between the dependent variables). If it is possible to find two experimental conditions having an inverse effect on dependent variables such that, say, the *first* condition improves performance for the first dependent variable and decreases performance for the second dependent variable, whereas the *second* experimental condition decreases performance for the first dependent variable and improves performance for the second dependent variable, then we will say that there is a *double dissociation* of the dependent variables. We will see, later in the chapter dealing with two-factor designs, that a double dissociation involves some kind of an *interaction* (do not panic—if you do not know the meaning of this term on the first reading of this text, it will be clear from the second reading on). When we observe dissociations between dependent variables that were supposed to measure the same thing, this is often a sign that the ‘thing’ in question was not a simple entity as was first thought. For the example of memory, dissociations can be interpreted as revealing that there are several *components* or *modules* in memory and that the independent variables act differentially on the components.

Actually, a dependent variable is built or even created by psychologists to reveal the effects they are looking at. This construction can be quite sophisticated. For example, the dependent variable could be a reaction time (measured in milliseconds), but also a score obtained by asking subjects to rate on a scale from 1 to 5 how confident they feel that they have seen a sentence previously and by assigning a negative value to the score when the subject made a mistake (cf. Bransford *et al.*, 1972). In many cases, quantifying the dependent variable will imply quite a number of more or less arbitrary decisions. For example, if we want to measure the number of errors in the recollection of a story two pages long, what are we going to call *an error*? If the original word is *comprehend*, is *understand* a mistake? (This is up to you, as a researcher, to decide, depending upon the goal of the experiment).

1.4.1 Good dependent variables

The first principle guiding the construction of a dependent variable refers to its essential function: to reflect the effect of what we want to study. A good dependent variable should be *relevant* to the problem at hand. That is, a good dependent variable should be *valid*.

A good dependent variable should be *sensitive*. It will reflect clearly the effect of the independent variable. For example, to evaluate the difficulty of a problem, to be able to solve it or not will be a measure less sensitive than the time taken to solve it, even though both measurements will reflect how hard the task was.

For obvious practical reasons, dependent variables easy to measure will be favored, as well as dependent variables easy to define such that other researchers will be able to replicate your work. This can be straightforward in some cases (e.g. reaction time, number of syllables), a bit less so in some cases (e.g. number of errors in the recollection of a story), and rather challenging in some others (e.g. degree of introversion in a child’s play). In brief, a good dependent variable should be easy to *operationalize*.

Finally, an honest dependent variable will be reliable, which means the *error* of measurement will be small. The problem of *reliability* for a dependent variable corresponds

to the problem of internal validity for an independent variable. A reliable dependent variable increases the internal validity (can you see why?).

1.5 Choice of subjects and representative design of experiments

Psychological experiments make use of subjects. In general, it is not the subjects themselves that interest us, but rather what we learn concerning the population of individuals that they represent. A large number of statistics manuals insists on the necessity, in order to accomplish this, of working with representative samples of subjects drawn from the population of interest. And the means of achieving this is to draw *random* samples of subjects. Unfortunately, it is all but impossible to find an experiment in psychology (or in any of the other behavioral sciences) that can really claim to have used a *truly random* sample—one that is truly representative in the statistical sense of the population of interest: the human race. Exception might be made in the case of political polls, where pollsters might come close to obtaining a representative sample of voters. Apart from that, such studies—useful though they might be—do not really fall in the range of scientific research. Should we then give way to despair?

Quick—an example: We have carried out an experiment to show that the administration of barbiturates increases reaction time on a series of sensori-motor and intellectual tests. We have used forty students as subjects. We divided the subjects randomly into two groups: the experimental group (who received the barbiturate) and the control group (who received a placebo). (Note in passing the difference between ‘dividing the subjects randomly into two groups’ and ‘selecting a random sample of subjects from a certain population’.) To return to our experiment, we observe a clear difference in performance between the two groups—in fact a highly significant difference. The essential problem is to know whether we can have confidence in this difference. Will we find it for a population of Americans, of French, of Africans, of Eskimos, etc.? In this example, the sample is representative only of the population of students at the time of the experiment. Nevertheless we probably will be willing to generalize beyond the limits of the population of students, because we believe that the phenomena involved in our experiment are identical in the other populations aimed at. (Note that this decision depends on our knowledge of the phenomena and the comparability on relevant dimensions of the other populations). In fact, our experiment was designed to reveal essentially a *difference in behavior* of the subjects arising from differences in experimental conditions.

Apart from that, a large number of experiments aim at testing theoretical predictions, and will therefore be less sensitive to problems of the representativeness of the sample, as long as the assignment of subjects to groups is random. But the problem of ecological validity remains. It is also necessary to be aware of the sensitivity of classificatory variables to these problems. The warnings of McNemar (1949) remain valid: one part of the results of psychology only holds for American college students, and another part only holds for laboratory rats (we should note, following Neisser, 1982, that in response to this criticism several researchers stopped using rats).

A major problem linked to representativeness comes from *self-selection* of volunteers. In most studies subjects must be willing to participate in the experiment. Smart (1966) and Conrad and Maul (1981) estimate that two-thirds of the experiments reported in the

psychology journals use student volunteers. We find here again the *leitmotif* of confounded independent variables. The fact of volunteering risks being accompanied by other important characteristics. Rosenthal and Rosnow (1976) provide a description of the typical volunteer. The volunteer (in contrast to the non-volunteer) has:

- a higher education level;
- a greater need for approval;
- a higher score on intelligence tests;
- higher levels of authoritarianism (in the sense of Adorno *et al.*, 1950);
- a greater likelihood of showing signs of mental instability (in the sense of Lasagna and Von Felsinger, 1954).

Apart from that, volunteers tend, in contrast to non-volunteers, to produce results that agree with the research hypotheses espoused by the experimenters in the studies in which they participate (for a review, see Cowles and Davis, 1987).

Consequently, studies employing these independent variables have the potential of being biased, and should be examined very carefully.

Chapter summary

1.6 Key notions of the chapter

Below are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

Independent variable	Randomization
Dependent variable	Experimental error
Experimental factor	'Steam roller' method
Manipulated independent variable	Sensitivity, validity and reliability of dependent variables
Confounded independent variable	Operational definition
Classificatory or tag independent variable	Converging operations
Quasi-experimental study	Internal and external (ecological) validity

2

Correlation

2.1 Introduction

Correlation is often used as a descriptive tool in non-experimental research. We say that two measures are *correlated* if they have something in common. The *intensity* of the correlation is expressed by a number called the *coefficient of correlation* which is almost always denoted by the letter r . Although usually called the Pearson coefficient of correlation, it was first introduced by Galton (in a famous paper published in 1886) and later formalized by Karl Pearson (1896) and then by Fisher (1935). In this chapter we explain this coefficient, its rationale, and computation.

The main idea behind the coefficient of correlation is to compute an index which reflects how much two series of measurements are related to each other. For convenience, this coefficient will take values from -1 to $+1$ (inclusive). A value of 0 indicates that the two series of measurements have nothing in common. A value of $+1$ says that the two series of measurements are measuring the same thing (e.g. we have measured the height of a group of persons with metric and imperial¹ units). A value of -1 says that the two measurements are measuring the same thing but one measurement varies inversely to the other (e.g. one variable measures how rich you are, and the other one measures how poor you are: so the less rich you are, the poorer you are. Both scales are measuring the same financial status but with ‘opposite’ points of view).

2.2 Correlation: overview and example

The coefficient of correlation is a tool used to evaluate the similarity of two sets of measurements (i.e. two dependent variables) obtained on the same observations.² The coefficient of correlation indicates how much information is shared by two variables, or in other words, how much these two variables have in common.

For example, suppose that we take a (random) sample of $S = 20$ words from a dictionary and that, for each word, we count: (1) its number of letters and (2) the number of lines used

¹ (for U.S. readers) i.e. ‘British’ but they do not that anymore!

² The technical term is in fact ‘basic unit of measurement’, but here we will reserve the term ‘unit of measurement’ to indicate the unit in which an observation is measured.

Word	Length	Number of Lines
bag	3	14
across	6	7
on	2	11
insane	6	9
by	2	9
monastery	9	4
relief	6	8
slope	5	11
scoundrel	9	5
with	4	8
neither	7	2
pretentious	11	4
solid	5	12
this	4	9
for	3	8
therefore	9	1
generality	10	4
arise	5	13
blot	4	15
infectious	10	6
Σ	120	160
M	6	8

Table 2.1 Length (i.e. number of letters) and number of lines of the definition of a supposedly random sample of 20 words taken from the *Oxford English Dictionary*.

to define it in the dictionary. Looking at the relationship between these two quantities will show that, on the average, shorter words tend to have more meanings (i.e. ‘longer entries’) than longer words. In this example, the measurements or dependent variables that we compare are, on the one hand, the *length* (number of letters) and the *number of lines of the definition* on the other hand. The observations are the words that we measure. Table 2.1 gives the results of this survey.

What we would like to do is to express in a *quantitative* way the relationship between length and number of lines of the definition of the words. In order to do so, we want to compute an index that will summarize this relationship, and this is what the coefficient of correlation does.

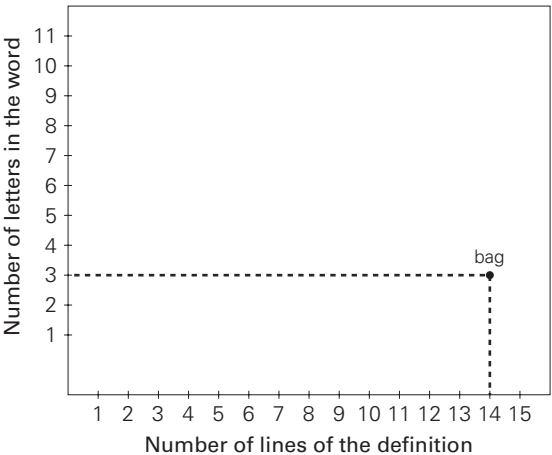


Figure 2.1 Plot of the word ‘bag’ which has 3 letters and 14 lines for its definition.

Let us go back to our example with data coming from a sample of 20 words taken (supposedly³) randomly from the *Oxford English Dictionary*.

A rapid perusal of Table 2.1 gives the impression that longer words tend, indeed, to have shorter definitions than shorter words (e.g. compare ‘by’ with ‘therefore’). In order to have a clearer representation of the data, the first step is to plot them in a *scatterplot*. To draw a scatterplot, we decide arbitrarily to use one of the dependent variables as the vertical axis (here the ‘word length’ or number of letters of the word) and the other dependent variable as the horizontal axis (here the ‘number of lines of the definition’). Each word is represented as a point whose coordinates correspond to its number of letters and number of lines. For example, the word ‘bag’ with 3 letters and 14 lines is represented as the point (14, 3) as illustrated in Figure 2.1.

The whole set of words is represented in Figure 2.2. The labels of the points in the graph can be omitted in order to make the graph more readable (see Figure 2.3). Looking at Figure 2.3 confirms our intuition that shorter words tend to have more meanings than longer words. The purpose of the coefficient of correlation is to quantify precisely this intuition.

2.3 Rationale and computation of the coefficient of correlation

Because, on the average, shorter words have longer definitions than longer words, the shape of the set of points⁴ displayed in Figure 2.3 is roughly elliptical, oriented from the upper left to the lower right corner.

If the relationship between length of words and number of lines of their definition were perfect, then all the points would be positioned on a line sloping downwards, as shown in

³ The truth is that we helped chance to get nice numbers and a beautiful story.

⁴ Statisticians call the set of points the ‘cloud’ of points.

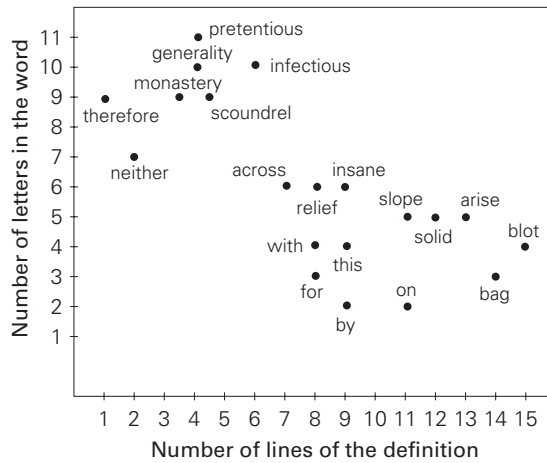


Figure 2.2 Plot of the words as points with vertical coordinates being the length of the words (letters) and with horizontal coordinates representing the number of lines of the definition of the words. Each point is labeled according to the word it represents.

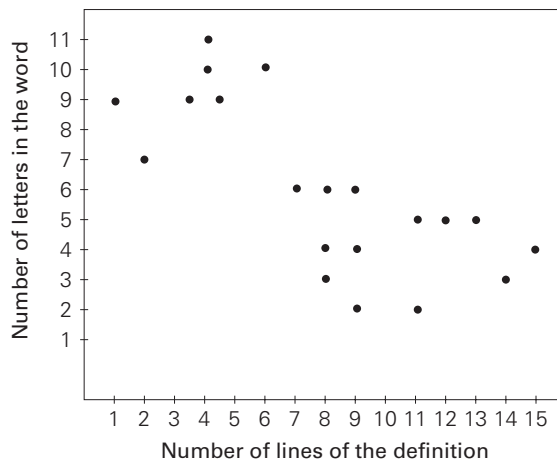


Figure 2.3 Plot of the data from Table 2.1 without labels.

Figure 2.4. This perfect relationship would give rise to a coefficient of correlation of $r = -1$, and we will call such a relationship a perfect *negative* correlation.

In our case, even though the trend is quite clear, the points are not strictly aligned, and so, the relationship between length of words and number of meanings is not perfect. The problem is to know how far from perfect this relationship is. Or, in other words, how far (or how close, depending upon your point of view) from the line (represented in Figure 2.4) are the data points representing the words.

2.3.1 Centering

The first step in computing the coefficient of correlation is to transform the data. Instead of using the raw data, we will use the *distance* or *deviation* from the means on the

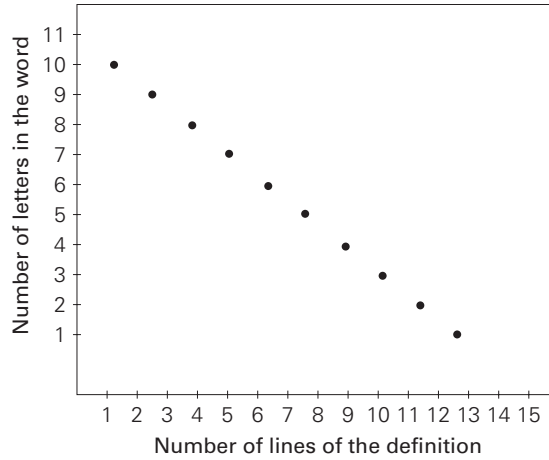


Figure 2.4 A perfect (negative) linear relationship.

two dimensions. Formally, instead of representing the observations as points whose coordinates are the length and the number of lines, we subtract from each length measurement the mean length and we subtract from each number of lines the mean number of lines. So if we denote by Y the dependent variable ‘length’ and by W the dependent variable⁵ ‘number of lines’, each word will be represented by two numbers

$$W - M_W \quad \text{and} \quad Y - M_Y .$$

This approach is, actually, equivalent to moving the axes so that the origin of the graph is now placed at the average point⁶ (i.e. the point with coordinates $M_W = 8$ and $M_Y = 6$). The graph, now, is said to be *centered*. Figures 2.5 and 2.6 show the effect of this centering.

2.3.2 The four quadrants

The position of the axes in Figure 2.5 defines four quadrants. These quadrants are numbered from 1 to 4 as shown in Figure 2.7. The data points falling into Quadrant 1 are above the average length but below the average number of lines. The data points falling into Quadrant 2 are above both the average length and the average number of lines. The data points falling into Quadrant 3 are below the average length but above the average number of lines. And, finally, the data points falling into Quadrant 4 are below both the average length and the average number of lines. Counting the number of data points falling in each quadrant shows that most points fall into Quadrants 1 and 3. Specifically, seven data points fall into Quadrant 1, zero in

⁵ You may be wondering why we do not use X for the horizontal axis the way it is often done. We use W in this text because X is *always* used for denoting an independent variable. The length of the words is not an independent variable but a dependent variable hence the use of W . The distinction between these notations will be clearer in Chapter 4 on regression.

⁶ Technically, this point is called the ‘center of gravity’ or ‘barycenter’, or ‘centroid’ of the cloud of points. This is because if each point were represented by a weight, the average point would coincide with the center of gravity.

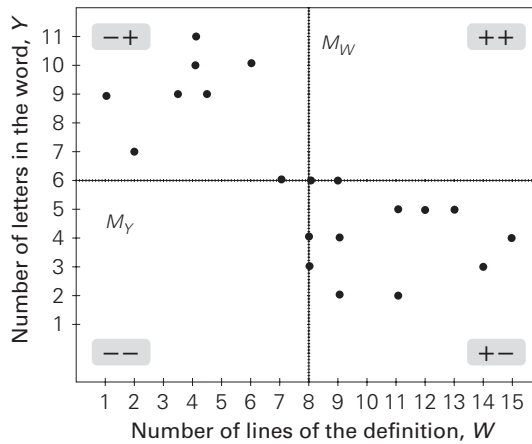


Figure 2.5 Centered plot of the data from Table 2.1. The data points (words) are now represented as deviations from the means.

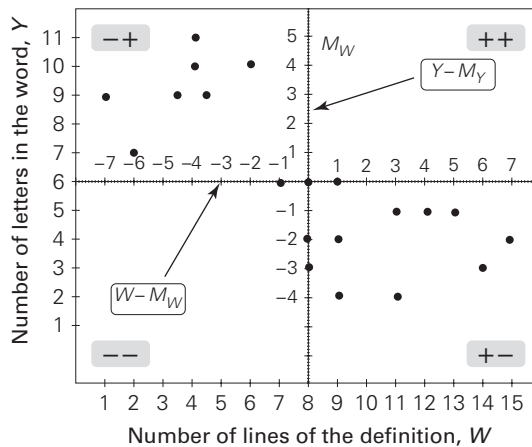


Figure 2.6 The data points in Figure 2.5 have now the values $W - M_W$ for horizontal coordinates and $Y - M_Y$ for vertical coordinates.

Quadrant 2, eight in Quadrant 3, and zero in Quadrant 4.⁷ This indicates that for fifteen words (seven in Quadrant 1, plus eight in Quadrant 3) we have an inverse relationship between length and number of lines: small values of W are associated with large values of Y (Quadrant 1) and large values of W are associated with small values of Y (Quadrant 3). The main idea, now, is to summarize this information by associating *one* number to the location of each of the data points and then combining (i.e. summing) all these numbers into one single index.

As we just saw, a first idea for assessing the relationship between two variables is to try to count the number of points falling in each quadrant. If larger values of W are associated with smaller values of Y (as is the case with the length and the number of lines of the words) and smaller values of W are associated with larger values of Y , then most observations will fall in

⁷ And five data points fall on the borderline between two or more quadrants. These points do not provide information about the correlation between W and Y .

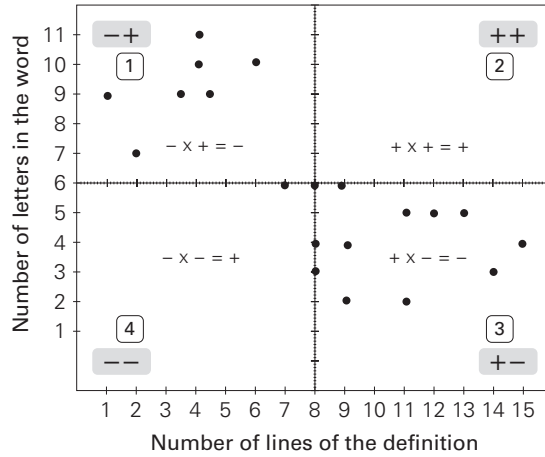


Figure 2.7 The four quadrants defined by centering the plot as in Figure 2.5. Each quadrant displays the sign of the product of the coordinates (in terms of deviation from the mean) of the data points that fall in this quadrant. For example, the observations falling in Quadrant 1 will have a value of $(W - M_W) \times (Y - M_Y)$ negative because $(W - M_W)$ is negative and $(Y - M_Y)$ is positive, and hence the sign of the product will be given by the rule ‘minus times plus equal minus’, or $- \times + = -$.

Quadrants 1 and 3. If, on the contrary, larger values of W are associated with larger values of Y and smaller values of W are associated with smaller values of Y , then most observations will fall in Quadrants 2 and 4. Finally, if there is no relationship between W and Y , the data points will be—roughly—evenly distributed in all the quadrants (this approach—of counting the points in each quadrant—can indeed be used, and would give a non-parametric test called the ‘corner test’).

The problem, however, with this approach is that it does not use all the information available in the scatterplot. In particular, it gives the same importance to each observation (cf. observations *a* and *b* in Figure 2.8), whereas extreme observations are more indicative of a relationship than observations close to the center of the scatterplot.

2.3.3 The rectangles and their sum

In order to give to each data point an importance that reflects its position in the scatterplot we will use two facts. First, the coordinates on a given axis of all the data points in the same quadrant have the same sign; and second, the product of the coordinates of a point gives an *area* (i.e. the rectangle in Figure 2.8) that reflects the importance of this point for the correlation. This area associated with a point will be positive or negative depending upon the position of the point.

Before going further it may be a good idea to remember the rule: ‘plus times plus is plus’, ‘minus times minus is plus’, ‘minus times plus is minus’, and ‘plus times minus is minus’.

With this rule in mind, we can determine the sign for each quadrant. For example, in Quadrant 1, all the W coordinates are negative because they correspond to words having a number of lines smaller than the average number of lines; and all Y coordinates are positive because they correspond to words having a length greater than the average length. And, because ‘minus times plus equals minus’ the product of the coordinates of the points

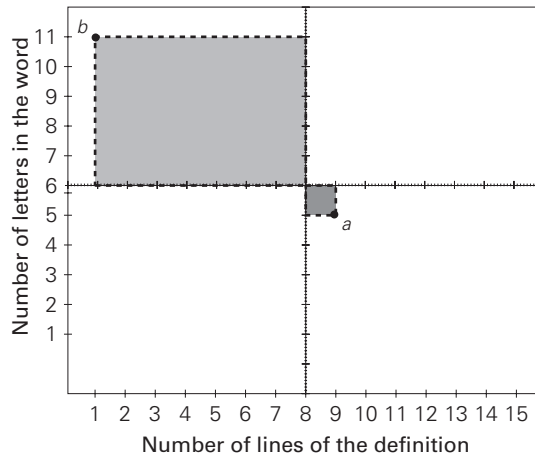


Figure 2.8 Two observations that should *not* be given the same importance when evaluating the relationship between dependent variables W and Y . Observation a is closer to the mean of the distribution than observation b . Observation b , which is more extreme than a , should be given a greater importance in assessing the relationship between W and Y . This is done by the coefficient of correlation. Note that the importance of an observation can be evaluated by the *area* (i.e. the 'rectangle') obtained by multiplying the deviations to the means.

falling in Quadrant 1 will be negative (i.e. it is obtained by multiplying a negative quantity (the horizontal axis: $W - M_W$) with a positive quantity (the vertical axis: $Y - M_Y$). We call the term

$$(W_s - M_W) \times (Y_s - M_Y)$$

the *cross-product* of the s th observation (or point or word, etc.). It is also called the '*rectangle*' term for the s th observation (it is a rectangle because the result of the multiplication of these two numbers gives the value of the 'area' of the rectangle whose sides correspond to these two numbers). The area corresponding to the cross-product for a point reflects its eccentricity: The further away from the center a point is, the larger its cross-product. So the cross-product for a point shows how much a point provides evidence for the correlation between the two variables of interest.

To summarize the previous paragraph, all the observations falling in Quadrant 1 have a cross-product with the same direction (or sign): *negative*, but with a different magnitude. Observations with extreme coordinates have a larger cross-product than those falling close to the center of the plot. With the same reasoning, all the observations falling in Quadrant 3 have a *negative* cross-product whereas all the observations falling in Quadrants 2 or 4 have a *positive* cross-product.

2.3.4 Sum of the cross-products

In order to give more importance to extreme observations than to central observations, we have associated to each observation its cross-product. In order to integrate all these cross-products

into one index, we simply compute their sum. And we call it the *sum of the cross-products*⁸ (very often abbreviated simply as the ‘cross-products’ or SCP_{WY} or simply SCP when there is no ambiguity). With a formula, we get:

$$SCP = \sum_s (W_s - M_W)(Y_s - M_Y).$$

Because extreme observations will have larger cross-products than central ones, they will contribute a more important part to the sum of all the cross-products.

In our example, with $M_W = 8$, and $M_Y = 6$, the cross-product is computed as

$$\begin{aligned} SCP_{WY} &= \sum_s (W_s - M_W)(Y_s - M_Y) \\ &= (14 - 8)(3 - 6) + (7 - 8)(6 - 6) + \dots \\ &\quad \dots + (15 - 8)(4 - 6) + (6 - 8)(10 - 6) \\ &= (6 \times -3) + (-1 \times 0) + \dots + (7 \times -2) + (-2 \times 4) \\ &= -18 + 0 + \dots - 14 - 8 \\ &= -154. \end{aligned}$$

The column labelled $w \times y$ in Table 2.2 gives the cross-product corresponding to each word and so the sum of this column gives the SCP (i.e. the sum of the cross-products). A positive value for the sum of the cross-products indicates a positive relationship between variables, and a negative value for the sum of the cross-products indicates a negative relationship. Note that, strictly speaking, the value of the SCP should be expressed with the measurements used. Therefore, we should say that the value of the SCP is -154 ‘number of letters of the word per number of lines of the definition of the word’ (e.g. the same way that we talk about ‘miles per hour’).

2.3.5 Covariance

The big problem, now, is to interpret the magnitude of the SCP . What does a value of -154 mean? Two things make the interpretation difficult. The first one is the *number of observations*: the larger the number of observations, the larger the value of the SCP . The second one is the *unit of measurement*: How could we compare the result of a study reporting a value of -154 ‘number of letters of the word per number of lines of the definition of the word’ with another one reporting a value of 87 ‘strokes per size of city’ (i.e. people living in large cities tend to suffer more strokes than people living in small cities). Is the relationship between variables stronger in the first study or in the second?

⁸ How original!

Word	Length Y	Number of Lines W	y	w	w × y	y ²	w ²
bag	13	14	−3	−6	-18	29	36
across	16	17	−0	−1	-20	20	21
on	12	11	−4	−3	-12	16	19
insane	16	19	−0	−1	-20	20	21
by	12	19	−4	−1	-4	16	11
monastery	19	14	−3	−4	-12	29	16
relief	16	18	−0	−0	-20	20	20
solid	15	12	−1	−4	-4	21	16
this	14	19	−2	−1	-2	24	21
for	13	18	−3	−0	-10	29	10
therefore	19	11	−3	−7	-21	29	49
generality	10	14	−4	−4	-16	16	16
arise	15	13	−1	−5	-5	21	25
blot	14	15	−2	−7	-14	24	49
infectious	10	16	−4	−2	-8	16	24
Σ	120	160	0	0	-154	150	294
					SCP	SS _Y	SS _W

Table 2.2 Raw scores, deviations from the mean, cross-products and sums of squares for the example length of words and number of lines. $M_W = 8$, $M_Y = 6$. The following abbreviations are used to label the columns: $w = (W - M_W)$; $y = (Y - M_Y)$; $w \times y = (W - M_W) \times (Y - M_Y)$; SS stands for sum of squares (see Appendix A, page 417).

In order to take into account the effect of the number of observations, the solution is to divide the *SCP* by the number of observations. This defines a new statistic called the *covariance*⁹ of W and Y . It is abbreviated as cov_{WY} . With a formula:

$$cov_{WY} = \frac{SCP}{\text{Number of Observations}} = \frac{SCP}{S}.$$

For our example, the covariance equals:

$$cov_{WY} = \frac{SCP}{S} = \frac{-154}{20} = -7.70.$$

⁹ We divide by the number of observations to compute the covariance of the set of observations. To estimate the covariance of the population from the sample we divide by $S - 1$. This is analogous to the distinction between σ and $\hat{\sigma}$ presented in Appendix A.

2.3.6 Correlation: the rectangles and the squares

The covariance does, indeed, take into account the problem of the number of observations, but it is still expressed in the original unit of measurements. In order to eliminate the original unit of measurement, the idea is to normalize the covariance by dividing it by the standard deviation of each variable.¹⁰ This defines the coefficient of correlation denoted $r_{W \cdot Y}$ (read ‘ r of W and Y ’, or ‘ r of W dot Y ’). The coefficient of correlation is also abbreviated as r when there is no ambiguity about the name of the variables involved. With a formula, the coefficient of correlation is defined as

$$r_{W \cdot Y} = \frac{COV_{WY}}{\sigma_W \sigma_Y} . \quad (2.1)$$

By rewriting the previous formula, a more practical formula for the coefficient of correlation is given by

$$r_{W \cdot Y} = \frac{SCP_{WY}}{\sqrt{SS_W SS_Y}} , \quad (2.2)$$

where SS stands for sum of squares (see Appendix A, page 417 for details; if you feel that you do not understand how we go from formula 2.1 to formula 2.2, this is explained in the next section). As a mnemonic, if we remember that the cross-products are rectangles, this formula says that the coefficient of correlation is the *rectangles divided by the square-root of the squares*.

This is the formula that we will use in general. For our example (cf. Table 2.2 for these quantities), we find the following value for r :

$$r_{W \cdot Y} = \frac{SCP_{WY}}{\sqrt{SS_W SS_Y}} = \frac{-154}{\sqrt{294 \times 150}} = \frac{-154}{\sqrt{44,200}} = \frac{-154}{210} = -.7333 . \quad (2.3)$$

This value indicates a negative linear relationship between the length of words and their number of meanings.

2.3.6.1 For experts: going from one formula to the other one

How do we transform Equation 2.1 into Equation 2.2? Basically we substitute and simplify as follows:

$$\begin{aligned} r_{W \cdot Y} &= \frac{COV_{WY}}{\sigma_W \sigma_Y} \\ &= \frac{SCP_{WY} / S}{\sqrt{\frac{SS_W}{S}} \sqrt{\frac{SS_Y}{S}}} \\ &= \frac{SCP_{WY}}{S \sqrt{\frac{SS_W}{S}} \sqrt{\frac{SS_Y}{S}}} \\ &= \frac{SCP_{WY}}{S \sqrt{\frac{SS_W SS_Y}{S^2}}} \end{aligned}$$

¹⁰ Remember that the standard deviation is expressed in the same unit of measurement as the variable it describes. For example, the standard deviation of W is expressed as number of lines, like W .

$$\begin{aligned}
&= \frac{SCP_{WY}}{S \times \frac{1}{S} \sqrt{SS_W SS_Y}} \\
&= \frac{SCP_{WY}}{\sqrt{SS_W SS_Y}}. \tag{2.4}
\end{aligned}$$

2.3.7 Some properties of the coefficient of correlation

When we compute the coefficient of correlation (see Equations 2.1 and 2.2), we divide the units of the numerator by the same units in the denominator, this process eliminates these units and therefore the coefficient of correlation is a number *without unit*. Hence, the coefficient of correlation can be used to compare different studies performed with different variables measured with different units.

Another very interesting property of the coefficient of correlation is that its maximum magnitude is very convenient and easy to remember. Precisely, a coefficient of correlation is restricted to the range of values between -1 and $+1$. The closer to -1 or $+1$ the coefficient of correlation is, the stronger the relationship. For example, $r = -.7333$ indicates a stronger negative relationship than, let's say $r = -.52$.

The magnitude of the coefficient of correlation is always smaller or equal to 1. This happens because the numerator of the coefficient of correlation (see Equation 2.2) is always smaller than or equal to its denominator. This property is detailed in the next section which can be skipped by less advanced readers.

2.3.8 For experts: why the correlation takes values between -1 and $+1$

The magnitude of the coefficient of correlation is always smaller or equal to 1. This is a consequence of a property known as the *Schwartz inequality*.¹¹ If we take two sets of numbers (call them X_s and T_s), this inequality can be expressed as:

$$\left| \sum X_s T_s \right| \leq \left(\sum |X_s| \right) \times \left(\sum |T_s| \right)$$

(where the vertical bars $|$ mean 'absolute value' or 'magnitude'). The Schwartz inequality shows that for the coefficient of correlation the following inequality always holds (we just need to square each term of the inequality to get the result):

$$\left[\sum (W - M_W)(Y - M_Y) \right]^2 \leq \left[\sum (W - M_W)^2 \right] \times \left[\sum (Y - M_Y)^2 \right],$$

which implies (as we said above) that $r_{W,Y}$ can take values between -1 and $+1$ only (because the numerator of r is always less than or equal to its denominator).

We could show (cf. section 2.7, page 32) that the value of -1 or $+1$ is obtained when the points corresponding to the (W_s, Y_s) observations lie on a straight line. In other words, the value of -1 or $+1$ is obtained when the shapes of the W and Y distributions are identical (cf. Section 2.6, page 30).

¹¹ The proof can be found in most calculus textbooks.

2.4 Interpreting correlation and scatterplots

We have seen that the coefficient of correlation varies between the values $+1$ and -1 . When it reaches these extreme values, we say that the dependent variables are *perfectly correlated*. In this case, the dependent variables are essentially measuring the same thing.

A *positive* value of the coefficient of correlation is said to reflect a *positive* linear¹² relationship between the dependent variables: those observations or individuals who score high on one variable tend to score high on the other and vice versa. A *negative* value of the coefficient of correlation is said to reflect a *negative* linear relationship between the dependent variables: those observations or individuals who score high on one variable tend to score low on the other. When the coefficient of correlation is null, the dependent variables are said to be *uncorrelated*.

2.5 The importance of scatterplots

Even if the coefficient of correlation gives a number that reflects as best as possible the relationship between variables, it can be misleading sometimes. We look here at two such problematic cases: first when the relationship between the variables is non-linear and second when the presence of outliers distorts the value of the correlation. In both cases, looking at the scatterplot is enough to detect the problem.

2.5.1 Linear and non-linear relationship

The coefficient of correlation measures the *linear* relationship between two variables. This means that it evaluates how close to a *line* the scatterplot is. An error—easy to make—is to think that the coefficient of correlation is evaluating *any* type of relationship between the two variables. This is not the case as shown in Figure 2.10 which displays an example of a perfect non-linear relationship between two variables (i.e. the data points show a *U*-shaped relationship with Y being proportional to the square of W). But the coefficient of correlation is equal to zero. Obviously, in such cases the coefficient of correlation does not give a good indication of the intensity of the relationship between the variables. In some cases, the non-linearity problem can be handled by transforming one or both variables (here, for example we can take the square root of Y instead of Y , or, alternatively, we could square W).

2.5.2 *Vive la différence?* The danger of outliers

As we have shown in Figure 2.8, observations that are far from the center of the distribution contribute a lot to the sum of the cross-products (this happens because these observations have large values for the deviations to the means; and when these values are multiplied we get a very large value for the cross-product). In fact, one extremely deviant observation (often called an ‘outlier’) can substantially change the value of r . An example is given in Figure 2.11.

¹² The relationship is termed linear because when plotted one against the other the values tend to fall on a straight line (cf. Figures 2.4 and 2.9).

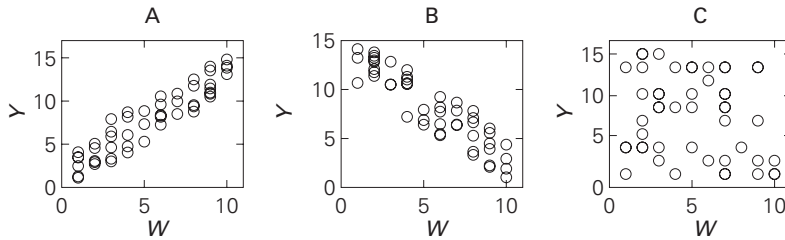


Figure 2.9 Examples of relationship between two variables: (A) positive linear relationship $0 < r \leq 1$; (B) negative linear relationship $-1 \leq r < 0$; and (C) no linear relationship $r = 0$.

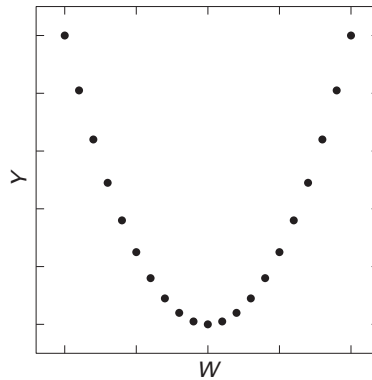


Figure 2.10 A perfect non-linear (e.g. *U*-shaped) relationship with a coefficient of correlation equal to zero.

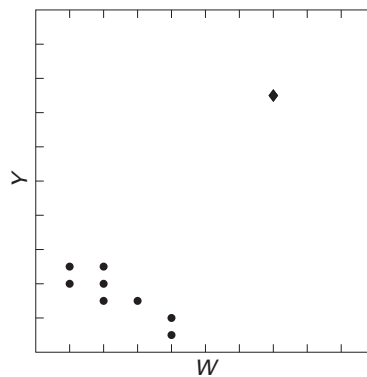


Figure 2.11 The dangerous effect of outliers on the value of the coefficient of correlation. The correlation of the set of points represented by the circles is equal to -0.87 , when the point represented by the diamond is added to the set, the correlation is now equal to $+0.61$. This example shows that one outlier can dramatically influence the value of the coefficient of correlation.

When such an extreme value is present in the data, we need to understand why this happened: Is it a typographical error, is it another error of some sort (such as an equipment failure), a data recording error, or is it an observation coming from a different population, etc.?

It is always a good idea to examine closely such outliers in order to decide what to do with them. When the outlier is the consequence of some error, it is a good idea to fix the problem and replace the erroneous value with the correct one. Other cases may require more

sophistication; in fact dealing with outliers is an important problem which is unfortunately too large to be detailed here (but see Barnett and Lewis 1994, if you want to explore this topic).

2.6 Correlation and similarity of distributions

The coefficient of correlation is also a way of quantifying the similarity between two distributions or between two shapes. Consider the picture drawn in Figure 2.12. The distributions on the left of the picture have the same shape. The units on the horizontal axis are the same for both distributions. Therefore, for each unit on the horizontal axis we have two values: one for W , and one for Y . As a consequence, we can consider that each horizontal unit is a point with coordinates W and Y . And we can plot each of these points in a scatterplot. This is illustrated in the right part of Figure 2.12 which shows that the points representing the horizontal units are perfectly aligned when plotted with W and Y coordinates.

2.6.1 The other side of the mirror: negative correlation

When the coefficient of correlation is negative, this means that the two variables vary in opposite direction. Plotting them, as illustrated in Figure 2.13 shows that one variable displays the *inverse* shape of the other. Because our eye is better at detecting similar shapes rather than opposite shapes, when the correlation is negative, it is convenient to change the direction of one of the variables before looking at the shapes of the two variables. This change

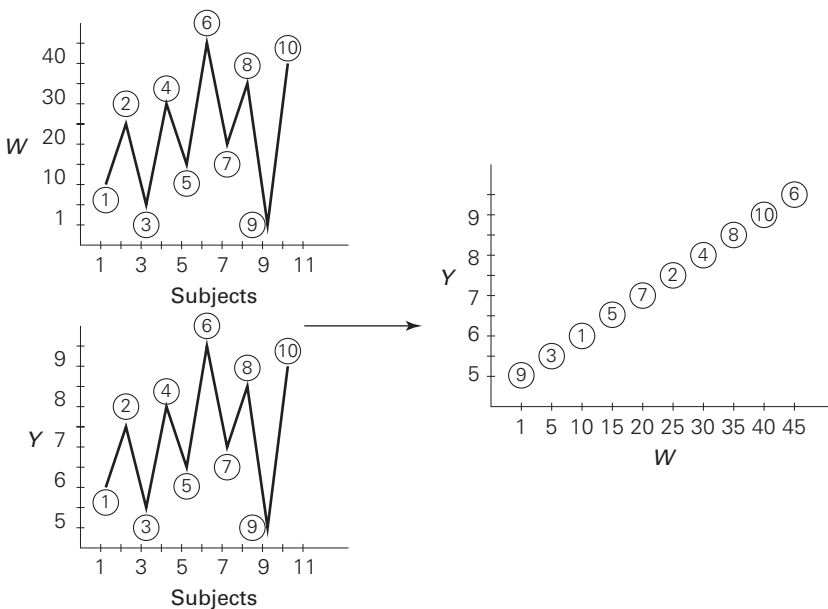


Figure 2.12 The points describing two identical shapes (when adjusted for scaling differences) will lie on a straight line when plotted as couples of points (W , Y). Note that the coordinates of the points in the left panels are given by the line, not by the circle (e.g. Subject 1 has a W coordinate of '10', and '6' for the Y coordinate, Subject 2 has '25' for the W coordinate, and '7.5' for the Y coordinate).

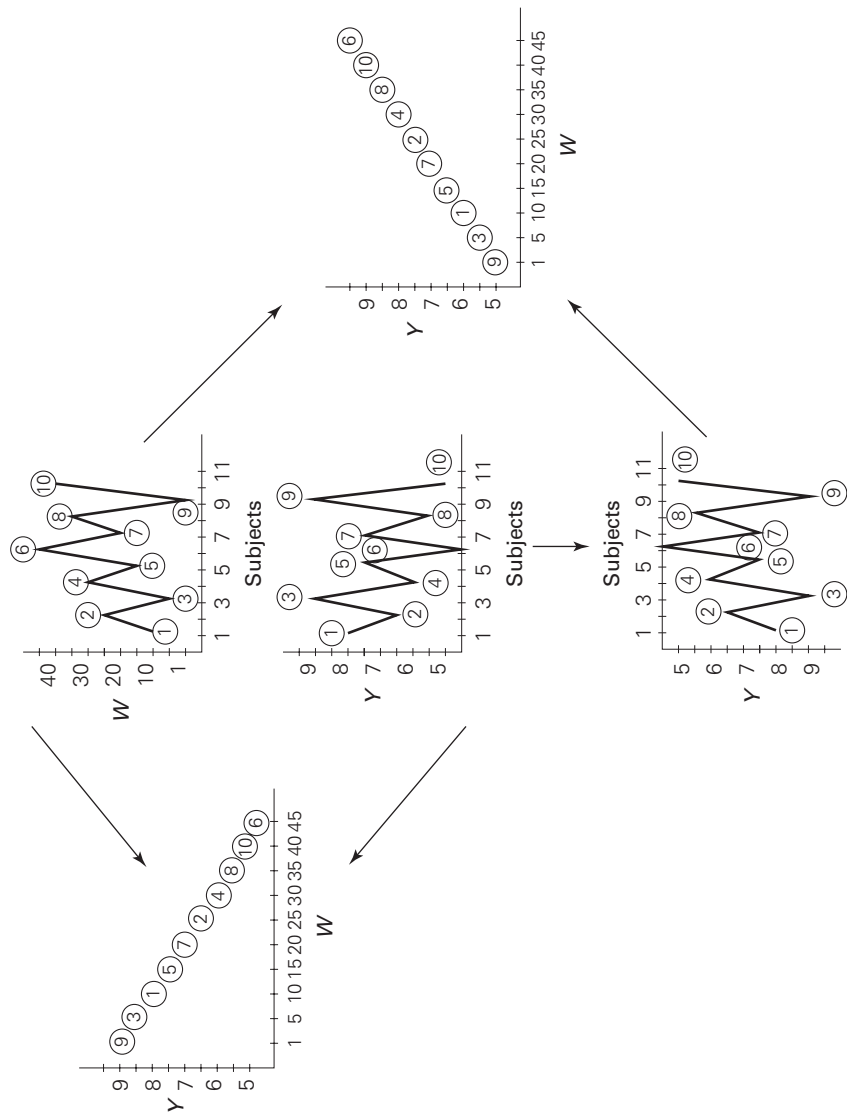


Figure 2.13 The points describing two *opposite* shapes (when adjusted for scaling differences) will lie on a *negative* line when plotted as couples of points (W , Y). To look at the similarity between the shapes of the distributions, is it convenient to use the mirror image of one of the variables (here we have 'flipped' the Y variable). After this transformation, the relationship between the two variables is positive and the shapes of the distributions show similar trends.

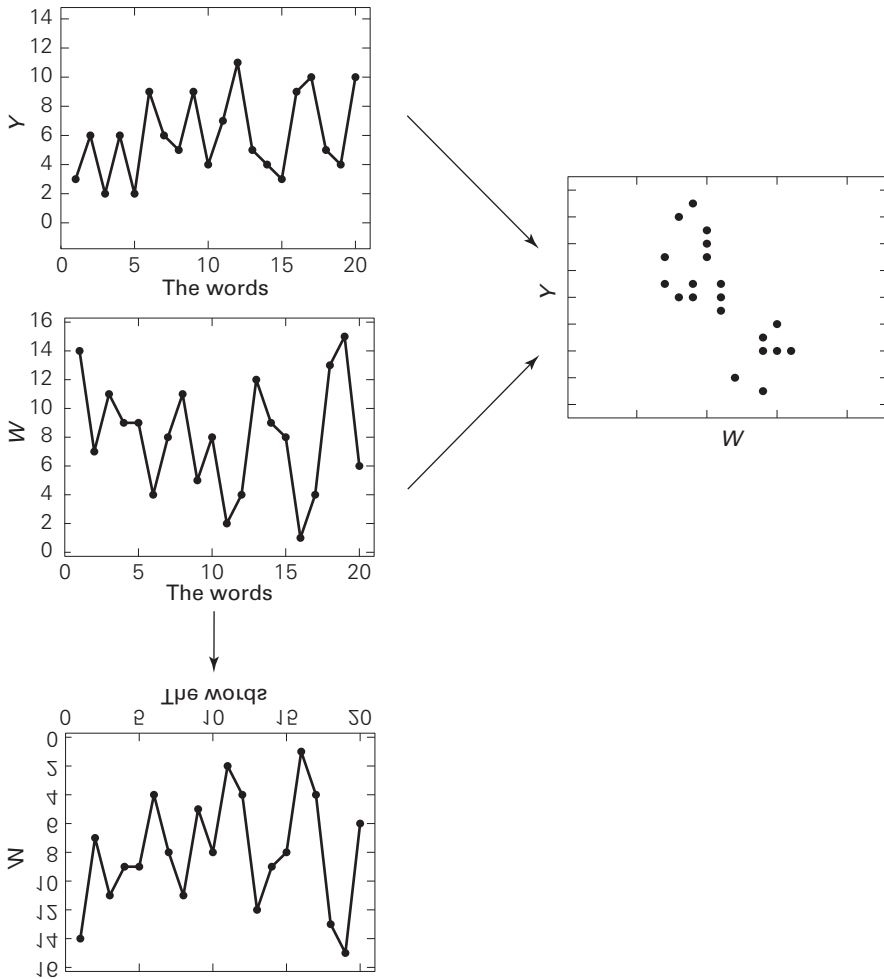


Figure 2.14 Polygons and scatterplot for the example length and number of lines of the definition of words. In the bottom graph, the W variable has been transformed by reversing in a mirror. Now we can more easily compare the shape of the Y distribution and the W distribution. This shows that the shapes of these two distributions are somewhat similar but not identical. This similarity is indicated by a value of the coefficient of correlation of $r = -.7333$.

can be performed in several ways; here we just turned Y upside down. This procedure is illustrated in Figure 2.13. After this transformation has been performed, the relationship between variables is positive and it is easier to compare the shapes of the variables.

This procedure can also be used with the words data. Here each of the words is assigned an arbitrary position on the horizontal axis. The plot of the distribution for the length of the words and the number of meanings is displayed in Figure 2.14.

2.7 Correlation and Z-scores

The coefficient of correlation can be seen as a way of normalizing the cross-product (by eliminating the measurement units). Doing so makes possible the comparison of data gathered

with different units. One way of comparing data measured with different units was to transform these data into Z-scores (see Appendix A for a refresher on Z-scores). Recall that a variable, say Y , is transformed into a Z-score by centering (i.e. subtracting the mean) and normalizing (i.e. dividing by the standard deviation). With a formula, the s th observation of variable Y is transformed into a Z-score as¹³

$$Z_{Y_s} = \frac{Y_s - M_Y}{\sigma_Y} .$$

We need a subscript here because we will use Z-scores for Y and W .

When variables have been transformed into Z-scores, the formula for the correlation becomes conceptually much simpler. However, this is not, in general, a very practical way of computing the coefficient of correlation¹⁴ even if it helps us to understand several of its properties. Specifically, if Z_{Y_s} denotes the Z-score for the s th observation of the variable Y , and Z_{W_s} denotes the Z-score for the s th observation of the variable W , then the coefficient of correlation can be obtained from the Z-scores as

$$r_{W \cdot Y} = \frac{1}{S} \times \sum Z_{W_s} Z_{Y_s} . \quad (2.5)$$

This can be shown, quite simply, by developing Equation 2.5:

$$\begin{aligned} r_{W \cdot Y} &= \frac{1}{S} \times \sum Z_{W_s} Z_{Y_s} = \frac{1}{S} \sum \left(\frac{W_s - M_W}{\sigma_W} \right) \left(\frac{Y_s - M_Y}{\sigma_Y} \right) \\ &= \sum \frac{(W_s - M_W)(Y_s - M_Y)}{S \times \sigma_W \times \sigma_Y} . \end{aligned}$$

$$\text{But, } S \times \sigma_W \times \sigma_Y = \sqrt{\sum (W_s - M_W)^2 \sum (Y_s - M_Y)^2} ,$$

and therefore we get,

$$\begin{aligned} \frac{1}{S} \times \sum Z_{W_s} Z_{Y_s} &= \frac{\sum (W_s - M_W)(Y_s - M_Y)}{\sqrt{\sum (W_s - M_W)^2 \sum (Y_s - M_Y)^2}} \\ &= \frac{SCP_{WY}}{\sqrt{SS_W SS_Y}} \\ &= r_{W \cdot Y} . \end{aligned}$$

¹³ In what follows, we use in the formula for Z-score the population standard deviation σ . We could have used the sample standard deviation ($\hat{\sigma}$) as well. In this case, however, we need to substitute the value $(S - 1)$ for S in the formulas of this section.

¹⁴ So don't use it for computing! Using Z-scores requires more computations and is likely to generate rounding errors.

2.7.1 Computing with Z-scores: an example

To illustrate the computation of the coefficient of correlation using Z-scores, we are going to use again the data from the example 'length of words and number of lines'. The first step is, indeed, to compute the Z-score. For this we need the mean and standard deviation of each variable. From Table 2.2, we find the following values:

- For the means,

$$M_W = \frac{160}{20} = 8.00$$

$$M_Y = \frac{120}{20} = 6.00$$

- For the variances,

$$\sigma_W^2 = \frac{294}{20} = 14.70$$

$$\sigma_Y^2 = \frac{150}{20} = 7.50$$

- For the standard deviations,

$$\sigma_W = \sqrt{\sigma_W^2} = \sqrt{14.70} \approx 3.83$$

$$\sigma_Y = \sqrt{\sigma_Y^2} = \sqrt{7.50} \approx 2.74 .$$

With these values, we can fill in Table 2.3 which gives, in turn, the quantities needed to compute the coefficient of correlation using Z-scores. You can check that

$$\begin{aligned} r_{W \cdot Y} &= \frac{1}{S} \times \sum Z_{W_s} Z_{Y_s} \\ &= \frac{1}{S} \times (-1.7143 + 0 + \dots - 1.3333 - 0.7619) \\ &= \frac{-14.6667}{20} \\ &= -.7333 , \end{aligned}$$

which, indeed, is the same value as was found before.

2.7.2 Z-scores and perfect correlation

Another advantage of expressing the coefficient of correlation in terms of Z-scores is that it makes clear the fact that the values of +1 and -1 correspond to a perfect correlation between two variables. Two variables are perfectly correlated when they vary in the exact same way, or in other words, when the shapes of their distributions are identical. In that case the Z-scores corresponding to the first variable, W , are equal to the Z-scores corresponding to the second variable Y :

$$Z_{W_s} = Z_{Y_s} \quad \text{for all } s .$$

Hence, the formula for r becomes

$$r_{W \cdot Y} = \frac{1}{S} \times \sum Z_{W_s} Z_{W_s} = \frac{1}{S} \times \sum Z_{W_s}^2 . \quad (2.6)$$

Word	Length Y	Number of Lines W	Z_{Y_s}	Z_{W_s}	$Z_{Y_s} \times Z_{W_s}$
bag	3	14	-1.0954	1.5649	-1.7143
across	6	7	0	-0.2608	0
on	2	11	-1.4606	0.7825	-1.1429
insane	6	9	0	0.2608	0
by	2	9	-1.4606	0.2608	-0.3810
monastery	9	4	1.0954	-1.0433	-1.1429
relief	6	8	0	0	0
slope	5	11	-0.3651	0.7825	-0.2857
scoundrel	9	5	1.0954	-0.7825	-0.8571
with	4	8	-0.7303	0	0
neither	7	2	0.3651	-1.5649	-0.5714
pretentious	11	4	1.8257	-1.0433	-1.9048
solid	5	12	-0.3651	1.0433	-0.3810
this	4	9	-0.7303	0.2608	-0.1905
for	3	8	-1.0954	0	0
therefore	9	1	1.0954	-1.8257	-2.0000
generality	10	4	1.4606	-1.0433	-1.5238
arise	5	13	-0.3651	1.3041	-0.4762
blot	4	15	-0.7303	1.8257	-1.3333
infectious	10	6	1.4606	-0.5216	-0.7619
Σ	120	160	0	0	-14.6667

Table 2.3 Raw scores, Z-scores and Z-score cross-products for the example: 'Length of words and number of lines of the definition'. $M_W = 8$, $M_Y = 6$, $\sigma_W \approx 3.83$, and $\sigma_Y \approx 2.74$.

Remember that Z-scores have a standard deviation equal to 1, and therefore a variance also equal to 1 (because $1^2 = 1$). From the formula of the variance of a set of scores, we find that for any set of Z-scores we have the following equality (remember that the mean of the Z-scores is zero):

$$\sigma_Z^2 = 1 = \frac{\sum_s (Z_s - M_Z)^2}{S} = \frac{\sum_s (Z_s - 0)^2}{S} = \frac{\sum_s Z_s^2}{S}.$$

This last equation implies that for any set of Z-scores

$$\sum Z_s^2 = S. \quad (2.7)$$

Now if we plug the result from Equation 2.7 into Equation 2.6, we find that when the shapes of two dependent variables are the same, the value of their coefficient of correlation becomes

$$r_{W \cdot Y} = \frac{1}{S} \times \sum Z_{W_s}^2 = \frac{1}{S} \times S = 1,$$

which shows that when the variables Y and W are perfectly correlated, the magnitude coefficient of correlation will be equal to 1.

2.8 Correlation and causality

An important point to keep in mind when dealing with correlation is that *correlation does not imply causality*. The fact that two variables are correlated does not mean that one variable causes the other. There are, in fact, quite a lot of examples showing the consequences of mixing correlation and causality. Several of them can be somewhat silly or funny (it is almost a professional hazard for statisticians to look for these cases). For example, the intellectual development of children as measured by the IQ or developmental quotient, is highly and significantly correlated with the size of their big toes—the higher the intellectual ability, the larger the big toe. Does that mean that intelligence is located in the big toe?

Another example: in France, the number of Catholic churches, as well as the number of schools, in a city is highly correlated with the incidence of cirrhosis of the liver (and the number of alcoholics), the number of teenage pregnancies, and the number of violent deaths. Does that mean that (French) churches and schools are sources of vice? Does that mean that (French) newborns are so prone to violence as to be murderers? Actually, it is more reasonable to realize that the larger a city is, the larger the number of churches, schools and alcoholics, and so on, is going to be. In this example, the correlation between number of churches/schools and alcoholics is called a *spurious* correlation because it reflects only their mutual correlation with a third variable (here the size of the city).

However, the existence of a correlation between two dependent variables can be used as a practical way to predict¹⁵ the values of a dependent variable from another one. For example, Armor (1972, see also Pedhazur, 1982) after a re-analysis of a very important educational survey called the Coleman report (1966), found that a composite index—derived from the possession of several domestic appliances (e.g. TV sets, telephone, refrigerator) and some other items (such as an encyclopedia)—shows a correlation of around .80 with a score of verbal intelligence. As a consequence, this index can be used to predict the verbal score of children (if it is impossible or impractical to measure it directly, of course). But it would be silly to infer from this correlation that a fridge *causes* verbal intelligence (maybe fridges are talkative around children?), or that buying appliances will increase the verbal intelligence of our children (but what about the encyclopedia?).

Correlations between several dependent variables are in general studied using techniques known as factor analysis, principal component analysis, or structural equation modelling (see e.g. Loehlin, 1987; Pedhazur and Pedhazur-Schmelkin, 1991; Jolliffe, 2003). The main goal of these methods is to reveal or express the relations between the dependent variables in terms of hidden (independent) variables often called factors, components, or latent variables.

¹⁵ How to use the coefficient of correlation to predict one variable from the other one is explained in Chapter 4 on regression analysis.

It is worth noting, in passing, how much the naturalistic and the experimental approach differ in the way they consider individual differences. The naturalistic approach essentially relies upon individual differences in order to detect the effects of variables of interest (cf. Chateau, 1972). On the other hand, the experimental approach treats individual differences as a nuisance (i.e. error) and actively tries to eliminate or control them.

However, even though a correlational approach is founded on individual differences, it can be used as a first test (or a *crucible* to use the pretty word of Underwood, 1975) for theories. Actually a good number of theories tested in the laboratory are also amenable to testing with a correlational approach (pre-tested, Underwood would say). For example, if a theory predicts that imagery affects memory, then subjects spontaneously using imagery (or ‘highly visual subjects’) should exhibit a better level of performance on a memory task than subjects who do not use imagery. As a consequence, we should be able to find a correlation between the dependent variable ‘spontaneous use of imagery’ and the dependent variable ‘memory performance’ (both are dependent variables because we are just measuring them, not controlling them). If the correlation between imagery and memory is close to zero, then the theory should be abandoned or amended. If the correlation is high, the theory is not proven (remember the slogan: ‘correlation is not causation’). It is at best supported, but we feel more confident in moving to an experimental approach in order to test the causal aspect of the theory. So a correlational approach can be used as a first preliminary step.

2.9 Squared correlation as common variance

When squared, the coefficient of correlation can be interpreted as the *proportion of common variance* between two variables. The reasoning behind this interpretation will be made clear after we have mastered regression analysis, so at this point we will just mention this property. For example, the correlation between the variables length (W) and number of meanings (Y) is equal to $r_{W.Y} = -0.7333$ (cf. Equation 2.3), therefore the proportion of common variance between W and Y is equal to $r_{W.Y}^2 = (-0.7333)^2 = .5378$. Equivalently, we could say that W and Y have 54% of their variance in common. Because a coefficient of correlation takes values between -1 and $+1$, the squared coefficient of correlation will always take values between 0 and 1 and its magnitude will always be smaller than the magnitude of the coefficient of correlation.

Chapter summary

2.10 Key notions of the chapter

Below are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

Linear relationship
Scatterplot

Cross-product
Covariance

Pearson coefficient of correlation
Perfect correlation

Correlation and causality

2.11 Key formulas of the chapter

Below are the main formulas introduced in this chapter: try to go through them and understand what they mean.

$$SCP_{WY} = \sum_s (W_s - M_W)(Y_s - M_Y)$$

$$COV_{WY} = \frac{SCP}{S}$$

$$r_{W \cdot Y} = \frac{COV_{WY}}{\sigma_W \sigma_Y}$$

$$= \frac{SCP_{WY}}{\sqrt{SS_W SS_Y}}$$

$$= \frac{1}{S} \times \sum Z_{W_s} Z_{Y_s}$$

2.12 Key questions of the chapter

Below are some questions about the content of this chapter. All the answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- * What types of information can you derive from a scatterplot ?
- * What is the goal of the coefficient of correlation?
- * In which case would you use a correlation analysis?
- * What can we conclude from a positive value of the coefficient of correlation and why?
- * What can we conclude from a negative value of the coefficient of correlation and why?
- * What can we conclude from a zero value of the coefficient of correlation and why?
- * Why do we use the coefficient of correlation rather than the cross-product or the covariance to assess the relationship between two variables?
- * What are the limitations of the coefficient of correlation?
- * There is a strong correlation between the number of fire trucks sent to fires and the amount of damage done by those fires. Therefore, to minimize fire damage we should send fewer fire trucks to fires. Are you convinced?

3

Statistical test: the F test

3.1 Introduction

In the previous chapter, we have seen that a linear relationship between two variables can be quantified by the Pearson coefficient of correlation (r) between these two variables. This coefficient indicates the *direction* of the relationship (positive or negative) and its *magnitude*. The closer the value of r to $+1$ or -1 , the clearer the indication of a linear relationship between these two variables.

The problem we would like to solve now is to decide whether this relationship really exists, or if it could be attributed to a fluke. So the question is: for a given value of a coefficient of correlation, can we *infer* that it indicates a real relationship between variables, or can such a value be obtained by chance? This question corresponds to what is called a *null hypothesis* statistical test. The term 'null hypothesis' is a statistical way of saying 'can be due to chance'. If we decide that the coefficient of correlation reflects a real relationship between variables, we say that its value cannot be due to chance, and the corresponding statistical expression is 'to reject the null hypothesis'. From a formal point of view, when we perform a statistical test we consider that the observations constitute a *sample* from some *population* of interest. The goal of the statistical test is to *infer* from the value of the correlation computed on the sample if the correlation in the population is equal to zero or not. In order to do so, we compute a statistic called F (after *Fisher*, the statistician who first studied this problem). The F is built to make a decision about the effect of the independent variable. It is called a criterion (from the Greek *κριτεριον* meaning 'a means of judgment' or a 'decision'). This F ratio is made of two components. The first one expresses the effect of the correlation, the second one depends on the number of observations. Here is the formula for the F ratio.

$$F = \frac{r^2}{1 - r^2} \times (S - 2). \quad (3.1)$$

The term $(S - 2)$ is called the number of degrees of freedom of error.¹ We explain the procedure of such a statistical test in the first part of this chapter.

¹ The reason for that will be clear in the chapters on regression (Chapter 4, page 75) and analysis of variance (Chapter 7, page 140).

So, a statistical test tells us if the correlation in the population is equal to zero or not. This is often not enough. We may want to know how *large* the correlation is likely to be in the population. This is a problem of estimating the value of the correlation of the population from the sample. In this context we also often want to compute a confidence interval around the estimation. We explain this procedure in the second part of this chapter.

3.2 Statistical test

In order to illustrate the notion of a statistical test, suppose we have a sample of 6 observations for which we have measured the values of two dependent variables (W and Y). We want to know, from this sample, if we can infer that there exists a real correlation in the population from which this sample was extracted. The data are given in Table 3.1.

As an exercise, you can check that we find the following values:

- The means of W and Y are

$$M_W = 4 \quad \text{and} \quad M_Y = 10.$$

- The sum of the cross-product of the deviations is

$$SCP_{YW} = -20.$$

- The sum of squares of W and Y are

$$SS_W = 20 \quad \text{and} \quad SS_Y = 80.$$

- Using the sum of cross-products and the sums of squares we can compute the coefficient of correlation between W and Y as

$$\begin{aligned} r_{W,Y} &= \frac{SCP_{YW}}{\sqrt{SS_Y \times SS_W}} \\ &= \frac{-20}{\sqrt{80 \times 20}} = \frac{-20}{\sqrt{1600}} = -\frac{20}{40} \\ &= -.5. \end{aligned} \tag{3.2}$$

This value of $r = -.5$ indicates a negative linear relationship between W and Y . Can we get such a value by chance alone or does this value show that W and Y are really correlated in the population?

3.2.1 The null hypothesis and the alternative hypothesis

Statistically, the goal is to *estimate* the values for the whole population (called the *parameters*) from the results obtained from a sample (called the *statistics*). Specifically, we want to

$W_1 = 1$	$W_2 = 3$	$W_3 = 4$	$W_4 = 4$	$W_5 = 5$	$W_6 = 7$
$Y_1 = 16$	$Y_2 = 104$	$Y_3 = 12$	$Y_4 = 4$	$Y_5 = 8$	$Y_6 = 10$

Table 3.1 Six observations with their values for the dependent variables W and Y .

determine if there is any correlation between the two variables for the whole population on the basis of the correlation calculated on a sample. We say that we want to *infer* the values of the parameters for the population on the basis of the statistics calculated on the sample. In a statistical test for correlation, we simply want to know if $r = 0$ or not in the population of interest. The decision problem involved here comes down to two alternatives:

1. The value of r computed on the sample is a fluke.
2. The value of r computed on the sample reflects a linear relationship between the two variables in the population.

In other words, we must choose between two hypotheses describing the population from which our sample of observations is drawn, in order to ‘explain’ the observed value of r :

- The first is called the *null hypothesis*, abbreviated H_0 : there is no linear relationship between the variables for the population of observations as a whole (i.e. the correlation is null, or inexistent² in the population of observations, hence the term null hypothesis). Note that the null hypothesis gives a precise value for the intensity of the effect—the *effect is absent or null*: its intensity is zero. We say that the null hypothesis is an *exact (or precise) statistical hypothesis*.
- The second is called the *alternative hypothesis*, abbreviated H_1 : there is a linear relationship between the two variables for the population of observations as a whole. (Note that the intensity of the correlation is not specified.) The observed value of r reflects that effect. This amounts to saying that the correlation is ‘not null’ (that is, it is greater or less than zero). We emphasize: H_1 *does not provide* a precise value for the intensity of the correlation (since there is an infinity of values greater or less than zero). We say that H_1 is an *inexact (or imprecise) statistical hypothesis*. If H_1 specified the expected value of the intensity of the correlation *before* the experiment, then H_1 would be an exact hypothesis. (That is, H_1 is not by nature inexact.) But that eventuality is rarely encountered in the domain of psychology (not to say never). Because of this, except when the contrary is specifically indicated, H_1 is here taken to be an inexact statistical hypothesis.

3.2.2 A decision rule: reject H_0 when it is ‘unlikely’

An ideal decision strategy would be to calculate the probability of observing the obtained value of r as a function of *each* statistical hypothesis (H_0 and H_1), and then adopt the hypothesis giving the best prediction—that is, to adopt the most *likely* hypothesis. However, we can only use this procedure when both H_0 and H_1 are *exact* statistical hypotheses.

Instead of this ideal strategy (impossible to carry out because H_1 is inexact), in order to show that there is a correlation between W and Y , we try to show that it is very unlikely to have obtained the results that we did if there had been no correlation in the population. More formally we apply the decision rule of rejecting the null hypothesis if it appears *unlikely*. So the first step is to evaluate the probability associated (see Appendix D, pages 451ff.) of obtaining

² Fisher (1966, p. 13) refers to the null hypothesis as ‘the hypothesis that the phenomenon to be demonstrated is in fact absent’.

the results we have obtained if the null hypothesis is true. We denote this probability by p . If p is small enough then we will consider that the null hypothesis is unlikely. Because the null hypothesis and the alternative hypothesis are contradictory, rejecting the null hypothesis implies accepting the alternative hypothesis (of the existence of a linear relationship between the two variables in the population). On the other hand, if we fail to reject the null hypothesis, we cannot really determine anything, and so we say that we suspend judgment or that our results are inconclusive. So:

If the null hypothesis is unlikely, reject it, and accept the H_1 .

Otherwise:

Fail to reject the null hypothesis and suspend judgment.

3.2.3 The significance level: specifying the ‘improbable’

But at what point can we estimate that an event is improbable? When it has less than one chance in five of happening? Less than one chance in ten? In twenty? In a hundred or a million? We need to choose a specific value for what we call improbable, and we call this value the *significance level*. This level is symbolized by the lower-case Greek letter α (alpha). For example, suppose that for us, unlikely means ‘less than five chances out of one hundred’, we will say that the α level is $\alpha = .05$. More technically, in order to accept the alternative hypothesis (H_1) we need to show that the results we have obtained are less likely than $\alpha = .05$ to occur if the null hypothesis (H_0) were true.

Traditionally two levels of significance have been in use. The first, $\alpha = .05$, indicates that the improbable event will occur by chance less than five times out of a hundred. The second, $\alpha = .01$, corresponds to one chance in a hundred.

Why do we use these two levels? Most of it is due to tradition. When the statistical test procedure was established (in the early part of the 20th century), computational facilities were rather primitive and a lot of computations needed to be done to validate a test for a given α level. This pushed to keep the number of levels to a minimum. So, the choice of $\alpha = .05$ and $\alpha = .01$ is in part an accident of history. Sir Ronald Fisher—who more than anyone is responsible for the modern statistical practices—happened to favor these two levels, and tradition plus habit conspired to make his choice a *de facto* standard. But these two levels stayed also because they correspond to obvious numbers in our base 10 counting system ($\alpha = .05$ is 20, and $\alpha = .01$ is 100). As our base 10 system is likely to derive from our 10 fingers, maybe, an additional reason for these standard levels is our physiology. In any case, it is worth keeping in mind that these standard values of .05 and .01 are somewhat arbitrary and subjective.

3.2.4 Type I and Type II errors: α and β

In fact, setting the level of significance implies taking into account the (subjective) weight attached to the consequences of two possible types of errors.

- On the one hand, there is the error of accepting the existence of a linear relationship between the variables in the population when that relationship does not really exist. We call this error a *Type I error* or a *false alarm*.

- On the other hand, there is the error of failing to detect the existence of a linear relationship between the variables when it really does exist. We call this error a *Type II error* or *miss*).

So we are in a situation where we cannot make no error. And therefore the problem becomes: Which type of error should be avoided most? Or what type of error can be acceptable? We can minimize one type of error, but this will in general imply that we will accept making the other type of error: There is always a trade-off. For example, in order to minimize Type I errors, we can set a strict significance level. So, with $\alpha = .00001$ we will commit fewer Type I errors than with $\alpha = .20$. Conversely, to avoid Type II errors we should set a looser significance level: With $\alpha = .20$ we shall make fewer Type II errors than with $\alpha = .00001$.

As you can see, these errors (i.e. Type I and II) vary as an inverse function of the α level. Therefore, the choice of α represents a compromise between these two types of errors. If you wish above all to avoid believing wrongly in the existence of the effect of a linear relationship between the variables in the population, then you should above all minimize Type I errors and choose a strict α level (.01, .001, or even smaller). In contrast, if you wish above all to detect an actual effect, then you should above all minimize Type II errors and choose a large α level (.05, .10, .20, or, why not, .50?).

Notice the strict relationship between the significance level and Type I errors. *The significance level corresponds to the probability of committing a Type I error.* In effect, with a fixed α level, we will erroneously reject the null hypothesis each time the probability associated with the statistical index F (i.e. the index used to determine the likelihood that a correlation exists in the population) is less than or equal to α , which comes down to saying that α gives the probability of committing a Type I error.

Just as α denotes the probability of committing a Type I error, β denotes the probability of committing a Type II error (β is the lower-case Greek beta). However, the alternative hypothesis being inexact, the exact value of β cannot be known precisely. As a consequence we can *never* reject the alternative hypothesis, nor can we *ever* accept the null hypothesis. Not rejecting the null hypothesis is not equivalent to accepting it. Recall, however, the inverse relationship of α and β : the smaller α , the larger β is, and *vice versa*.

However, several factors other than α influence β :

- First, there is *the intensity of the relationship between the variables*. When H_1 is inexact, this quantity remains unknown. We can nevertheless estimate the parameter *intensity of the relationship in the population* on the basis of the statistics calculated on the sample (more on that later on).
- Second, there is *the number of observations*. The larger the number of observations, the smaller β is. Remember that the statistical index F is made of two terms: one reflecting the correlation, and the other the number of observations.
- Third, there is *the intensity of the experimental error* (individual differences, measurement error, etc.). The smaller the experimental error, the smaller β is.

Thus, in order to minimize the probability of a Type II error of a statistical test, we can play with three factors. While we do not have much influence on the first one (intensity of the relationship), we can easily play with the second or the third. This is obvious for the case of the number of observations, and we can manipulate various aspects of the research methods to minimize measurement error.

Remember, however, that because the alternative hypothesis is inexact, the exact value of β remains unknown, and this it is impossible to evaluate the likelihood of the alternative hypothesis. So:

**We can *never* reject H_1 , nor can we *ever* accept H_0 .
Not rejecting H_0 is not equivalent to accepting it.**

Nevertheless, we can—in some cases—compute the *probable values* of the effect on the basis of the results of a measurement. In other words, how likely is it that these results will occur if the measurement is conducted repeatedly with different observations? We can then decide whether the possible intensity of the effect is *uninteresting* or *unimportant* (cf. Introduction).

In summary:

- α and β represent the probabilities of error. $1 - \alpha$ and $1 - \beta$ represent the probabilities of success. Thus:
- $1 - \alpha$ = probability of not rejecting H_0 when it is true. This is the probability of correctly deciding that there is no effect when in fact there is none.
- $1 - \beta$ = probability of rejecting H_0 when it is false. This is the probability of correctly detecting an effect when there is one. This probability is so important that we give it a specific name: We call the quantity $(1 - \beta)$, the *power* of the statistical test.

The more powerful the test, the more sensitive it is to the presence of an effect—the more easily it detects an effect. With all these probabilities [i.e. α , β , $(1 - \alpha)$, and $(1 - \beta)$], we can build a table summarizing the various possible outcomes of a statistical decision (cf. Table 3.2).

3.2.5 Sampling distributions: Fisher's F

Remember that in our previous numerical example we found a correlation $r = -.5$ on a sample of 6 observations. How can we evaluate the likelihood of obtaining the same correlation when the null hypothesis is true (i.e. when there is no correlation in the population)? There are several ways. The first one is to use an empirical approach (also called Monte-Carlo approach) and generate a large number of samples of 6 observations according to the model stated by the null hypothesis. The *frequency* of the values of r^2 can then be used to evaluate how likely is the value of $r^2 = .25$ when there is, in fact no relationship between W and Y . For historical

Experimental decision	State of nature	
	H_0 true (H_1 false)	H_0 false (H_1 true)
Do not reject H_0	Correct non-reject probability = $1 - \alpha$	Miss (Type II error) probability = β
Reject H_0	False alarm (Type I error) probability = α	Correct detection probability = $1 - \beta$

Table 3.2 Outcomes of a statistical decision.

reasons, however, we do not use the coefficient of correlation but a statistic closely related to it: the index F . Recall that F is obtained as:

$$F = \frac{r^2}{1 - r^2} \times (S - 2)$$

which gives, for our example

$$F = \frac{.25}{1 - .25} \times (6 - 2) = \frac{.25}{.75} \times 4 = \frac{1}{3} \times 4 = \frac{4}{3} = 1.33.$$

The problem now is to find out what is the probability of finding this value of F if the null hypothesis is true, or in other words, if there is no relationship between W and Y for the population as a whole.

3.2.5.1 Empirical (Monte-Carlo) approach

If the null hypothesis is true, then the values of W are obtained *independently* of the values of Y . This is equivalent to saying that the values of W are obtained randomly from one population, and that the values of Y are obtained randomly from another population. If we want to be able to simulate that model, in addition to supposing the null hypothesis true, we need also to specify the shape of these populations. The favorite shape for statisticians is the *normal* (or Gaussian, see Appendix D, 460ff.) curve, so we will assume that the Z scores for W and Y come from normal populations when the null hypothesis is true. This model is illustrated in Figure 3.1

To evaluate the likelihood of obtaining a value of $r^2 = .25$ from 6 observations when the null hypothesis is actually true, we start by generating 6 random observations corresponding to the null hypothesis model. In order to do so, we extract from the first hat (cf. Figure 3.2) 6 values representing the W values and we extract from the second hat 6 values representing the Y values. For one example, here are the values we have obtained:

$$\begin{array}{ll} W_1 = -0.3325 & Y_1 = -1.5715 \\ W_2 = -0.9912 & Y_2 = -0.8114 \\ W_3 = -0.6802 & Y_3 = -0.3204 \\ W_4 = -0.6625 & Y_4 = -0.2992 \\ W_5 = -0.3554 & Y_5 = -1.2197 \\ W_6 = -1.6968 & Y_6 = -0.4384 \end{array} \quad (3.3)$$

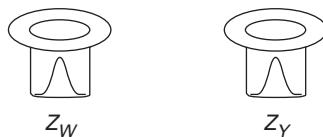


Figure 3.1 The model of score generation when the null hypothesis is true. Each pair of observations is obtained by taking randomly one score Z_W from a normal population (i.e. drawing it out of a hat) and one score Z_Y from another normal population. The populations are symbolized by hats.

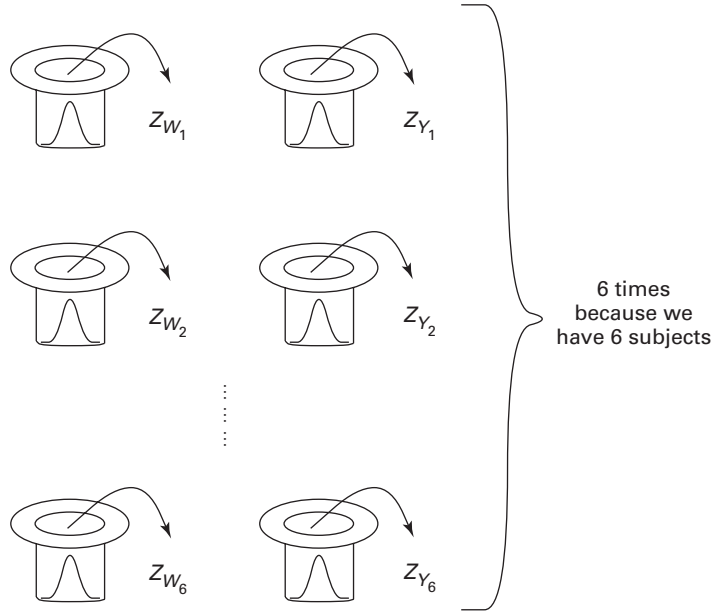


Figure 3.2 The model of score generation when the null hypothesis is true: 6 pairs of observations are obtained by taking one score Z_W from a normal population and one score Z_Y from another population. The populations are symbolized by hats.

The next step is to compute the values of r^2 and F for that sample. Here we obtain:

$$r_{W \cdot Y} = -.2086 \quad r_{W \cdot Y}^2 = .0435 \quad F = 0.1819$$

Suppose now that we repeat the procedure of computing r^2 and F for 1000 trials . Figure 3.3 shows the first 9 trials and their r values.

Figure 3.4 shows the histogram of the values of r^2 and F obtained for 1000 random trials (i.e. when the null hypothesis is true). The horizontal axes represent the different values of r^2 (top panel) and F (bottom panel) obtained for the 1000 trials and the vertical axis the number of occurrences of each value of r^2 and F . For example, you can read on the top panel that 160 samples (over the 1000 trials) had a value of $r^2 = .01$ which was between 0 and .01 (this corresponds to the first bar of the histogram in Figure 3.4).

As you can see in Figure 3.4 the number of occurrences of a given value of r^2 and F decreases as an inverse function of their magnitude: the greater the value, the less probable it is to obtain it when there is no correlation in the population (i.e. when the null hypothesis is true). However, we can see that the probability of obtaining a large value of r^2 or F is not null. In other words, even when the null hypothesis is true, we can obtain very large values of r^2 and F .

From now on, we will focus on the F distribution, but what we say could apply to the r^2 distribution as well. In order to decide to reject the null hypothesis we need to find the values corresponding to the 95th and 99th percentiles which correspond respectively to the significance levels of $\alpha = .05$ and $\alpha = .01$. These values are equal to 7.55 and 24.1, respectively. The probability associated to these values (i.e. the famous ' p ' value) is exactly equal to the probability of $\alpha = .05$ and $\alpha = .01$. Therefore these values correspond to probability thresholds, any F larger than these values has an associated probability smaller

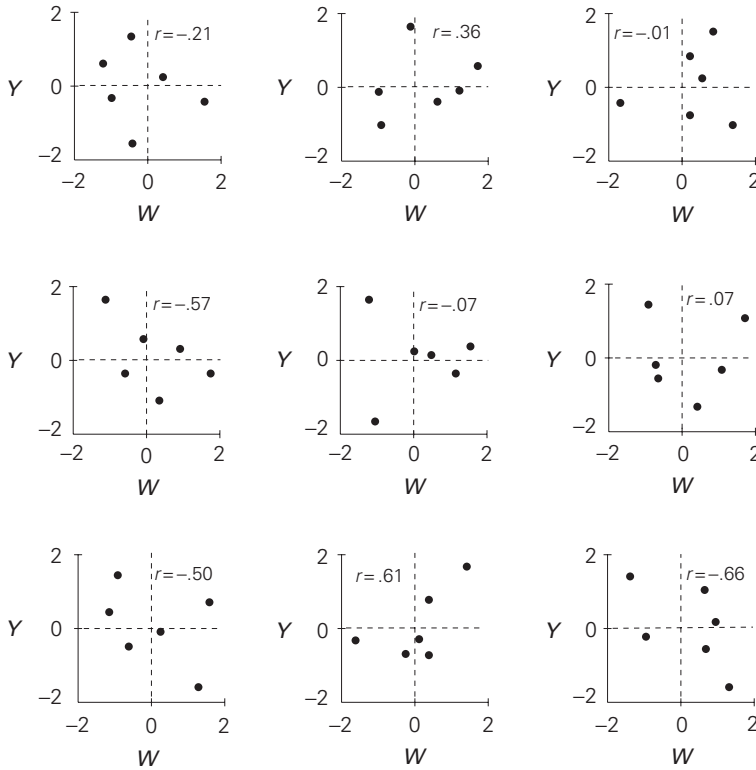


Figure 3.3 The first nine random samples and their correlation coefficient.

than the threshold. These special values of 7.55 and 24.1 are called the *critical values* of F for $\alpha = .05$ and $\alpha = .01$. Figure 3.5 shows the empirical distribution of F along with these critical values. Any value of F larger than the critical value corresponds to an event *less probable* (i.e. less likely) than the α level when the null hypothesis is true.

The rule when deciding between the two hypotheses H_0 and H_1 is to reject the null hypothesis when the F ratio is larger than the critical F ; in an equivalent manner, the null hypothesis is rejected when the probability associated with the F ratio is less than the α level chosen. Suppose that, for the previous numerical example, we have chosen α equal to .05; since at this significance level the computed F of 1.33 is smaller than the critical F of 7.55, we cannot reject the null hypothesis and we must suspend judgment.

3.2.5.2 Theoretical (traditional) approach

Another way of performing the F test is to try to derive mathematically the sampling distribution (see Appendix D) of the F ratio when the null hypothesis is true. This was done in the 1930s by Fisher who was elaborating on previous work done by Student.³ In order to do so, Fisher had to suppose that the populations from which the random samples are drawn are normal distributions (the same assumption that we have used to derive the

³ To know more about these creators of modern statistics see Vokey and Allen, 2005; and McCloskey and Ziliak, 2008.

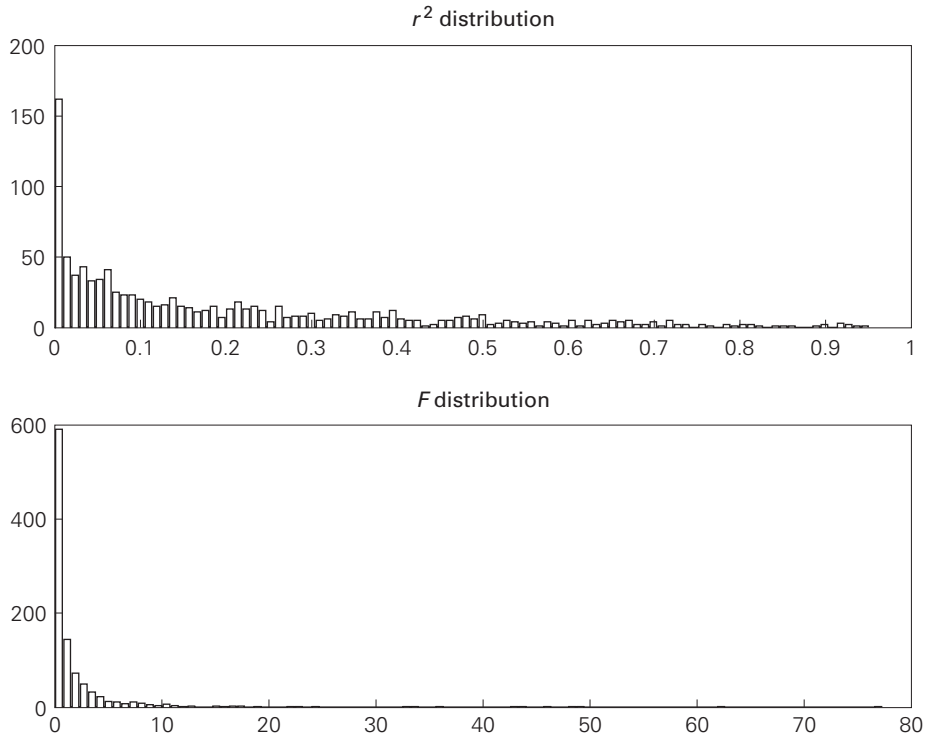


Figure 3.4 Histogram of values of r^2 and F computed from 1000 random samples when the null hypothesis is true. The histogram shows the empirical distribution of F and r^2 under the null hypothesis.

empirical distribution). Fisher found that there was not one distribution, but a whole family of distributions which depends upon the number of *independent* observations used to compute the F ratio.

We call these distributions Fisher's F distributions. They are derived from the normal distribution and are among the most important distributions in statistics, and are the most popular distributions in psychological papers (you will find that almost every paper in the American Psychological Association—a.k.a. APA—journals quotes them). We will use them repeatedly in this book. Figure 3.6 shows three different Fisher's distributions.

These distributions depend upon two parameters, ν_1 and ν_2 (pronounced 'nu-one,' and 'nu-two'), which are called the numbers of degrees of freedom⁴ (abbreviated as df) of the distribution. The parameter ν_1 corresponds to the number of degrees of freedom of correlation, and for testing a coefficient of correlation (like the ones we have seen so far), it is always equal to 1. The parameter ν_2 corresponds to the number of degrees of freedom of error, and is equal to $S - 2$. Thus, when we are testing the correlation between two variables, ν_1 is always equal to 1, and ν_2 is equal to $S - 2$. In our example, we find the following values:

$$\nu_1 = df_{\text{correlation}} = 1 \quad \nu_2 = df_{\text{error}} = S - 2 = 4.$$

⁴ The rationale for this name will be explained later on in Chapter 7 on analysis of variance, Section 7.3.6, pages 140ff.

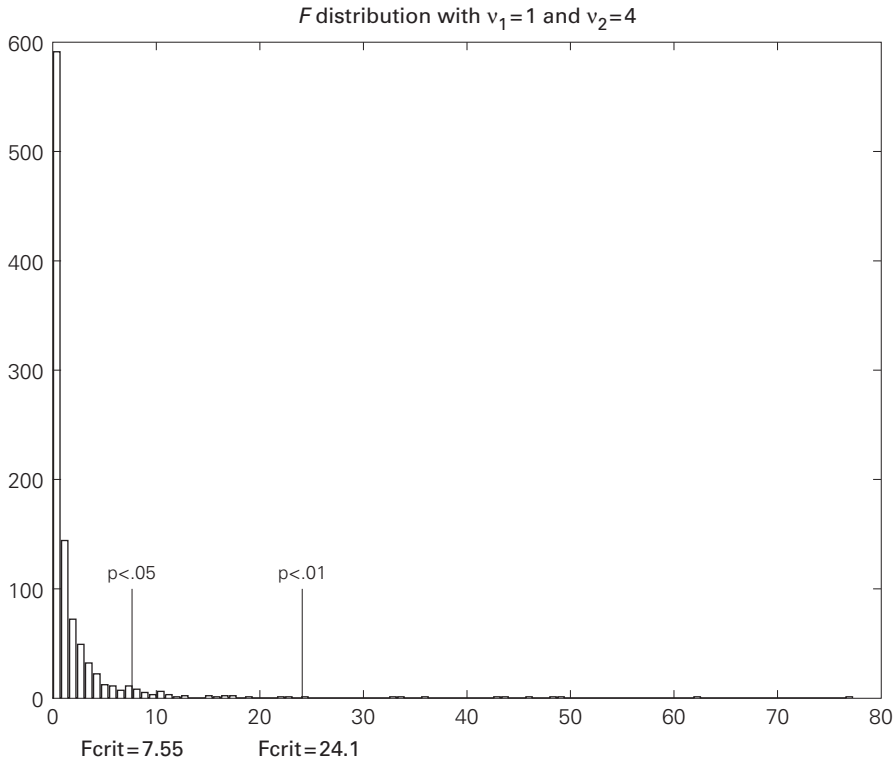


Figure 3.5 Histogram of F values computed from 1000 random samples when the null hypothesis is true, along with the critical values for $\alpha = .01$ and $\alpha = .05$.

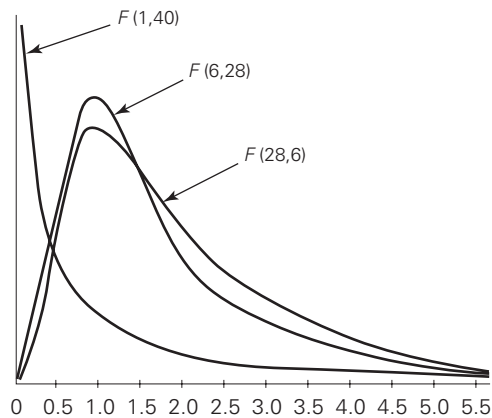


Figure 3.6 Some of Fisher's F distributions: The distributions shown here are Fisher's F distribution with the following sets of parameters (1) $\nu_1 = 1$ and $\nu_2 = 40$, (2) $\nu_1 = 6$ and $\nu_2 = 28$, and (3) $\nu_1 = 28$ and $\nu_2 = 6$. These distributions are used to test statistical hypotheses.

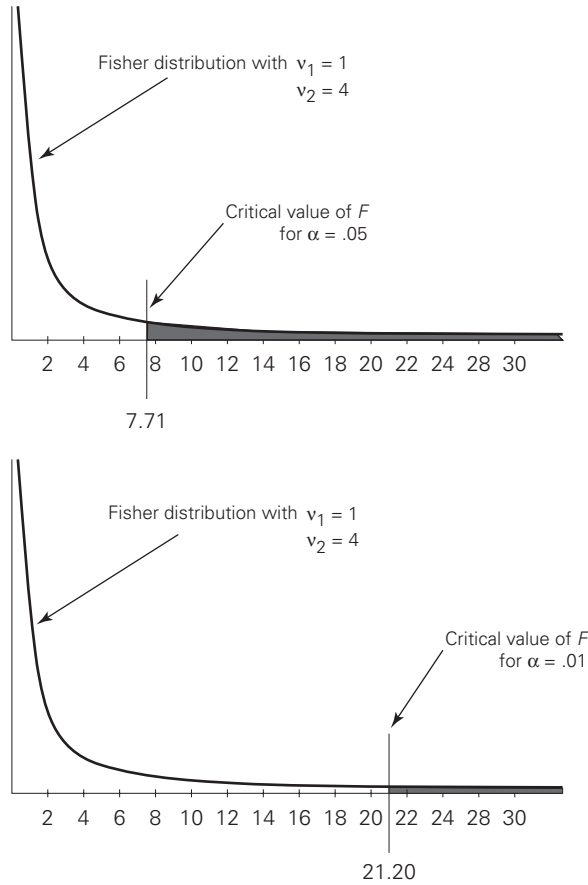


Figure 3.7 Fisher's F distribution with $\nu_1 = 1$ and $\nu_2 = 4$. This is the theoretical distribution of F when the null hypothesis is true, along with the critical values for $\alpha = .05$ and $\alpha = .01$.

This theoretical distribution is shown in Figure 3.7, along with the critical values of 7.71 corresponding to the α level of .05 (i.e. 5% of the values exceed 7.71 when the null hypothesis is true) and 21.20 corresponding to $\alpha = .01$ (i.e. 1% of the values exceed 21.20 when the null hypothesis is true). Note that these theoretical values are very close to the empirical values of 7.55 and 24.1 (with a larger number of trials used to compute the empirical distribution, the approximation would have been better).

As shown in Figure 3.8, the same decision of failing to reject the null hypothesis is reached with the theoretical distribution as with the empirical distribution.

3.2.6 Region of rejection, region of suspension of judgment, and critical value

Let us return to the rule for statistical decisions. We decide to reject H_0 when p (which is the probability associated with F) is less than the significance level α , or equivalently when the computed value of F is larger than the critical value corresponding to the

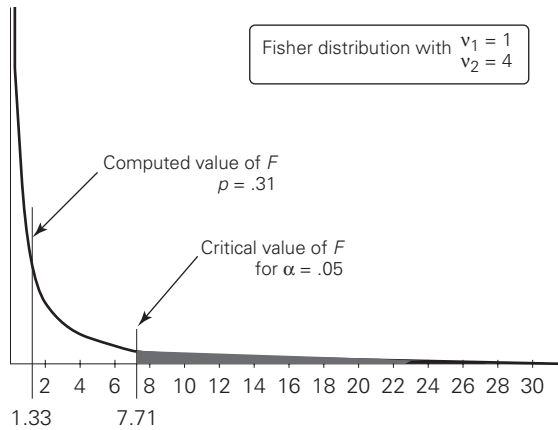


Figure 3.8 Fisher's F distribution with $v_1 = 1$ and $v_2 = 4$. This is the theoretical distribution of F when the null hypothesis is true, along with the critical values for $\alpha = .05$, and the value of F computed from the example.

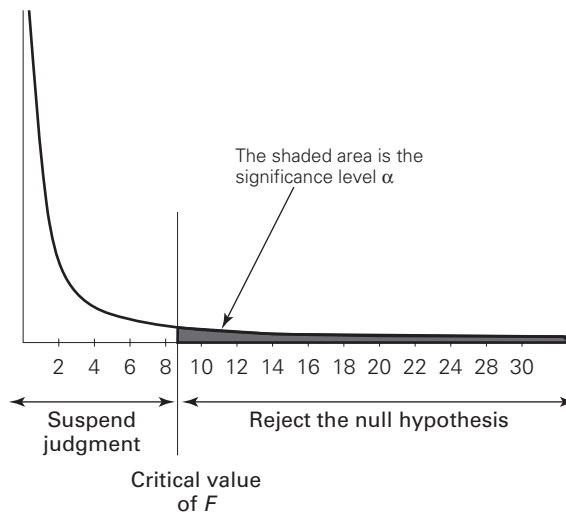


Figure 3.9 Statistical decision: region of rejection, region of suspension of judgment, critical value of Fisher's F .

chosen α -level. By doing so, we separate the values of F into two regions as illustrated in Figure 3.9.

- First, the values of F are associated with a probability less than or equal to the α level. In that case we reject H_0 and consequently accept H_1 . These values are in the *rejection region for the null hypothesis*, or simply *rejection region*.
- Second, the values of F are associated with a probability greater than the α level. In that case we cannot reject H_0 . As noted above, this does not lead us to accept H_1 . We are therefore unable to decide. These values are in the *region of suspension of judgment*.

The value of F that divides these two regions (the ‘frontier’ value) is called the *critical value*, and is denoted F_{critical} . It corresponds to the value of F associated with the probability α .

Because F is always positive (being obtained by the division of two positive quantities), the values of F greater than or equal to the critical value are associated with probabilities less than or equal to the α level. Those lead to the rejection of H_0 , and constitute the region of rejection. Conversely, the values of F less than the critical value are associated with probabilities greater than the α level. Those values do not lead to rejection of H_0 , and constitute the region of suspension of judgment.

Because it is difficult to compute by hand the probability associated with a given value of F , the critical values of F have been calculated for various α levels and differing values of ν_1 and ν_2 and are given in Table 2 in the Appendix on page 499 (see below for how to use the table). The critical value defines a decision rule:

- If the F calculated for the sample is greater than or equal to F_{critical} , its associated probability is less than α . We therefore decide to *reject the null hypothesis* and consequently accept the alternative hypothesis.
- If the F calculated for the sample is less than F_{critical} , its associated probability is greater than α , and so we cannot reject H_0 . Strictly speaking, we decide not to decide. Because we can never accept H_0 , all we can do is suspend judgment.

We can dispense with the step of finding the critical value when we can obtain the probability associated with F directly.

3.2.7 Using the table of critical values of Fisher’s F

One could describe each of the distributions of F in detail, and provide the associated probabilities under H_0 for a large number of possible values of F . Such tables exist, but they are rather voluminous and difficult to use, since we need a separate table for each pair of possible values of ν_1 and ν_2 . Using computer programs to calculate the probability associated with a critical value is a relatively recent development and is not yet in frequent use. Researchers have therefore looked for practical condensed ways of presenting the information in the sampling distributions for Fisher’s F .

As a result of the traditional practice of using only a small number of α levels (i.e. $\alpha = .05$, $\alpha = .01$), we can be happy with using only the critical values of F corresponding to these levels. In fact most tables will give only the values for the two ‘magical α levels.’ And you will find Table 2 in the Appendix which is a ritual table of F_{critical} which are the *critical values* of the distribution of Fisher’s F corresponding to the two levels $\alpha = .05$ and $\alpha = .01$. The critical values are at the intersection of the row corresponding to ν_2 and the column corresponding to ν_1 . The critical value for $\alpha = .05$ is above the critical value for $\alpha = .01$.

In our example, the critical values of F for $\nu_1 = 1$ and $\nu_2 = 4$ are found at the intersection of the column for $\nu_1 = 1$ and the row for $\nu_2 = 4$. We find at that intersection two numbers:

- 7.71: the critical value for $\alpha = .05$,
- 21.20: the critical value for $\alpha = .01$.

Values for F (calculated for 1 and 4 df) greater than the critical value 7.71 permit us to reject the null hypothesis at the .05 level. Values for F greater than 21.20 permit rejection of the null hypothesis at the .01 level.

3.2.8 Summary of the procedure for a statistical test

1. Express the statistical hypotheses in such a way that all the possible cases are covered by the two hypotheses (that is, so that either one or the other is true). This amounts to stating two hypotheses:
 - On the one hand, the *null hypothesis*: in the population as a whole there is no linear relationship between the two variables; and the observed value of the test statistic can be attributed to chance.
 - On the other hand, the *alternative hypothesis* (which will most often be an inexact statistical hypothesis): in the population as a whole there is a linear relationship between the two variables; and the observed value of the test statistic can be attributed to this relationship.

Note that the statistical hypotheses H_0 and H_1 are not statements about the sample, but about the population from which the sample is drawn.

2. Define a statistical index which will allow us to decide between the two hypotheses. In the framework of correlation, we could choose either the squared coefficient of correlation, r^2 , or the F of Fisher.
3. Choose a level of significance. The existence of the two conventional significance levels ($\alpha = .05$ and $\alpha = .01$) should not make this a routine choice. The choice of a significance level involves a compromise between two desirable goals: on the one hand minimizing Type I errors (the probability α) and on the other hand minimizing Type II errors (the probability β , which can be known only if H_1 is exact).
4. Compute or define the sampling distribution under H_0 . Here, this is the distribution of Fisher's F for $\nu_1 = 1$ and $\nu_2 = S - 2$.
5. Specify the value F_{critical} (if you decide to use a table rather than a computer program to obtain the probability of F) and consequently specify the regions of rejection of H_0 and of suspension of judgment.
6. Compute F and decide: reject H_0 if the probability associated with F is less than or equal to the α level; suspend judgment if the probability associated with F is greater than the α level. If we follow the procedure of looking up F_{critical} in the table, then reject H_0 if F is greater than or equal to F_{critical} , and suspend judgment if it is not.

3.2.9 Permutation tests: how likely are the results?

For both the Monte-Carlo and the traditional (i.e. Fisher) approaches, we need to specify the shape of the distribution under the null hypothesis. The Monte-Carlo approach can be used with any distribution (but we need to specify which one we want) and the traditional approach requires the normal distribution. An alternative way to look at a null hypothesis test is to evaluate if the pattern of results for the experiment is a *rare* event by comparing it with all the other patterns of results that could have arisen from these data. This approach originates with Student and Fisher (1935, see also, Pitman, 1937, 1938) who developed the (now standard) F approach because it was possible then to compute one F but very impractical to compute all these F s for all these permutations (nevertheless it seems that both Student and Fisher spent some inordinate amount of time doing so, in order to

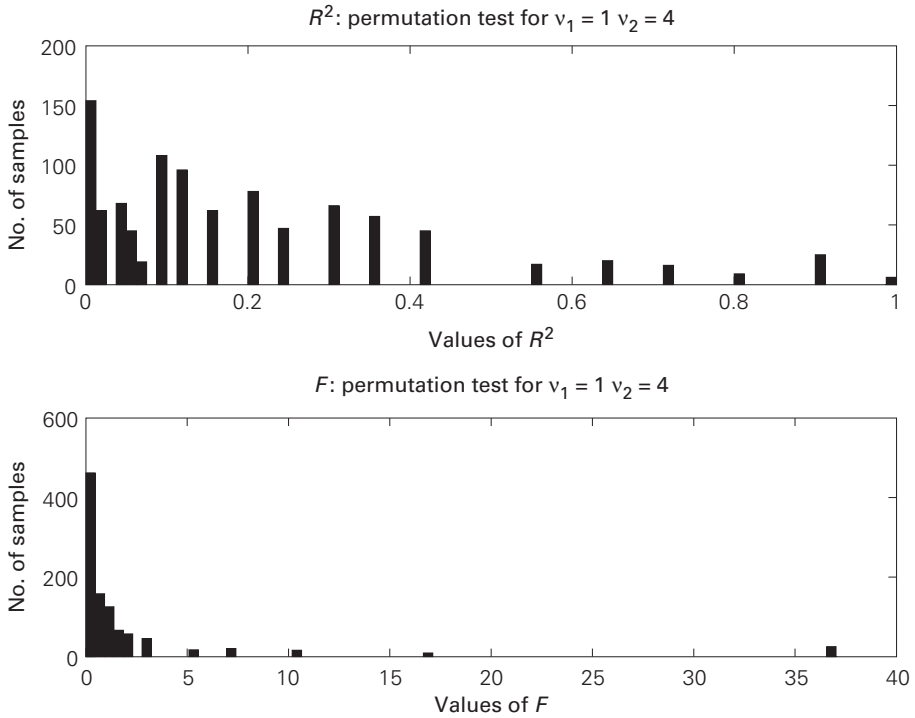


Figure 3.10 Histogram of F values computed from the $6! = 720$ possible permutations of the 6 scores from Table 3.1.

derive a ‘feel’ for the sampling distribution they were looking for). If Fisher could have had access to modern computers, it is likely that permutation tests would be the standard procedure.

For our example, we have 6 observations and therefore there are

$$6! = 6 \times 5 \times 4 \times 3 \times 2 = 720$$

different possible patterns of results. Each of these patterns corresponds to a given *permutation* of the data. For example, here is a possible permutation of the results from Table 3.1:

$W_1 = 1$	$W_2 = 3$	$W_3 = 4$	$W_4 = 4$	$W_5 = 5$	$W_6 = 7$
$Y_1 = 8$	$Y_2 = 10$	$Y_3 = 16$	$Y_4 = 12$	$Y_5 = 10$	$Y_6 = 4$

(Note: we just need to permute one of the two series of numbers, here we permuted Y .) This permutation gives a value of $r_{W \cdot Y} = -.30$ and of $r_{W \cdot Y}^2 = .09$. We computed the value of $r_{W \cdot Y}$ for the remaining 718 permutations. The histogram is plotted in Figure 3.10, where, for convenience, we have also plotted the histogram of the corresponding F values.

For our example, we want to use the permutation test in order to compute the probability associated with $r_{W \cdot Y}^2 = .25$. This is obtained by computing the *proportion* of $r_{W \cdot Y}^2$ larger than .25. We counted 220 $r_{W \cdot Y}^2$ out of 720 larger or equal to .25, this gives a probability of

$$p = \frac{220}{720} = .306.$$

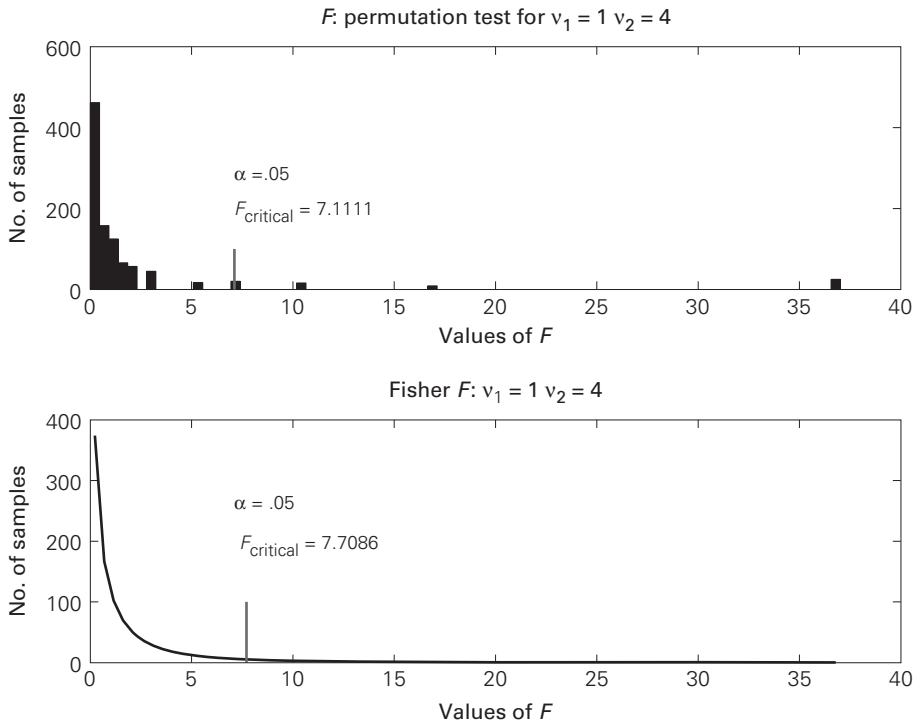


Figure 3.11 Histogram of $r_{W.Y}$ (top) and F (bottom) values computed from the $6! = 720$ possible permutations of the 6 scores from Table 3.1 and theoretical Fisher distribution with $\nu_1 = 1$ and ν_2 .

Interestingly this value is very close to the values found with the other approaches (cf. Monte-Carlo $p = .310$ and Fisher distribution $p = .313$). This similarity is confirmed by Figure 3.11 where we have plotted the permutation histogram for F along with the Fisher distribution.

When the number of observations is small (as is the case for this example with 6 observations), it is possible to compute all the possible permutations. In this case we have an *exact* permutation test. However, the number of permutations grows very quickly when the number of observations increases. For example, with 20 observations (cf. our ‘word example’ in Chapter 2) the total number of permutations is close to 2.4×10^{18} . Such large numbers obviously prohibit computing all the permutations and therefore in these cases we *approximate* the permutation test by using a large number (say 10,000 or 100,000) of random permutations. In this case we have an *approximate* permutation test.

3.3 Not zero is not enough!

The null hypothesis test that we perform with the F test ends up with a binary decision: either we reject the null hypothesis or we do not; either we conclude that r is different from 0 or not. Such a conclusion is often too narrow: we want to know more than that. And current APA recommendations agree: most psychologists want to know (1) how large the population correlation really is, and (2) how much they can trust such an estimation. For the first question, the problem is to estimate the intensity of the population correlation from the value computed

on a sample. For the second question, the problem is to evaluate the confidence interval of the coefficient of correlation.

3.3.1 Shrunk and adjusted r values

The value of the coefficient of correlation computed on a given sample is a *descriptive* statistic. If we want to *estimate* the value of the correlation in the population from the value obtained on the sample, then the sample coefficient of correlation is not a good estimate because it always overestimates the population correlation. This problem is similar to the problem of the estimation of the variance of a population from the variance computed on a sample: Here also, the variance of the sample overestimates the population variance. For the variance problem, the solution is rather simple: we obtain the correct estimation by dividing the sum of squares by $S - 1$ instead of S (see Appendix A). Unfortunately, the problem is much more complex for the coefficient of correlation. In fact, there is not one simple correction formula and therefore there are several correction formulas available (and some can be quite daunting, see, for example, Abdi, 1987, p. 115 ff., where one of the correction formulas takes more than half a page!).

The corrected values of r go under different names: corrected r , shrunk r , or adjusted r (there are some subtle differences between these different appellations; if you want to know more, consult Darlington, 1990, or Stevens, 1996). Here we will describe only two correction formulas. The first one is the one most often used (most statistical programs use it). The second one (sometimes called Stein's correction, see Hersberg, 1969; Fowler, 1985) is slightly more complex, more stringent also, but is preferable because it is more accurate.

We will illustrate these two formulas with our introductory example where we had obtained a value of $r = -.50$ with a sample made of $S = 6$ observations. These formulas do not use r but r^2 which, for our example is equal to $r^2 = (-.50)^2 = .25$. We will denote the corrected value of the squared coefficient of correlation $\tilde{r}^2$ (read ' r tilde'). The first formula estimates the value of the population correlation as

$$\tilde{r}^2 = 1 - \left[(1 - r^2) \left(\frac{S-1}{S-2} \right) \right], \quad (3.4)$$

for our example, this gives:

$$\tilde{r}^2 = 1 - \left[(1 - r^2) \left(\frac{S-1}{S-2} \right) \right] = 1 - \left[(1 - .25) \times \frac{5}{4} \right] = 1 - \left[.75 \times \frac{5}{4} \right] = .06.$$

If we want to get back to the value of r (instead of r^2), the first formula suggests that the value of the correlation in the population is around⁵

$$\tilde{r} = -\sqrt{\tilde{r}^2} = -\sqrt{.06} = -.24.$$

The second formula estimates the value of $\tilde{r}^2$ as

$$\tilde{r}^2 = 1 - \left[(1 - r^2) \left(\frac{S-1}{S-2} \right) \left(\frac{S-2}{S-3} \right) \left(\frac{S+1}{S} \right) \right], \quad (3.5)$$

⁵ Remember that the value of the sample coefficient of correlation was equal to $-.5$, so we need to put back the minus sign.

for our example, this gives:

$$\begin{aligned}
 \hat{r}^2 &= 1 - \left[(1 - r^2) \left(\frac{S-1}{S-2} \right) \left(\frac{S-2}{S-3} \right) \left(\frac{S+1}{S} \right) \right] \\
 &= 1 - \left[(1 - .25) \times \left(\frac{5}{4} \right) \left(\frac{4}{3} \right) \left(\frac{7}{6} \right) \right] \\
 &= 1 - \left[(.75) \times \left(\frac{5}{4} \right) \left(\frac{4}{3} \right) \left(\frac{7}{6} \right) \right] \\
 &= 1 - 1.46 \\
 &= -.46 .
 \end{aligned}$$

Oooops! Notice that we have obtained a *negative* value for a squared coefficient of correlation! This means that we should estimate that the value of the correlation in the population is in fact equal to zero. A value of the correlation of zero is clearly smaller than .06. This illustrates that the second formula always gives a *smaller* value than the first formula for the estimated value of the magnitude of the correlation in the population.

3.3.2 Confidence interval

The value of r computed from a sample is an estimation of the correlation of the population from which this sample was obtained. Suppose that we obtain a new sample from the same population and that we compute the value of the coefficient of correlation for this new sample. In what range is this value likely to fall? This question is answered by computing the *confidence interval* of the coefficient of correlation. This gives a lower bound and a higher bound between which the population coefficient of correlation is likely to stand. So, for example, taking into account the value of the correlation that we have found on a sample of a given size, we want to say that the value of the correlation in the population had 95 out of 100 chances of being between .30 and .50.

Using confidence intervals is more general than a null hypothesis test because if the confidence interval excludes the value zero then we can reject the null hypothesis. But, in addition, a confidence interval gives a *range* of probable values for the correlation. Using confidence intervals has another big advantage: we can act *as if* we could accept the null hypothesis.⁶ In order to do so, we first compute the confidence interval of the coefficient of correlation and look at the largest magnitude it can have. If we consider that this value is small, then we can say that even if the magnitude of the population correlation is not zero, it is too small to be of interest.

Conversely, we can give more weight to a conclusion if we show that the smallest possible value for the coefficient of correlation will still be large enough to be impressive.

The problem of computing the confidence interval for r has been explored by Student and then by Fisher (like so many statistical things!). Fisher found that the problem could be simplified by transforming r into another variable called Z . This transformation is called the Fisher's Z transform. The new Z variable has the beautiful property of having a sampling

⁶ Remember that one of the best insults for psychologists is 'to try to accept the null'.

distribution which is close to the normal distribution.⁷ Therefore we can use the normal distribution to compute the confidence interval of Z and this will give a lower and a higher bound for the population values of Z . Then we can transform these bounds back into r values (using the inverse transformation) and this gives a lower and upper bound for the possible values of r in the population.

3.3.2.1 Fisher's Z transform

Fisher's Z transform gives a variable called Z from a given coefficient of correlation r according to the following formula:

$$Z = \frac{1}{2} [\ln(1 + r) - \ln(1 - r)] . \quad (3.6)$$

In this formula, the expression $\ln$ means 'use the *natural* logarithm' (most hand calculators have this operation, and it is often denoted $\ln$).

The inverse transformation, which gives r from Z , is obtained using the following formula:

$$r = \frac{\exp\{2 \times Z\} - 1}{\exp\{2 \times Z\} + 1} . \quad (3.7)$$

In this formula, the expression $\exp\{x\}$ means 'raise the number e to the power x ' (i.e. $\exp\{x\} = e^x$, where e is Euler's constant which is approximately $e \approx 2.71828\dots$).

Most hand calculators can be used to compute both transformations, another simple method is to use Table 3 on page 502 in the Appendix.

Fisher showed that the new Z variable has a sampling distribution which is normal with a mean of zero and a variance of $S - 3$. From this distribution we can compute directly the upper and lower bounds of Z and then transform them back into values of r .

3.3.3 How to transform r to Z : an example

We will illustrate the computation of the confidence interval for the coefficient of correlation with the previous example where we computed a coefficient of correlation of $r = -.5$ on a sample made of $S = 6$ observations. The procedure can be decomposed into seven steps which are detailed below.

1. Before doing any computation we need to choose an α level that will correspond to the probability of finding the population value of r in the confidence interval. Suppose we chose the value $\alpha = .05$. This means that we want to obtain a confidence interval such that there are $(1 - \alpha) = (1 - .05) = .95$ chances of having the population value being in the confidence interval that we will compute.
2. Find in the table of the normal distribution the critical values corresponding to the chosen α level. Call this value Z_{α} . The most frequently used values are:

- $Z_{\alpha=.10} = 1.645$ (for $\alpha = .10$)
- $Z_{\alpha=.05} = 1.960$ (for $\alpha = .05$)
- $Z_{\alpha=.01} = 2.575$ (for $\alpha = .01$)
- $Z_{\alpha=.001} = 3.325$ (for $\alpha = .001$).

⁷ And this is why we call it Z !

3. Transform r into Z . This can be done with Table 3 in the Appendix or with Equation 3.6. For the present example, with $r = -.5$, we find that $Z = -0.5493$.
4. Compute a quantity called Q as

$$Q = Z_{\alpha} \times \sqrt{\frac{1}{S-3}}.$$

For our example we obtain:

$$Q = Z_{.05} \times \sqrt{\frac{1}{6-3}} = 1.960 \times \sqrt{\frac{1}{3}} = 1.1316.$$

5. Compute the lower and upper limits for Z as:

$$\text{lower limit} = Z_{\text{lower}} = Z - Q = -0.5493 - 1.1316 = -1.6809$$

$$\text{upper limit} = Z_{\text{upper}} = Z + Q = -0.5493 + 1.1316 = 0.5823.$$

6. Transform Z_{lower} and Z_{upper} into r_{lower} and r_{upper} . This can be done with Table 3 in the Appendix or with Equation 3.7. For the present example, we find that

$$\text{lower limit} = r_{\text{lower}} = -.9330$$

$$\text{upper limit} = r_{\text{upper}} = .5243.$$

As you can see, the range of possible values of r is very large: the value of the coefficient of correlation that we have computed could come from a population whose correlation could have been as low as $r_{\text{lower}} = -.9330$ or as high as $r_{\text{upper}} = .5243$. Also, because zero is in the range of possible values, we cannot reject the null hypothesis (which is the conclusion we reached with the null hypothesis test).

It is worth noting that because the Z transformation is non-linear, the confidence interval is *not* symmetric around r .

Finally, current APA practice recommends the use of confidence intervals routinely because this approach is more informative than standard hypothesis testing.

3.3.4 Confidence intervals with the bootstrap

A modern Monte-Carlo approach for deriving confidence intervals was proposed by Efron (1979, 1981, see also Efron and Tibshirani, 1993). This approach is probably the most important advance for inferential statistics in the second part of the 20th century.

The idea is marvelously simple but can be implemented only with modern computers and this explains why it was developed only recently. With the bootstrap approach, to estimate the sampling distribution of a statistic we just treat the sample as if it were the population. Practically this means that in order to estimate the sampling distribution of a statistic, we just need to create bootstrap samples obtained by drawing observations *with replacement*⁸ from the original sample. The distribution of the bootstrap samples is taken as the population distribution. Confidence intervals are then computed from the percentile of this distribution.

⁸ When we draw observation with replacement, each observation is put back into the sample after it has been drawn, therefore an observation can be drawn several times.

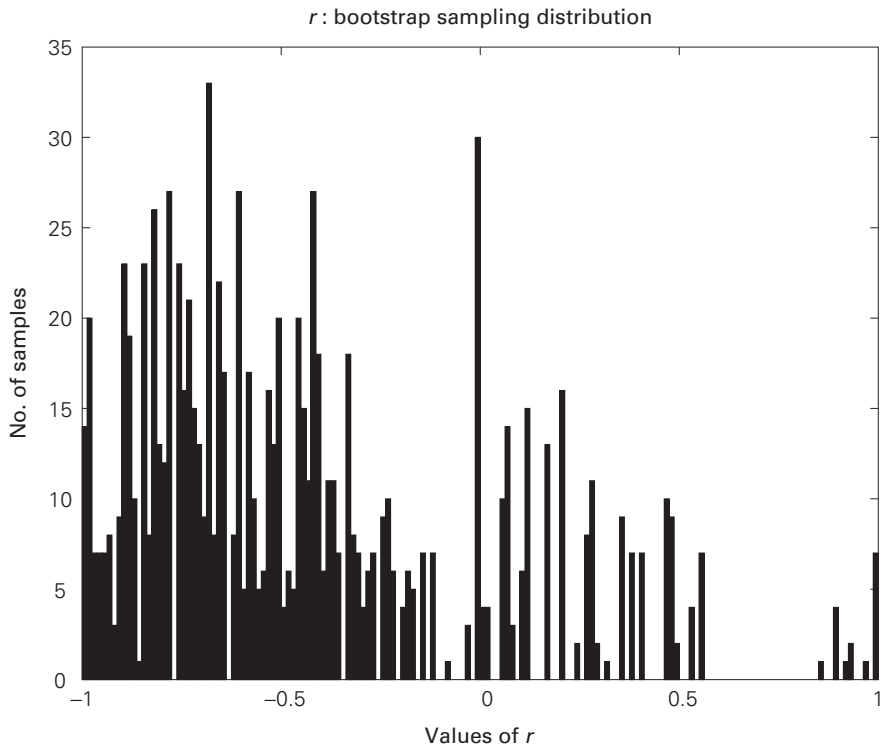


Figure 3.12 Histogram of $r_{W.Y}$ values computed from 1,000 bootstrapped samples drawn with replacement from the data of Table 3.1.

For our example, the first bootstrap sample that we obtained comprised the following observations (note that some observations are missing and some are repeated as a consequence of drawing with replacement):

$$s_1 = 5, \; s_2 = 1, \; s_3 = 3, \; s_4 = 2, \; s_5 = 3, \; s_6 = 6 .$$

This gives the following values for the first bootstrapped sample obtained drawing with replacement from Table 3.1 (page 40):

$W_1 = 5$	$W_2 = 1$	$W_3 = 4$	$W_4 = 3$	$W_5 = 4$	$W_6 = 7$
$Y_1 = 8$	$Y_2 = 16$	$Y_3 = 12$	$Y_4 = 10$	$Y_5 = 12$	$Y_6 = 10$

This bootstrapped sample gives a correlation of $r_{W.Y} = -.73$.

If we repeat this bootstrapping procedure for 1,000 samples, we obtain the sampling distribution of $r_{W.Y}$ as shown in Figure 3.12. From this figure, it is obvious that the value of $r_{W.Y}$ varies a lot with such a small sample (in fact, it covers the whole range of possible values, from -1 to $+1$). In order to find the upper and the lower limits of a confidence interval we look for the corresponding percentiles. For example, if we select a value of $\alpha = .05$, we look at the values of the bootstrapped distribution corresponding to the 2.5th and the 97.5th percentiles. In our example, we find that 2.5% of the values are smaller than $-.9487$ and that 2.5% of the values are larger than $.4093$. Therefore, these two values constitute the lower and the upper limits of the 95% confidence interval of the population estimation of $r_{W.Y}$

(cf. the values obtained with Fisher's Z transform of $-.9330$ and $.5243$). Contrary to Fisher's Z transform approach, the bootstrap limits are not dependent upon assumptions about the population or its parameters (but it is comforting to see that these two approaches concur for our example). Because the value of zero is in the confidence interval of $r_{W.Y}$, we cannot reject the null hypothesis. This shows once again that the confidence interval approach provides more information than the null hypothesis approach.

Chapter summary

3.4 Key notions of the chapter

Below are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

Null hypothesis	False-alarm and miss
Alternative hypothesis	Critical value
Inferential statistic	Region of rejection of the null hypothesis
Statistical decision	Region of suspension of judgment
Population and sample	Table of critical values
Statistics and parameters	Decision rule
Exact and inexact statistical hypotheses	Six steps of a statistical test
Power of a test	Shrunken r
Probability associated with the F ratio	Confidence interval
Sampling distribution	Permutation test
Distributions of Fisher's F	Bootstrap
Rejection of the null hypothesis	Z -transform
Type I and Type II errors	

3.5 New notations

Below are the new notations introduced in this chapter. Test yourself on their meaning.

F	β , $(1 - \alpha)$, and $(1 - \beta)$
H_0 and H_1	$\tilde{r}$
ν_1 and ν_2	Z_{lower} and Z_{upper} ; r_{lower} and r_{upper}
α -level and ritual values $\alpha = .05$ and $\alpha = .01$	

3.6 Key formulas of the chapter

Over the page are the main formulas introduced in this chapter: try to go through them and understand what they mean.

$$F = \frac{r^2}{1 - r^2} \times (S - 2).$$

$$\nu_1 = 1 \quad \text{and} \quad \nu_2 = S - 2.$$

First correction formula for $\tilde{r}^2$:

$$\tilde{r}^2 = 1 - \left[(1 - r^2) \left(\frac{S - 1}{S - 2} \right) \right]. \quad (3.8)$$

Second correction formula for $\tilde{r}^2$:

$$\tilde{r}^2 = 1 - \left[(1 - r^2) \left(\frac{S - 1}{S - 2} \right) \left(\frac{S - 2}{S - 3} \right) \left(\frac{S + 1}{S} \right) \right]. \quad (3.9)$$

Transformation from r to Z :

$$Z = \frac{1}{2} [\ln(1 + r) - \ln(1 - r)]. \quad (3.10)$$

Transformation from Z to r :

$$r = \frac{\exp\{2 \times Z\} - 1}{\exp\{2 \times Z\} + 1}. \quad (3.11)$$

Confidence interval for Z :

$$Z_{\text{limits}} = Z \pm \left(Z_{\alpha} \times \sqrt{\frac{1}{S - 3}} \right). \quad (3.12)$$

3.7 Key questions of the chapter

Below are some questions about the content of this chapter. All the answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- * Why do we want to infer conclusions about the population and not about the sample actually tested?
- * Why is the alternative hypothesis an inexact hypothesis?
- * Why is it impossible to accept the null hypothesis (or reject the alternative hypothesis) from the data gathered from a sample of observations?
- * Why do α and β vary in opposite directions?
- * What are the three main factors having an influence on β ?
- * Why do we use critical values?
- * Why do we favor the α -levels of .01 and .05?
- * What does the null hypothesis test tell us about the value of r in the population?
- * How can we estimate the value of r in the population?
- * What are the differences between a confidence interval approach and a null hypothesis test?
- * Do you think that the confidence interval approach should replace null hypothesis testing?

4

Simple linear regression

4.1 Introduction

Linear regression is a technique used for analyzing experimental data, when both the independent variable and the dependent variable are quantitative variables.

The main idea behind regression analysis is to *predict* the dependent variable (denoted by the letter Y) from the independent variable (denoted by the letter X). In other words, we want to express the dependent variable as a *function* of the independent variable. Because the independent variable is quantitative, the simplest functional relation will be the equation for a straight line:

$$\hat{Y} = a + bX,$$

where $\hat{Y}$ (read Y -hat) is the *predicted* value of the dependent variable¹, X is the independent variable, a is the intercept, and b is the slope. Because $\hat{Y}$ is predicted from a line given as an equation involving X , we say that $\hat{Y}$ is a *linear* function of X .

4.2 Generalities

4.2.1 The equation of a line

Figure 4.1 shows an example of a linear relationship. The values of X , the independent variable, are represented on the horizontal axis. The values of Y , the dependent variable, are represented on the vertical axis. As you can see, the values of $\hat{Y}$ corresponding to the different values of X draw a line, and so the equation

$$\hat{Y} = a + bX$$

is called a *linear* equation. In other words, $\hat{Y}$ is a *linear function* of X . The meaning of the intercept and the slope can be derived from the graph. The intercept, noted a , is the value of $\hat{Y}$ corresponding to a value of $X = 0$. This is the value of $\hat{Y}$ when the line crosses (i.e. intercepts) the Y -axis. The slope, noted b , gives the *rate* of change of $\hat{Y}$ corresponding to a change of one unit in X .

¹ You may wonder why we are using the notation $\hat{Y}$, instead of Y ; this is because we need to distinguish the predicted value $\hat{Y}$ from the actual or observed value Y .

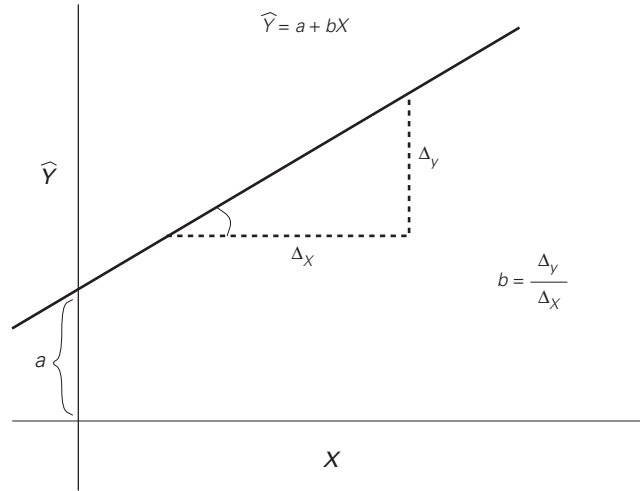


Figure 4.1 Plot of the linear functional relation $\hat{Y} = a + bX$. The *intercept* (a) is the value of $\hat{Y}$ corresponding to a value of $X = 0$. The *slope* of the line (b) is given by dividing Δ_Y by Δ_X or ‘the rise over the run’.

The value of the slope is found by dividing the change in $\hat{Y}$ by the corresponding change in X . For example if we have two values X_1 and X_2 with corresponding values $\hat{Y}_1$ and $\hat{Y}_2$, the slope is found as

$$b = \frac{\hat{Y}_2 - \hat{Y}_1}{X_2 - X_1}.$$

This relation is very often written as

$$b = \frac{\Delta \hat{Y}}{\Delta X},$$

where $\Delta \hat{Y}$ (read Delta y-hat) represents the rate of change in $\hat{Y}$, and ΔX (read Delta x) represents the rate of change in X .

4.2.2 Example of a perfect line

Suppose that we have obtained the hypothetical data gathered in Table 4.1. These data are plotted in Figure 4.2. As you can see, they describe a perfect line. In order to find the intercept we just need to look at the value of $\hat{Y}$ when the line crosses the $\hat{Y}$ axis which is 1. In order to find the value of the slope we need to take two values of X and the corresponding values of $\hat{Y}$. For example, when $X = 1$, $\hat{Y} = 3$ and when $X = 5$, $\hat{Y} = 11$ (any other set would give the same result for b), therefore

$$\begin{aligned} b &= \frac{\Delta \hat{Y}}{\Delta X} = \frac{\hat{Y}_2 - \hat{Y}_1}{X_2 - X_1} \\ &= \frac{11 - 3}{5 - 1} \\ &= \frac{8}{4} \\ &= 2. \end{aligned} \tag{4.1}$$

X	1	2	3	4	5	6
$\hat{Y}$	3	5	7	9	11	13

Table 4.1 A set of data plotting a perfect line (plotted in Figure 4.2).

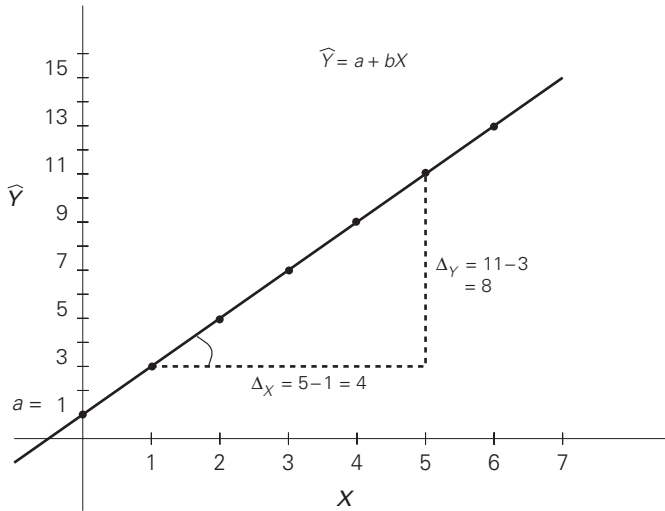


Figure 4.2 Plot of the data from Table 4.1. The equation of the line is $\hat{Y} = a + bX = 1 + 2X$.

Hence the data from Table 4.1 can be obtained from the equation:

$$\hat{Y} = a + bX = 1 + 2 \times X.$$

4.2.3 An example: reaction time and memory set

When dealing with real data, it is very rare that they fall nicely on a line. Hence most of the time, our problem is to find the *best* possible line that fits the data. To illustrate the argument, suppose that we have been performing the following replication of an experiment originally designed by Sternberg (1969). The goal of this experiment was to find out if subjects were ‘scanning’ material stored in short term memory.

Subjects are asked to memorize a set of random letters (like *lqwh*) called the *memory set*. The number of letters in the set is called the *memory set size*. The subjects are presented with a *probe* letter (say *q*). They should give the answer *Yes* if the probe is present in the memory set and *No* if the probe is not present in the memory set (here the answer should be *Yes*). The time it takes the subjects to answer is recorded.

In this replication, each subject is tested one hundred times with a constant memory set size. For half of the trials, the probe is present, whereas for the other half the probe is absent. Four different set sizes are used: 1, 3, 5, and 7 letters. Twenty (fictitious) subjects are tested (five per condition). For each subject, we use as a dependent variable the mean reaction time for the correct *Yes* answers. The research hypothesis is that subjects need to scan serially the letters in the memory set and that they need to compare each letter in turn with the probe. If this is the case, then each letter would add a given time to the reaction time. Hence the

slope of the line would correspond to the time needed to process one letter of the memory set. The time needed to produce the answer and encode the probe should be constant for all conditions of the memory set size. Hence it should correspond to the intercept.

The results of this experiment are given in Table 4.2. A first way of examining the relationship between two series of measurements (or variables) is to plot the results of the dependent variable against the independent variable. In a scatterplot for a regression analysis, we always plot the independent variable (X) as the horizontal axis and the dependent variable (Y) as the vertical axis. This is done in Figure 4.3. Just by looking at this figure it seems that the relationship that was hypothesized between the independent variable ‘memory set size’ and the dependent variable ‘reaction time’ holds. Precisely, there is a positive relationship between the independent variable and the dependent variable. The larger the memory set size, the longer it takes the subjects to produce an answer. It seems also that the points are falling along a line that describes the shape of the functional relationship between X and Y . This line is called the *regression line*.

Memory set size			
$X = 1$	$X = 3$	$X = 5$	$X = 7$
433	519	598	666
435	511	584	674
434	513	606	683
441	520	605	685
457	537	607	692

Table 4.2 Data from a replication of a Sternberg (1969) experiment. Each data point represents the mean reaction time for the Yes answers of a given subject. Subjects are tested in only one condition. Twenty (fictitious) subjects participated in this experiment. For example the mean reaction time of subject one who was tested with a memory set of 1 was 433 ($Y_1 = 433, X_1 = 1$).

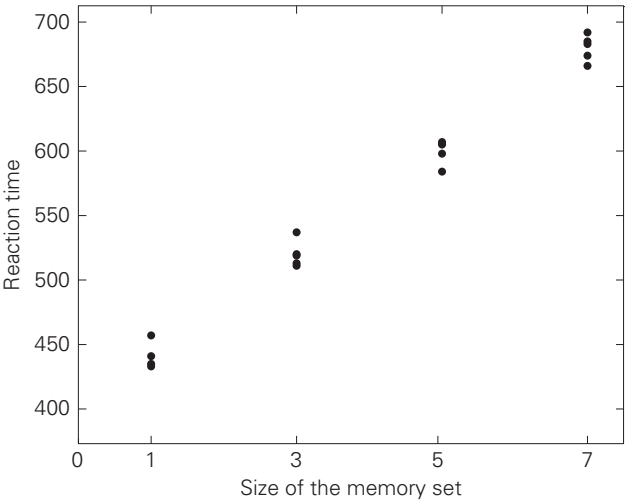


Figure 4.3 Plot of the results of a replication of Sternberg (1969). Each data point represents the mean reaction time for the Yes answers of a given subject. Data from Table 4.2.

4.3 The regression line is the 'best-fit' line

Because the data points seem to be quite close to the regression line, it makes sense to try to predict the value of the dependent variable as a linear function of the independent variable. The problem that we are now faced with is how could we find the line that describes best the linear relationship between X and Y ? The main idea is to suppose that the values we observe *should* have been positioned on a straight line, and that the deviations from the data points to this straight line are due to some random effects that we call the *measurement error*. This error could be due to the fact that subjects do not react all alike, that some subjects may have been tired and some others in a very frisky mood, that our measuring apparatus can be flawed, that we could have made some clerical mistakes when entering the data, etc. The important point is that we suppose that the observation points should have been positioned on the line. A characteristic of a random error is that it is as likely to be positive as to be negative. In other words, error measurement is not systematic. Hence, its effect should be to increase the observed values as often as to decrease these values. This implies that the average error should be zero (because it is as often positive as it is negative). A final point is that the error can affect only the dependent variable. It cannot affect the independent variable because we *control* it, whereas we *measure* the dependent variable.

With that in mind, we can define the regression line as being the line that describes best the shape of the data. Intuitively, this line will be as much in the 'middle of the data' as possible. Such an intuition can be sufficient to try to draw an approximate line by hand. We need however a more formal criterion or definition, in order to be able to compute the parameters (i.e. a and b) of the regression line. Formally, we say that the predicted values of the dependent variable should be obtained as

$$\hat{Y} = a + bX, \quad (4.2)$$

and that the difference between what we observe and the prediction,

$$Y - \hat{Y}, \quad (4.3)$$

can be considered as an error. The term $(Y - \hat{Y})$ is also called the *residual of the prediction* or simply the *residual* (because this quantity is what is left out after we have been trying to predict Y).

For the regression line to be as much as possible in 'the middle of the data points', the magnitude of the differences between the observed values and the predicted values should be as small as possible. Precisely, we want to find the values of a and b such that the sum of the squared deviations from the line that they determine will be as small as possible.² Formally, we say that we are looking for the regression line that *minimizes* (i.e. gives the smallest value of) the sum of the squared deviations. In other words, we say that we are looking for the line that provides the best fit for the Y scores.

With a formula, we say that we are looking for the parameters a and b such that

$$\min = \sum_s (Y_s - \hat{Y}_s)^2 = \sum_s [Y_s - (a + bX_s)]^2. \quad (4.4)$$

² The reason why we choose the squared deviations instead of, say, the magnitude, is essentially for practical reasons. As shown in a digression (see Section 4.7.1, page 77) later on, elementary calculus makes it quite easy to find a and b for squared values. It is much more difficult in most cases to find the values of the parameters when dealing with magnitude.

We show later on in a digression (that you can skip if you feel comfortable with accepting these results at face value) that the values of a and b that define the regression line are:

- Slope

$$b = \frac{SCP_{XY}}{SS_X} = \frac{\hat{\sigma}_Y}{\hat{\sigma}_X} r_{X \cdot Y} , \quad (4.5)$$

where SCP_{XY} is the sum of cross-products of X and Y [$SCP_{XY} = \sum (X - M_X)(Y - M_Y)$]; SS_X is the sum of squares of X [$SS_X = \sum (X - M_X)^2$]; $r_{X \cdot Y}$ is the coefficient of correlation between Y and X ; $\hat{\sigma}_Y$ and $\hat{\sigma}_X$ denote the standard deviation of Y and X respectively.

As you can see, we indicate two equivalent ways to compute the slope of the regression line. Depending upon the data at hand, one formula can be more handy. In general, the first is the easiest (but the second is the one found most often in textbooks!).

- Intercept

$$a = M_Y - bM_X . \quad (4.6)$$

4.4 Example: reaction time and memory set

Table 4.3 gives most of the quantities needed to compute the values of the slope and the intercept for our example. As a first step we can gather all the statistics that we need to compute b and a in Table 4.4.

If we use these values to determine the equation line for the memory set example, we obtain the following values:

- Slope

$$\begin{aligned} b &= \frac{SCP_{XY}}{SS_X} = \frac{4,000}{100} \\ &= \frac{\hat{\sigma}_Y}{\hat{\sigma}_X} r_{X \cdot Y} = \frac{92.22}{2.29} \times .9950 \\ &= 40 . \end{aligned} \quad (4.7)$$

- Intercept

$$\begin{aligned} a &= M_Y - bM_X \\ &= 560 - 40 \times 4 \\ &= 560 - 160 \end{aligned} \quad (4.8)$$

$$= 400 . \quad (4.9)$$

We can now write the equation that predicts the values of the dependent variable (reaction time) from the independent variable (memory set size):

$$\begin{aligned} \hat{Y} &= a + bX \\ &= 400 + 40 \times X \end{aligned} \quad (4.10)$$

Subject	Set size X	RT Y	x	x^2	$x \times y$	y	y^2
1	1	433	-3	9	381	-127	16,129
2	1	435	-3	9	375	-125	15,625
3	1	434	-3	9	378	-126	15,876
4	1	441	-3	9	357	-119	14,161
5	1	457	-3	9	309	-103	10,609
6	3	519	-1	1	41	-41	1,681
7	3	511	-1	1	49	-49	2,401
8	3	513	-1	1	47	-47	2,209
9	3	520	-1	1	40	-40	1,600
10	3	537	-1	1	23	23	529
11	5	598	1	1	38	38	1,444
12	5	584	1	1	24	24	576
13	5	606	1	1	46	46	2,116
14	5	605	1	1	45	45	2,025
15	5	607	1	1	47	47	2,209
16	7	666	3	9	318	106	11,236
17	7	674	3	9	342	114	12,996
18	7	683	3	9	369	123	15,129
19	7	685	3	9	375	125	15,625
20	7	692	3	9	396	132	17,424
Σ	80	11,200	0	100	4,000	0	161,600
	ΣX	ΣY		SS_X	SCP_{XY}		SS_Y

Table 4.3 The different quantities needed to compute the value of the parameters a and b of a regression line $\hat{Y} = a + bX$. The following abbreviations are used to label the columns of the Table: $x = (X - M_X)$, $y = (Y - M_Y)$.

X	1	2	3	4	5	6	7	8	6
ΣX	M_X	SS_X	$\hat{\sigma}_X$	ΣY	M_Y	SS_Y	$\hat{\sigma}_Y$	SCP_{XY}	r_{XY}
80	4	100	2.29	11,200	560	161,600	92.22	4,000	.9950

Table 4.4 The basic statistics needed to compute a and b .

or

$$\text{predicted reaction time} = 400 + 40 \times \text{memory set size} .$$

For example, with a memory set size of 3, the predicted reaction time should be

$$\begin{aligned}
 \hat{Y} &= a + bX \\
 &= 400 + 40 \times X \\
 &= 400 + 40 \times 3 \\
 &= 520 .
 \end{aligned}
 \tag{4.11}$$

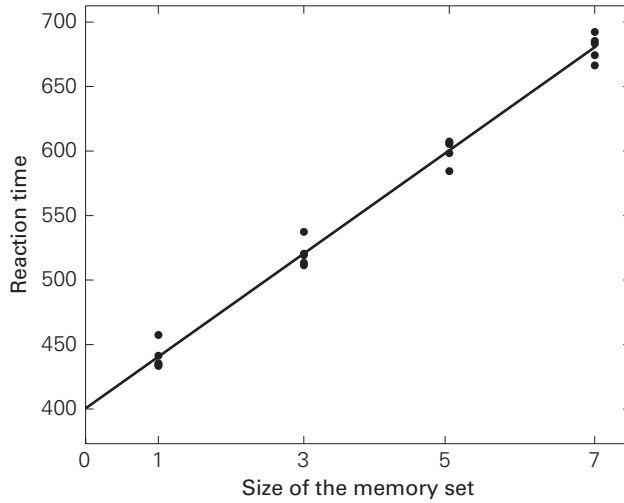


Figure 4.4 Plot of the results of a replication of Sternberg (1969). Each data point represents the mean reaction time for the Yes answers of a given subject. The regression line to predict reaction time from memory set size is $\hat{Y} = a + bX = 400 + 40X$. The line passes through the values 440, 520, 600, 680 that correspond respectively to the predictions for a memory set size of 1, 3, 5 and 7 digits. Data from Table 4.2.

Figure 4.4 shows the original data points with the regression line superimposed. From the value of the slope and the intercept of the line, we can say that adding one letter in the memory set increases reaction time by 40 ms (which corresponds to the time needed to ‘scan’ the letter), and that to respond and to initiate the task takes 400 ms.

4.5 How to evaluate the quality of prediction

We know that the prediction $\hat{Y}$ of the dependent variable computed from the independent variable is the best possible one. But how can we evaluate the quality of the prediction? Is the prediction very good or just mediocre?

In order to evaluate the quality of the prediction, we need to have a numerical index quantifying the quality of the prediction. This index should evaluate the similarity between two distributions: the original distribution, made of the Y values, and the new distribution, made of the $\hat{Y}$ values. And we have seen previously in Chapter 2 on correlation (p. 17) that the coefficient of correlation does exactly that! Therefore, in order to evaluate the quality of the prediction it suffices to compute the coefficient of correlation between the predicted values and the actual values. Hence, the first thing we need to do is to compute $r_{\hat{Y}Y}$.

Table 4.5 gives all the information needed to compute this coefficient of correlation. For our example, we can find that:

$$\begin{aligned}
 r_{\hat{Y}Y} &= \frac{SCP_{\hat{Y}Y}}{\sqrt{SS_{\hat{Y}}SS_Y}} \\
 &= \frac{160,000}{\sqrt{160,000 \times 161,600}} \\
 &= .9950 .
 \end{aligned} \tag{4.12}$$

Subject	Y	y	y^2	$\hat{Y}$	$\hat{y}$	$\hat{y}^2$	$y \times \hat{y}$
1	433	-127	16,129	440	-120	14,400	15,240
2	435	-125	15,625	440	-120	14,400	15,000
3	434	-126	15,876	440	-120	14,400	15,120
4	441	-119	14,161	440	-120	14,400	14,280
5	457	-103	10,609	440	-120	14,400	12,360
6	519	-41	1,681	520	-40	1,600	1,640
7	511	-49	2,401	520	-40	1,600	1,960
8	513	-47	2,209	520	-40	1,600	1,880
9	520	-40	1,600	520	-40	1,600	1,600
10	537	-23	529	520	-40	1,600	920
11	598	38	1,444	600	40	1,600	1,520
12	584	24	576	600	40	1,600	960
13	606	46	2,116	600	40	1,600	1,840
14	605	45	2,025	600	40	1,600	1,800
15	607	47	2,209	600	40	1,600	1,880
16	666	106	11,236	680	120	14,400	12,720
17	674	114	12,996	680	120	14,400	13,680
18	683	123	15,129	680	120	14,400	14,760
19	685	125	15,625	680	120	14,400	15,000
20	692	132	17,424	680	120	14,400	15,840
Σ	11,200		161,600	11,200		160,000	160,000
	ΣY		SS_Y	$\Sigma \hat{Y}$		$SS_{\hat{Y}}$	$SCP_{Y\hat{Y}}$

Table 4.5 Quantities needed to compute the coefficient of correlation between Y and $\hat{Y}$ (data from Table 4.2). The following abbreviations are used: $y = Y - M_Y$, $\hat{y} = (\hat{Y} - M_{\hat{Y}})$.

In order to evaluate if regression analysis gives a better prediction than chance alone would give, the procedure is the same as for the standard coefficient of correlation seen before (cf. Chapter 3).

The first step is to compute an F ratio:

$$\begin{aligned}
 F &= \frac{r_{\hat{Y} \cdot Y}^2}{1 - r_{\hat{Y} \cdot Y}^2} \times (S - 2) \\
 &= \frac{.9950^2}{1 - .9950^2} \times (20 - 2) \\
 &= 1,800.00 .
 \end{aligned} \tag{4.13}$$

We need then to follow the standard test procedure to see if this value of F can be obtained by mere chance (H_0) or if it reflects the existence of a linear relationship between the dependent and the independent variables (H_1). For that, we need first to decide on a significance level α . Because we have no particular reason to chose a specific value of α we can use the traditional level of $\alpha = .05$ and $\alpha = .01$. The next step is to find the critical value of F for $\nu_1 = 1$ and $\nu_2 = S - 2 = 20 - 2 = 18$ in the Fisher's F table (see Table 2 in the Appendix on page 499). For $\nu_1 = 1$, and $\nu_2 = 18$ we find in the table a critical value of F equal to 4.41 for $\alpha = .05$ and 8.28 for $\alpha = .01$. Because the computed value of $F = 1,800.00$ is larger than both critical values we can reject the null hypothesis and accept the alternative hypothesis stating that there is a linear relationship between the size of the memory set and the time taken by a subject to respond.

We have now seen one way that can provide us with a measure of the likelihood that the regression line provides a better prediction of a dependent variable from an independent variable than chance alone. We will now examine another approach to this same problem. This second approach is of interest because it easily extends to material that will follow in this text concerning analysis of variance.

4.6 Partitioning the total sum of squares

4.6.1 Generalities

Imagine that we want to predict some unknown Y scores. If all we have is the Y scores recorded for a sample of subjects, then our best guess concerning an unknown Y score is the mean of Y : M_Y . We are very likely to be wrong with this guess, but it is still the best guess we can make. The degree to which we are likely to be wrong is given by the dispersion (or variance) of the Y scores around their mean. This variance is given by:

$$\hat{\sigma}_Y^2 = \frac{\sum (Y_s - M_Y)^2}{S - 1} = \frac{SS_Y}{S - 1}.$$

Note that the sum of squares of Y (SS_Y) used to compute this variance is generally referred to as the ‘total’ sum of squares (SS_{total}).

We can try to improve our guess concerning a Y score by using our knowledge of the independent variable X . In other words, we can try to predict a subject’s Y score from her or his X value using the regression equation. If the X and Y are correlated, the regression line will provide a better guess than the mean. The improvement in our new guess will reflect the intensity of the correlation between X and Y because the larger the magnitude of the correlation is, the better X predicts Y . How good or bad is our new guess? Previously the variance around the mean of Y was an index of the quality (or lack thereof) of the prediction. Now, as we are using the regression line as a predictor, the degree to which we are likely to be wrong is given by the dispersion (i.e. the variance) of the Y scores around the regression line. This variance is given by:

$$\hat{\sigma}_{\text{residual}}^2 = \frac{\sum (Y_s - \hat{Y}_s)^2}{S - 2} = \frac{SS_{\text{residual}}}{S - 2}.$$

That is, we will still not predict the Y scores perfectly, but our guess will be a lot better if there is effectively a relationship between Y and X . The stronger the relationship is (i.e. the closer to 1 or -1 the coefficient of correlation is) the better our guess, and hence, the smaller the residual variance.

The problem now is to find a way to assess if the prediction that we made from the regression line is better than the one we would have made from the mean of the Y scores (M_Y). In other words, we want to find out *how much* the knowledge of the independent variable X improves the prediction of the dependent variable Y as compared to the mean of Y . More formally we want to find a numerical index that will tell us if the dispersion of

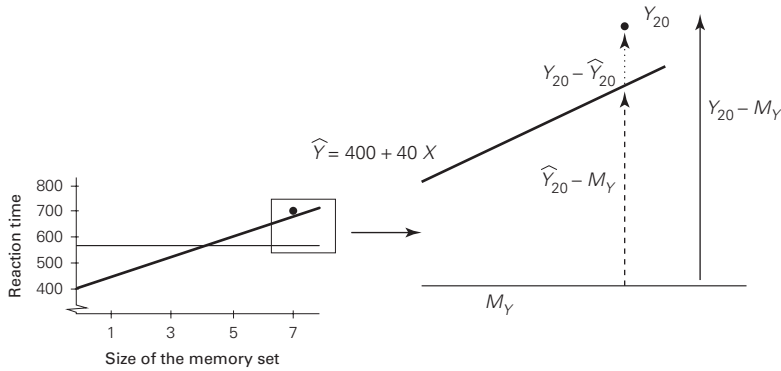


Figure 4.5 Partitioning the total sum of squares: two ways to arrive at the score for subject 20 in the memory set size experiment.

the Y scores around the regression line (i.e. the residual variance, $\hat{\sigma}_{\text{residual}}^2$) is smaller than the dispersion of the Y scores around the mean of the Y scores.

4.6.2 Partitioning the total sum of squares

In the previous example on reaction time and size of memory set, we saw that the variation in subjects' reaction time can be regarded as made of two quantities: variation in the size of the memory set (the independent variable), and residual variation (or variation attributable to other sources such as individual differences, measurement error, etc.).

Another way of stating this point, which will be very useful when we come to analysis of variance, is that a given Y score can be reached by different routes (cf. Figure 4.5).

- First, we can get to a particular Y score by starting at the mean of the Y scores, M_Y , and going straight to the score (solid arrow in Figure 4.5).
- Second, we could think of the Y score as constructed out of a set of components. In that case we would start at M_Y and proceed to the regression line (dashed arrow in Figure 4.5) to get the predicted value $\hat{Y}$. That move represents the effect of X on the score. The next move is to get to the score from the regression line (dotted arrow in Figure 4.5).

Note that these moves correspond to the series of guesses we were making concerning the value of Y . Looking at the moves by which you can reach a score, we see that each of those moves is in effect a deviation from a mean or a regression line:

- $Y - M_Y$: deviation of a Y score from the mean of the Y scores.
- $\hat{Y} - M_Y$: deviation of a predicted score from the mean of the Y scores.
- $Y - \hat{Y}$: deviation of an actual score from the regression line.

As such, each deviation makes a contribution to a sum of squared deviations:

- *Sum of squares total*: sum of the squared deviations of each subject's Y score from the mean of the scores.

$$SS_{\text{total}} = \sum_s (Y_s - M_Y)^2.$$

- *Sum of squares regression*: sum of the squared deviations of each subject's predicted $\hat{Y}$ to the mean of the Y scores.

$$SS_{\text{regression}} = \sum_s (\hat{Y}_s - M_Y)^2.$$

- *Sum of squares residual*: sum of the squared deviations of each subject's Y score from the predicted value $\hat{Y}$.

$$SS_{\text{residual}} = \sum_s (Y_s - \hat{Y}_s)^2.$$

It is an interesting and very useful fact that the sum of squared deviations comes out the same no matter which route you follow to the score. That is, the sum of squares total (i.e. the sum of squared deviations of Y scores around their mean, M_Y) is equal to the sum of squares regression (i.e. the sum of squared deviations of values on the regression line, $\hat{Y}$, from M_Y) plus the sum of squares residual (i.e. the sum of squared deviations of the Y scores from the values on the regression line $\hat{Y}$). More formally we would say that the total sum of squares can be partitioned into two parts:

$$SS_{\text{total}} = SS_{\text{regression}} + SS_{\text{residual}}$$

$$\sum_s (Y_s - M_Y)^2 = \sum_s (\hat{Y}_s - M_Y)^2 + \sum_s (Y_s - \hat{Y}_s)^2. \quad (4.14)$$

The sum of squares of regression expresses the variability of the predicted values ($\hat{Y}$) around the mean of the Y scores (M_Y) and the sum of squares residual expresses the variability of the Y scores around the regression line ($\hat{Y}$).

Note that by dividing the sum of squares of regression by the sum of squares total we can find out the proportion of total variance that can be accounted for by the independent variable. This proportion is generally referred to as the *coefficient of determination*. It is equal to the squared coefficient of correlation between the independent variable X and the dependent variable Y .

$$r_{X \cdot Y}^2 = \frac{SS_{\text{regression}}}{SS_{\text{total}}}.$$

In the numerical example presented above, $r = .99$, hence $r^2 = .98$ which indicates that 98% of the variance in Y is attributable to the relationship between the dependent and the independent variables, leaving 2% of the variance to be attributed to other factors. Another way of interpreting the squared coefficient of correlation is to say that an r_{XY}^2 of .98 means that knowing X and using it to predict Y will improve the quality of prediction by 98% over simply using the mean of Y as a predictor.

In summary, we have seen that the variability of the Y scores around their mean can be decomposed into two parts:

- variability of predicted values around the mean of the Y scores (sum of squares regression)
- variability of the actual values around the regression line (sum of squares residual).

The next step in finding a numerical index that will tell us how much the knowledge of X improves our prediction of Y is to transform these sums of squares into a variance-like

quantity called a mean square (MS for short or $\hat{\sigma}^2$). The general formula for a mean square is:

$$MS_{\text{something}} = \frac{SS_{\text{something}}}{df_{\text{something}}}$$

where df represents the number of degrees of freedom associated with the sum of squares.

4.6.3 Degrees of freedom

The notion of degrees of freedom will be detailed later on (in Chapter 7, page 140). For now, it suffices to know that since the point corresponding to M_X and M_Y belongs to the regression line, we need only to find one other point to position the regression line, and hence the number of degrees of freedom for the sum of squares regression is 1 (because we need only two points to draw a straight line):

$$df_{\text{regression}} = 1 .$$

For the residual sum of squares, if we try to make up an example, we will find that we can ‘play’ with only $S - 2$ values to make it work. This is the case because the residuals are forced to have a mean of zero and a sum of squares equal to SS_{residual} . Hence the number of degrees of freedom for the residual sum of squares is the number of subjects *per* group minus 2:

$$df_{\text{residual}} = S - 2 .$$

4.6.4 Variance of regression and variance of residual

Recall that a variance (mean square) is equal to a sum of squares divided by its degrees of freedom.

- The variance (or mean square) of regression is computed as

$$\hat{\sigma}_{\text{regression}}^2 = MS_{\text{regression}} = \frac{SS_{\text{regression}}}{df_{\text{regression}}} = \frac{SS_{\text{regression}}}{1} = SS_{\text{regression}} .$$

- The variance (or mean square) of residual is computed as

$$\hat{\sigma}_{\text{residual}}^2 = MS_{\text{residual}} = \frac{SS_{\text{residual}}}{df_{\text{residual}}} = \frac{SS_{\text{residual}}}{S - 2} .$$

4.6.5 Another way of computing F

Recall that we are trying to find a numerical index that will tell us if our regression line is a good predictor of the Y scores. Intuitively, the regression line will be a better predictor of the Y scores than the mean of the scores, M_Y , if the deviation from the regression line (residual variance) is smaller than the deviation of the predicted values from the mean (regression variance). A standard way of evaluating the magnitude of a quantity relative to another is to compute the ratio of these two quantities. The ratio of the regression variance to the residual variance is named F and is given by

$$F = \frac{\hat{\sigma}_{\text{regression}}^2}{\hat{\sigma}_{\text{residual}}^2} = \frac{MS_{\text{regression}}}{MS_{\text{residual}}} .$$

A value of F greater than 1 indicates that the regression variance is larger than the residual variance, and hence that the regression line is a better predictor than the mean of Y . For a fixed number of subjects, the greater the value of F , the better the prediction. The significance of F is then estimated using the statistical procedure described in the previous chapter with $\nu_1 = 1$ and $\nu_2 = S - 2$ degrees of freedom.

4.6.6 Back to the numerical example

In order to compute the index F , we need first to find the sums of squares of regression and residual. An easy way of doing that is to use a table like Table 4.6.

The sum of squares of regression is given by the sum of the right-most column of Table 4.6:

$$SS_{\text{regression}} = 160,000 ,$$

and the sum of squares of the residuals is given by the sum of the 7th column of Table 4.6:

$$SS_{\text{residual}} = 1,600 .$$

Note that the total sum of squares or sum of squares of Y (cf. Table 4.5) can be obtained by adding the sum of squares regression and the sum of squares residual:

$$SS_Y = SS_{\text{regression}} + SS_{\text{residual}} = 160,000 + 1,600 = 161,600 .$$

Subject	X	Y	M_Y	$\hat{Y}$	$Y - \hat{Y}$	$(Y - \hat{Y})^2$	$\hat{Y} - M_Y$	$(\hat{Y} - M_Y)^2$
1	1	433	560	440	-7	49	-120	14,400
2	1	435	560	440	-5	25	-120	14,400
3	1	434	560	440	-6	36	-120	14,400
4	1	441	560	440	1	1	-120	14,400
5	1	457	560	440	17	289	-120	14,400
6	3	519	560	520	-1	1	-40	1,600
7	3	511	560	520	-9	81	-40	1,600
8	3	513	560	520	-7	49	-40	1,600
9	3	520	560	520	0	0	-40	1,600
10	3	537	560	520	17	289	-40	1,600
11	5	598	560	600	-2	4	40	1,600
12	5	584	560	600	-16	256	40	1,600
13	5	606	560	600	6	36	40	1,600
14	5	605	560	600	5	25	40	1,600
15	5	607	560	600	7	49	40	1,600
16	7	666	560	680	-14	196	120	14,400
17	7	674	560	680	-6	36	120	14,400
18	7	683	560	680	3	9	120	14,400
19	7	685	560	680	5	25	120	14,400
20	7	692	560	680	12	144	120	14,400
Σ						1,600		160,000

Table 4.6 Quantities needed for the computation of the regression and residual sums of squares in a regression analysis.

The variance (or mean square) of regression is then computed as

$$\hat{\sigma}_{\text{regression}}^2 = MS_{\text{regression}} = \frac{SS_{\text{regression}}}{df_{\text{regression}}} = \frac{SS_{\text{regression}}}{1} = SS_{\text{regression}} = 160,000 .$$

The variance (or mean square) of the residual is computed as

$$\hat{\sigma}_{\text{residual}}^2 = MS_{\text{residual}} = \frac{SS_{\text{residual}}}{df_{\text{residual}}} = \frac{SS_{\text{residual}}}{S-2} = \frac{1,600}{20-2} = \frac{1,600}{18} = 88.89 .$$

4.6.6.1 Index F

Finally, the F ratio is equal to

$$F = \frac{\hat{\sigma}_{\text{regression}}^2}{\hat{\sigma}_{\text{residual}}^2} = \frac{MS_{\text{regression}}}{MS_{\text{residual}}} = \frac{160,000}{88.89} = 1,800.00 ,$$

which matches the value we found previously.

4.7 Mathematical digressions

In the text, we mentioned several results that justified the approach taken. Some readers like to know how these mathematical derivations are obtained. In what follows, we detail the rationale behind some of the procedures used.

4.7.1 Digression 1: finding the values of a and b

In this section, we show that the values of a and b that satisfy the equation

$$\min = \sum (Y - \hat{Y})^2 = \sum [Y - (a + bX)]^2$$

are

- for the slope:

$$b = \frac{SCP_{XY}}{SS_X} .$$

Actually, an alternative expression for the slope makes it easier to write this proof. As a warm-up exercise, we will show that the slope can be computed as

$$\begin{aligned} b &= \frac{SCP_{XY}}{SS_X} \\ &= \frac{\sum XY - SM_X M_Y}{\sum X^2 - SM_X^2} ; \end{aligned} \quad (4.15)$$

- for the intercept:

$$a = M_Y - bM_X .$$

The goal of this section is to give you an insight into the techniques used to derive the statistical tools. If you feel comfortable accepting this result³ at face value you can skip

³ Or if it seems obvious to you that this is the solution.

this section. If, however, you want to understand how these values are obtained, you'll find that a small amount of study can go quite a long way.

The general idea is to use a basic result from calculus stating that the minimum of a quadratic function is reached when its derivative is zero.

As a preliminary warm-up, let us start with rewriting the formula for the slope in order to find the result given in Equation 4.15:

$$\begin{aligned}
 b &= \frac{SCP_{XY}}{SS_X} \\
 &= \frac{\sum (X - M_X)(Y - M_Y)}{\sum (X - M_X)^2} \\
 &= \frac{\sum (XY - M_X Y - M_Y X + M_X M_Y)}{\sum (X^2 + M_X^2 - 2M_X X)} \\
 &= \frac{\sum XY - M_X \sum Y - M_Y \sum X + SM_X M_Y}{\sum X^2 + SM_X^2 - 2M_X \sum X} \\
 &= \frac{\sum XY - SM_X M_Y - SM_Y M_X + SM_X M_Y}{\sum X^2 + SM_X^2 - 2SM_X^2} \\
 &= \frac{\sum XY - SM_X M_Y}{\sum X^2 - SM_X^2} .
 \end{aligned} \tag{4.16}$$

The first step is to develop the quantity we want to minimize. Let us denote by $\mathcal{E}$ (like *Error*) the expression to minimize:

$$\mathcal{E} = \sum (Y - \hat{Y})^2 . \tag{4.17}$$

Developing $\mathcal{E}$ gives:

$$\begin{aligned}
 \mathcal{E} &= \sum (Y - \hat{Y})^2 \\
 &= \sum (Y^2 + \hat{Y}^2 - 2Y\hat{Y}) \\
 &= \sum Y^2 + \sum \hat{Y}^2 - 2 \sum Y\hat{Y} \\
 &= \sum Y^2 + \sum (a + bX)^2 - 2 \sum Y(a + bX) \\
 &= \sum Y^2 + Sa^2 + b^2 \sum X^2 + 2ab \sum X \\
 &\quad - 2a \sum Y - 2b \sum XY .
 \end{aligned} \tag{4.18}$$

It is now time to remember the magical formula from calculus: *a quadratic function reaches its minimum value when its derivatives vanish*. So the first step is to take the derivative of $\mathcal{E}$ with respect to a and b . This gives:

$$\frac{\partial \mathcal{E}}{\partial a} = 2Sa + 2b \sum X - 2 \sum Y \quad (4.19)$$

and

$$\frac{\partial \mathcal{E}}{\partial b} = 2b \sum X^2 + 2a \sum X - 2 \sum XY. \quad (4.20)$$

We are now trying to find the values of a and b for which the derivatives vanish. This gives the following set of equations (called the ‘normal’ equations):

$$\frac{\partial \mathcal{E}}{\partial a} = 0 \quad \Longleftrightarrow \quad Sa + b \sum X - \sum Y = 0 \quad (4.21)$$

and

$$\frac{\partial \mathcal{E}}{\partial b} = 0 \quad \Longleftrightarrow \quad b \sum X^2 + a \sum X - \sum XY = 0. \quad (4.22)$$

From Equation 4.21 the value for the intercept is found as

$$\begin{aligned} a &= \frac{1}{S} \left(\sum Y - b \sum X \right) \\ &= M_Y - bM_X. \end{aligned} \quad (4.23)$$

Plugging in the result of Equation 4.23 in Equation 4.20 and developing gives:

$$\begin{aligned} b \sum X^2 - \sum XY + a \sum X &= 0 \\ b \sum X^2 + a \sum X - \sum XY &= 0 \\ b \sum X^2 + (M_Y - bM_X) \sum X - \sum XY &= 0 \\ b \sum X^2 + M_Y \sum X - bM_X \sum X - \sum XY &= 0 \\ b \left(\sum X^2 - SM_X^2 \right) - \left(\sum XY - SM_Y M_X \right) &= 0 \end{aligned}$$

which gives the following value for the slope:

$$\begin{aligned} b &= \frac{\sum XY - SM_X M_Y}{\sum X^2 - SM_X^2} \\ &= \frac{SCP_{XY}}{SS_X}, \end{aligned} \quad (4.24)$$

which completes the proof.

An alternative formula for the slope comes with a bit of rewriting of Equation 4.24:

$$\begin{aligned}
 b &= \frac{SCP_{XY}}{SS_X} \\
 &= \frac{r_{X \cdot Y} \sqrt{SS_X SS_Y}}{SS_X} \\
 &= \frac{\sqrt{SS_Y}}{\sqrt{SS_X}} r_{X \cdot Y} \\
 &= \frac{\sqrt{SS_Y}/(S-1)}{\sqrt{SS_X}/(S-1)} r_{X \cdot Y} \\
 &= \frac{\hat{\sigma}_Y}{\hat{\sigma}_X} r_{X \cdot Y} .
 \end{aligned} \tag{4.25}$$

4.7.2 Digression 2: the mean of $\hat{Y}$ is equal to the mean of Y

The mean of the predicted values (the $\hat{Y}$ values) is the same as the mean of the observed values (the Y values). In order to demonstrate this, we will show that the sum of the predicted values is equal to the sum of the observed values. With a formula, we want to show that:

$$\sum \hat{Y} = \sum Y \quad \Longleftrightarrow \quad M_Y = M_{\hat{Y}} . \tag{4.26}$$

We can obtain this result by developing the defining formula of the $\hat{Y}$ values:

$$\begin{aligned}
 \sum \hat{Y} &= \sum (a + bX) \\
 &= \sum (M_Y - bM_X + bX) \\
 &= \sum M_Y - b \sum M_X + b \sum X \\
 &= SM_Y - bSM_X + bSM_X \\
 &= SM_Y \\
 &= \sum Y .
 \end{aligned} \tag{4.27}$$

Et voilà!

4.7.3 Digression 3: the residuals ($Y - \hat{Y}$) and the predicted values $\hat{Y}$ are uncorrelated

In this section, we show that the residuals from the prediction ($Y - \hat{Y}$) are uncorrelated with the predicted values $\hat{Y}$. This property justifies the additive decomposition of the total sum of squares into ‘regression sum of squares’ on the first hand and ‘residual sum of squares’ on the other hand.

To show that the residual and the predicted values are uncorrelated, it suffices to show that their cross-product is zero. That is to say we need to show that:

$$\sum (Y - \hat{Y})(\hat{Y} - M_Y) = 0 . \tag{4.28}$$

The first step is to develop Equation 4.28:

$$\begin{aligned}\sum (Y - \hat{Y})(\hat{Y} - M_Y) &= \sum Y\hat{Y} - \sum \hat{Y}^2 - M_Y \sum Y + M_Y \sum \hat{Y} \\ &= \sum Y\hat{Y} - \sum \hat{Y}^2 - SM_Y^2 + SM_Y^2\end{aligned}$$

(cf. Equation 4.26)

$$= \sum Y\hat{Y} - \sum \hat{Y}^2. \quad (4.29)$$

Hence the problem boils down to showing that

$$\sum Y\hat{Y} = \sum \hat{Y}^2. \quad (4.30)$$

We will show that both quantities are equal to

$$SM_Y^2 + \frac{SCP_{XY}^2}{SS_X}.$$

Let us start by the first term of Equation 4.30:

$$\begin{aligned}\sum Y\hat{Y} &= \sum Y(a + bX) \\ &= \sum Y(M_Y - bM_X + bX) \\ &= \sum Y[M_Y + b(X - M_X)] \\ &= M_Y \sum Y + b \sum Y(X - M_X) \\ &= SM_Y^2 + b \sum Y(X - M_X) \\ &= SM_Y^2 + \frac{SCP_{XY} \times SCP_{XY}}{SS_X} \\ &= SM_Y^2 + \frac{SCP_{XY}^2}{SS_X}.\end{aligned} \quad (4.31)$$

This completes the first part of the proof.

To deal with the left term of Equation 4.30 we use the same technique of developing the formula:

$$\begin{aligned}\sum \hat{Y}^2 &= \sum (a - bX)^2 \\ &= \sum (M_Y - bM_X + bX)^2 \\ &= \sum [M_Y + b(X - M_X)]^2 \\ &= \sum [M_Y^2 + b^2(X - M_X)^2 + 2M_Y b(X - M_X)]\end{aligned}$$

$$= SM_Y^2 + b^2 \sum (X - M_X)^2 + 2M_Y b \sum (X - M_X) .$$

But $\sum (X - M_X) = 0$ (this is the sum of the deviations to the mean), hence

$$\begin{aligned} \sum \hat{Y}^2 &= SM_Y^2 + b^2 \sum (X - M_X)^2 \\ &= SM_Y^2 + \frac{SCP_{XY}^2}{SS_X^2} \times SS_X \\ &= SM_Y^2 + \frac{SCP_{XY}^2}{SS_X} , \end{aligned} \tag{4.32}$$

which completes the proof.

4.7.4 Digression 4: $r_{\hat{Y}, Y} = r_{X, Y}$

In this section, we show that the correlation between the predicted values (the $\hat{Y}$ values) and the original values (the Y values) is actually equal to the correlation between the X values and the Y values. Or, with a formula, we want to show that

$$r_{\hat{Y}, Y} = r_{X, Y} .$$

As usual, there are several ways of showing that. The route we shall use derives from one of the ‘Z-score formulas’ for the coefficient of correlation. Recall that the correlation between two variables W and Y can be computed using Z-scores as:

$$r_{W, Y} = \frac{1}{S} \sum Z_W Z_Y .$$

Hence in order to show that

$$r_{\hat{Y}, Y} = r_{X, Y} ,$$

it suffices to show that the Z-score for each $\hat{Y}$ is the same as the Z-score for each X . With the formula, we want to show that for all observations the following identity is true:

$$Z_{\hat{Y}_s} = Z_{X_s} .$$

The first step is to express the mean of the $\hat{Y}$ in terms of the slope and intercept. Remember that

$$\hat{Y} = a + bX .$$

From this definition, we find that $M_{\hat{Y}}$ is equal to $a + bM_X$. To obtain this result we develop the formula

$$\begin{aligned} M_{\hat{Y}} &= \frac{1}{S} \sum \hat{Y} \\ &= \frac{1}{S} \sum (a + bX) \\ &= \frac{1}{S} \left(\sum a + \sum bX \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{S} (Sa + SbM_X). \\
&= a + bM_X.
\end{aligned} \tag{4.33}$$

We show now that $\sigma_{\hat{Y}} = b\sigma_X$. Doing so is equivalent to showing that $SS_{\hat{Y}} = b^2 SS_X$. Once again we substitute and develop:

$$\begin{aligned}
SS_{\hat{Y}} &= \sum (\hat{Y} - M_{\hat{Y}})^2 \\
&= \sum [(a + bX) - (a + bM_X)]^2 \\
&= \sum [a + bX - a - bM_X]^2 \\
&= \sum [bX - bM_X]^2 \\
&= \sum [b(X - M_X)]^2 \\
&= \sum b^2 (X - M_X)^2 \\
&= b^2 \left[\sum (X - M_X)^2 \right] \\
&= b^2 SS_X.
\end{aligned} \tag{4.34}$$

Now, to show that

$$Z_{\hat{Y}_s} = Z_{X_s},$$

as usual, we substitute and develop (to make the proof easier to read we have dropped the index s):

$$\begin{aligned}
Z_{\hat{Y}} &= \frac{\hat{Y} - M_{\hat{Y}}}{\sigma_{\hat{Y}}} \\
&= \frac{a + bX - (a + bM_X)}{b\sigma_X} \\
&= \frac{a + bX - a - bM_X}{b\sigma_X} \\
&= \frac{bX - bM_X}{b\sigma_X} \\
&= \frac{b(X - M_X)}{b\sigma_X} \\
&= \frac{(X - M_X)}{\sigma_X} \\
&= Z_X.
\end{aligned} \tag{4.35}$$

Note in passing that the result that we have proved is quite general. It shows that the correlation is *invariant for any linear transformation* of the variables.

Chapter summary

4.8 Key notions of the chapter

Below are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

Regression line	Estimation of the quality of prediction
Regression equation	Total, regression, and residual sums of squares
Residual or experimental error	Squared coefficient of correlation
Prediction of the dependent variable from the independent variable	

4.9 New notations

Below are the new notations introduced in this chapter. Test yourself on their meaning.

$$\hat{Y}$$

$$r_{X \cdot Y}^2$$

$$SS_{\text{total}}, SS_{\text{regression}}, SS_{\text{residual}}$$

$$df_{\text{total}}, df_{\text{regression}}, df_{\text{residual}}$$

$$MS_{\text{regression}}, MS_{\text{residual}}$$

4.10 Key formulas of the chapter

Below are the main formulas introduced in this chapter: try to go through them and understand what they mean.

$$\hat{Y} = a + bX$$

with

$$b = \frac{SCP_{XY}}{SS_X} = \frac{\hat{\sigma}_Y}{\hat{\sigma}_X} r_{XY}$$

and

$$a = M_Y - bM_X$$

$$SS_{\text{total}} = \sum_s (Y_s - M_Y)^2$$

$$SS_{\text{regression}} = \sum_s (\hat{Y}_s - M_Y)^2$$

$$SS_{\text{residual}} = \sum_s (Y_s - \hat{Y}_s)^2$$

$$SS_{\text{total}} = SS_{\text{regression}} + SS_{\text{residual}}$$

$$\sum_s (Y_s - M_Y)^2 = \sum_s (\hat{Y}_s - M_Y)^2 + \sum_s (Y_s - \hat{Y}_s)^2$$

$$r_{XY}^2 = \frac{SS_{\text{regression}}}{SS_{\text{total}}}$$

$$df_{\text{regression}} = 1$$

and

$$df_{\text{residual}} = S - 2$$

$$\hat{\sigma}_{\text{regression}}^2 = MS_{\text{regression}} = \frac{SS_{\text{regression}}}{df_{\text{regression}}} = \frac{SS_{\text{regression}}}{1} = SS_{\text{regression}}$$

$$\hat{\sigma}_{\text{residual}}^2 = MS_{\text{residual}} = \frac{SS_{\text{residual}}}{df_{\text{residual}}} = \frac{SS_{\text{residual}}}{S - 2}$$

$$F = \frac{\hat{\sigma}_{\text{regression}}^2}{\hat{\sigma}_{\text{residual}}^2} = \frac{MS_{\text{regression}}}{MS_{\text{residual}}}$$

4.11 Key questions of the chapter

Below are some questions about the content of this chapter. All the answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- ★ What is the purpose of regression analysis?
- ★ In which cases can we apply regression analysis?
- ★ What is the main difference between correlation and regression analysis?
- ★ What are the different components of a given subject score?
- ★ What is the rationale behind the F ratio?

5

Orthogonal multiple regression

5.1 Introduction

In the previous chapter we learned how to predict a dependent variable from a single independent variable. In this chapter, we extend this approach to the prediction of a dependent variable from two independent variables. This generalization of the simple regression analysis is called *multiple regression analysis* (it is sometimes written MRA for short) or multiple linear regression analysis (MLR for short).

Remember that when we were trying to predict a dependent variable, denoted Y , from a single independent variable, denoted X , we expressed Y as a linear function of X . This was equivalent to finding the line that would best describe the linear relationship between Y and X . This line was called the *regression line*. The quality of prediction of this line was estimated by computing the sum of the squared deviations from the line. Here we show that trying to predict a dependent variable Y from two independent variables denoted X and T , respectively, amounts to finding the plane that will best describe the linear¹ relationship that can be used to predict the dependent variable from the two independent variables. This plane is called the *regression plane* of Y on X and T .

In this chapter we examine the case of *uncorrelated* independent variables. Remember that two independent variables are uncorrelated if their coefficient of correlation is equal to zero. A synonym of uncorrelated variables, which we shall use in this book, is the term *orthogonal* variables. As we have noted in Chapter 4, in the framework of regression analysis, the independent variables are *controlled* by the experimenters. In general, when designing experiments with several independent variables, experimenters use orthogonal variables whenever possible (how to create designs with orthogonal variables is developed in Chapter 15 on experimental design). A most desirable feature of orthogonal independent variables is that they are free of confound (see Chapter 1).

¹ Although there is now more than one independent variable, the relationship is still termed linear (instead of say *planar*), because it is a sum of weighted components, and the tradition in mathematics is to call such a sum a *linear* combination of these components.

As an additional bonus, orthogonal independent variables facilitate both the computation and the interpretation of the analysis.

In what follows, we shall derive the general formulas for multiple regression analyses, and apply them to an experiment with orthogonal independent variables. We shall also look at the simplification resulting from the orthogonality of the independent variables.

5.2 Generalities

5.2.1 The equation of a plane

The notion of a linear equation can easily be generalized from a two-dimensional 2D space to a 3D space. Remember that, in the previous chapter, we expressed the relationship between two variables X and Y as a line in a 2D space in which the first dimension is equal to X and the second dimension is equal to Y . The equation of the line is given by

$$Y = a + bX$$

where a represents the intercept (i.e. the Y value corresponding to $X = 0$) and b represents the slope of the line (i.e. the rate of change in Y corresponding to a change of one unit in X). Similarly, we can express the linear relationship between the dependent variable Y and the independent variables X and T , as a plane in the 3D space formed by the three variables (see Figure 5.1 for an illustration). The equation of the plane is simply obtained by adding a dimension to the equation of a line:

$$Y = a + bX + cT,$$

where

- a represents the intercept or value of Y obtained when both X and T are equal to zero.

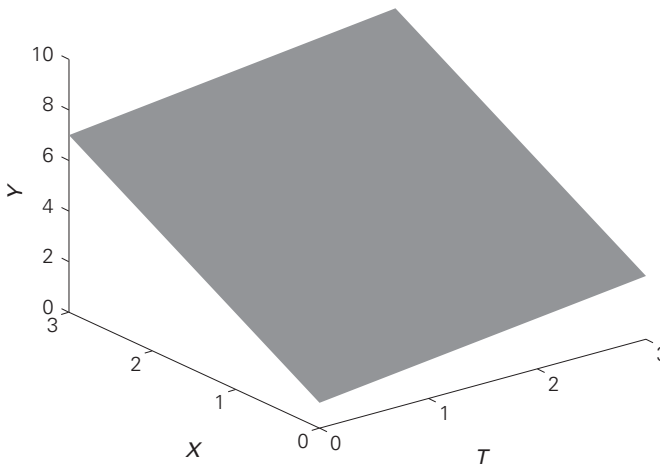


Figure 5.1 Example of a plane in the three-dimensional space formed by the two independent variables X and T , and the dependent variable Y . The plane separates the space into two subspaces, the first subspace being composed of all the points above the plane, and the second subspace being composed of all the points below the plane.

X	0	1	1	1	2	2	2	3	3	3
T	0	1	2	3	1	2	3	1	2	3
Y	1	3.5	4	4.5	5.5	6	6.5	7.5	8	8.5

Table 5.1 A set of data which can be used to draw a perfect plane (plotted in Figure 5.3).

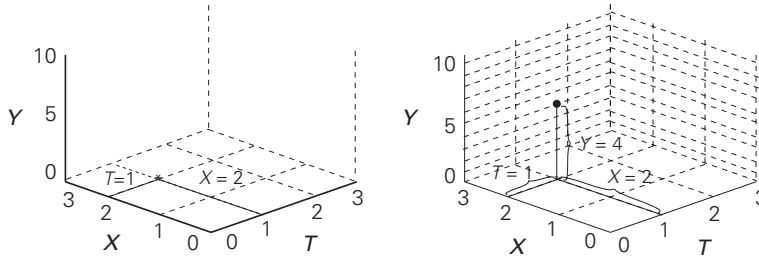


Figure 5.2 How to plot a data point in a 3D space.

- b represents the slope in the direction of X . It is the rate of change in Y corresponding to a change of one unit in X . The value of b is obtained by dividing the change in Y by the corresponding change in X :

$$b = \frac{\Delta_Y}{\Delta_X}.$$

- c represents the slope in the direction of T . It is the rate of change in Y corresponding to a change of one unit in T . The value of c is obtained by dividing the change in Y by the corresponding change in T :

$$c = \frac{\Delta_Y}{\Delta_T}.$$

In other words, we will say that the dependent variable Y is a *linear function* of the independent variables X and T with parameters a , b , and c .

5.2.2 Example of a perfect plane

Suppose that we have run an experiment in which we manipulated two independent variables X and T and recorded one dependent variable Y . The results of this hypothetical experiment are given in Table 5.1.

As you have probably noticed, it is not easy to ‘see’ the relationship between the independent variables and the dependent variable from this table. To have a clearer representation of the data we can plot them as a three-dimensional scatter plot (‘3D plot’) in which the vertical axis represents the dependent variable, and the two other (horizontal) axes represent the independent variables. Each element in the data set is represented as a point whose coordinates correspond to its X , T , and Y values. For example, Figure 5.2 shows how to draw a point with $X = 2$, $T = 1$, and $Y = 4$ coordinates. The first step is to find the projection of the point on to the 2D space made by the two independent variables X and T . As illustrated in the left panel of Figure 5.2 this is done by:

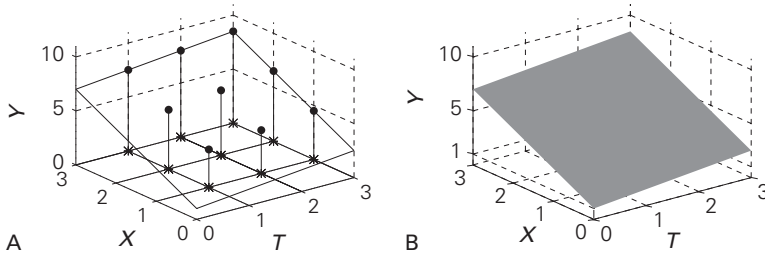


Figure 5.3 Plot of the data from Table 5.1. The equation of the plane is $Y = 1 + 2X + 0.5T$.

- locating the value 1 on the T axis and drawing a line parallel to the X axis passing through this point.
- locating the value 2 on the X axis and drawing a line parallel to the T axis passing through this point.
- marking the intersection between the two lines with a star.

The 3D projection is then obtained by

- drawing from the star a line parallel to the Y axis.
- measuring a value of 4 units on this line.
- marking the endpoint of the segment line with a black dot.

The black dot represents the projection of the data point in the 3D space. The length of the line between the star and the black dot gives the value of Y . The star indicates the experimental group (i.e. $X = 2$ and $T = 1$) from which the data point was extracted.

Figure 5.3 shows the 3D plot of the data presented in Table 5.1. As you can see, all the data points lie on a single plane. This plane is represented in gray on the right panel of Figure 5.3. To describe this plane more precisely we need to find its intercept, denoted a , and its slope in the X and T directions, called b and c respectively.

In order to find the intercept we just need to look at the value of Y when X and T are both equal to zero. From Table 5.1 we can read that the intercept is equal to 1. To find the slope of the plane for the X dimension, we need to take two values of X and the corresponding values of Y for a constant value of T . For example, we can use the value of Y for $X = 1$ and $X = 2$ when $T = 1$. From Table 5.1 we can see that for $T = 1$ and $X = 1$, $Y = 3.5$ and for $T = 1$ and $X = 2$, $Y = 5.5$. We can thus write

$$b = \frac{\Delta_Y}{\Delta_X} = \frac{Y_2 - Y_1}{X_2 - X_1} = \frac{5.5 - 3.5}{2 - 1} = 2.$$

Similarly, to find the slope of the plane for the T dimension, we need to take two values of T and the corresponding values of Y for a constant value of X . For example, we can use the value of Y for $T = 1$ and $T = 2$ when $X = 1$. From Table 5.1 we can see that for $X = 1$ and $T = 1$, $Y = 3.5$ and for $X = 1$ and $T = 2$, $Y = 4$. We can thus write

$$c = \frac{\Delta_Y}{\Delta_T} = \frac{Y_2 - Y_1}{T_2 - T_1} = \frac{4 - 3.5}{2 - 1} = 0.5.$$

It suffices now to plug the values of a , b , and c into the general equation of a plane to find the equation of the plane displayed in the left panel of Figure 5.3:

$$\begin{aligned} Y &= a + bX + cT \\ &= 1 + 2X + 0.5T. \end{aligned} \quad (5.1)$$

5.2.3 An example: retroactive interference

As was the case for the 2D space presented in the previous chapter, it is very rare, when dealing with real data, that they fall nicely on a plane. To illustrate what generally happens with real data, we present a hypothetical replication of an experiment on retroactive interference originally designed by Slamecka (1960). The term retroactive interference expresses the idea that what we learn now can make us forget what we have learned earlier. The general paradigm used to test the effect of retroactive interference is as follows. Experimental subjects are asked to learn a first list of words, and then they are asked to learn a second list of words. After they have learned the second list, the subjects are asked to recall the first list they learned. The number of words recalled by these subjects is then compared with the number of words recalled by control subjects who learned only the first list of words. Results, in general, show that learning a second list impairs the recall of the first list (i.e. experimental subjects recall fewer words than control subjects).

In Slamecka’s experiment subjects had to learn complex sentences such as

Communicators can exercise latitude in specifying meaning however they choose provided that such definition corresponds somewhat closely to regular usage.

The sentences were presented to the subjects two, four, or eight times (this is the first independent variable). In what follows we will refer to this variable as the *number of learning trials* or *X*. The subjects were then asked to learn a second series of sentences. This second series was again presented two, four, or eight times (this is the second independent variable). In what follows we will refer to this variable as the *number of interpolated lists* or *T*. After the second learning session, the subjects were asked to recall the first sentences presented. For each subject, the number of words correctly recalled was recorded (this is the dependent variable). We will refer to the dependent variable as *Y*.

In our replication of Slamecka’s experiment, we used a total of 18 subjects, two in each of the nine experimental conditions. The results of this hypothetical replication are presented in Table 5.2. In this replication we were interested in evaluating how well the two independent variables ‘number of learning trials’ and ‘number of interpolated lists’ predict the dependent variable ‘number of words correctly recalled’. A first way of examining the quality of prediction of the two independent variables is to plot the dependent variable against the two independent variables. This will allow us to visually estimate the linear relationship between the three variables. If there is a strong linear relationship (i.e. if the independent variables are good

Number of learning trials (X)	Number of interpolated lists (T)		
	2	4	8
2	35	21	6
	39	31	8
4	40	34	18
	52	42	26
8	61	58	46
	73	66	52

Table 5.2 Results of a hypothetical replication of Slamecka (1960)’s retroactive interference experiment.

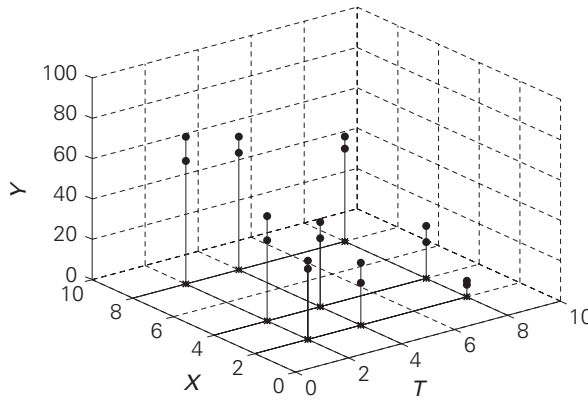


Figure 5.4 3D representation of the data presented in Table 5.2. X stands for the number of learning trials. T stands for the number of interpolated lists. Y stands for the number of words correctly recalled.

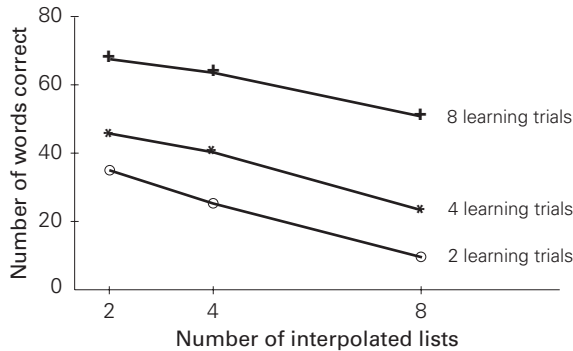


Figure 5.5 2D representation of the data presented in Table 5.2.

predictors of the dependent variable) all the points in the scatter plot should be positioned on a plane in the 3D space.

Figure 5.4 displays the 3D plot of the number of words correctly recalled as a function of both the number of learning trials, and the number of interpolated lists. In this figure, each data point is represented by a black circle, and its projection onto the space created by the two independent variables by a star. The projection from the star to the axes representing each independent variable indicates the level of the independent variable. The length of the line between the star and the black circle represents the number of words recalled. The longer the line, the better the recall.

Figure 5.4 shows that, although all the points do not perfectly lie on a single plane, they give the impression of being relatively clustered around a plane. This plane (to be found later) is called the *regression plane*. Even though we do not know its exact equation, we can guess from Figure 5.4 that the regression plane has a positive slope in the X direction and a negative slope in the T direction. In other words, it seems that there is a positive relationship between the number of words recalled (Y) and the number of learning trials (X), as well as a negative relationship between the number of words recalled and the number of interpolated lists (T).

The larger the number of learning trials and the smaller the number of interpolated lists, the better the recall.

Another more popular way of representing this type of result is shown in Figure 5.5. In this figure we use a 2D representation in which the vertical axis represents the values of the dependent variable and the horizontal axis the values of one of the independent variables (variable T for our example). The levels of the second independent variable are indicated by using different markers. For our example, the three levels of variable X are marked by an open circle 'o' (2 learning trials), a star '*' (4 learning trials), and a plus sign '+' (8 learning trials), respectively. In Figure 5.5 we can clearly see the strong positive relationship between the number of trials and the number of words recalled (i.e. the line connecting the plus signs is clearly above the line connecting the stars, which itself is above the line connecting the open circles), and the negative relationship between the number of interpolated lists and the number of words recalled.

5.3 The regression plane is the 'best-fit' plane

The idea of predicting the values of the dependent variable from their positions on the regression plane is the 2-independent variable analog of the 1-independent variable prediction of the values of the dependent variable from their position on the regression line (don't hesitate to refresh your memory in Chapter 4, Section 4.3, page 67). In the 1-independent variable case, we assumed that when the regression model is perfect, all the values of the dependent variable are positioned on the regression line. Likewise, in the 2-independent variable case, we assume that when the regression model is perfect, all the values of the dependent variable are positioned on the regression plane. The deviations from the data points to this plane are attributed, once again, to the measurement error. As in the 1-independent variable case, the error can affect only the dependent variable. The measurement error cannot affect the independent variables because we *control* them (using the standard statistical jargon, we would say the independent variables are *fixed*).

With that in mind, we can now define the regression plane as the plane that describes best the shape of the data. Intuitively, this plane should be as much 'in the middle of the data points' as possible. As we have already done for the regression line, we will formalize this intuition by a *least squares* criterion. Specifically, we define the regression plane as the plane which minimizes the sum of the squared deviations from the estimated values to the observed values. Formally, we say that the predicted values of the dependent variable should be obtained as

$$\hat{Y} = a + bX + cT \quad (5.2)$$

(compare with Equation 4.2, page 67).

The difference between what we observe and the prediction

$$Y - \hat{Y} \quad (5.3)$$

is considered as the error of measurement. The term $(Y - \hat{Y})$ is also called the *residual of the prediction* or simply the *residual*.

For the regression plane to be as much as possible in 'the middle of the data points', the magnitude of the differences between the observed values and the predicted values should be as small as possible. Precisely, we want to find the values of a , b , and c such that the sum of

the squared deviations from the plane that they determine will be as small as possible. With a formula, we say that we are looking for the parameters a , b , and c such that

$$\min \sum_s (Y_s - \hat{Y}_s)^2 = \sum_s [Y_s - (a + bX_s + cT_s)]^2 \quad (5.4)$$

(cf. Equation 4.4 on p. 67).

We show later on in a digression (which you can skip if you feel comfortable with accepting these results at face value) that the values of a , b , and c defining the regression plane are:

- For the slope parameter in the X direction

$$b = \frac{SS_T SCP_{XY} - SCP_{XT} SCP_{TY}}{SS_X SS_T - (SCP_{XT})^2}. \quad (5.5)$$

where

SCP_{XY} is the sum of cross-products of X and Y

$$SCP_{XY} = \sum (X - M_X)(Y - M_Y),$$

SCP_{XT} is the sum of cross-products of X and T

$$SCP_{XT} = \sum (X - M_X)(T - M_T),$$

SCP_{TY} is the sum of cross-products of T and Y

$$SCP_{TY} = \sum (T - M_T)(Y - M_Y),$$

SS_X is the sum of squares of X

$$SS_X = \sum (X - M_X)^2,$$

and

SS_T is the sum of squares of T

$$SS_T = \sum (T - M_T)^2.$$

- For the slope parameter in the T direction

$$c = \frac{SS_X SCP_{TY} - SCP_{XT} SCP_{XY}}{SS_X SS_T - (SCP_{XT})^2}. \quad (5.6)$$

- For the intercept

$$a = M_Y - bM_X - cM_T. \quad (5.7)$$

If you compare the computation of these parameters with the computation of the intercept and slope for the regression line, you will note their general similarity, but also that the computation of the parameters of the regression plane is a bit more complex than the computation of the parameters of the regression line (cf. Equation 4.5, page 68).

5.4 Back to the example: retroactive interference

Table 5.3 gives the quantities needed to compute the values of the intercept and the slopes for the retroactive interference example. If we use these values to determine the equation plane for the retroactive interference example, we obtain the following values:

<i>Y</i>	<i>y</i>	<i>y</i> ²	<i>X</i>	<i>x</i>	<i>x</i> ²	<i>T</i>	<i>t</i>	<i>t</i> ²	<i>yx</i>	<i>yt</i>	<i>xt</i>
35.00	−4.33	18.78	2.00	−2.67	7.11	2.00	−2.67	7.11	11.56	11.56	7.11
21.00	−18.33	336.11	2.00	−2.67	7.11	4.00	−0.67	0.44	48.89	12.22	1.78
6.00	−33.33	1,111.11	2.00	−2.67	7.11	8.00	3.33	11.11	88.89	−111.11	−8.89
40.00	0.67	0.44	4.00	−0.67	0.44	2.00	−2.67	7.11	−0.44	−1.78	1.78
34.00	−5.33	28.44	4.00	−0.67	0.44	4.00	−0.67	0.44	3.56	3.56	0.44
18.00	−21.33	455.11	4.00	−0.67	0.44	8.00	3.33	11.11	14.22	−71.11	−2.22
61.00	21.67	469.44	8.00	3.33	11.11	2.00	−2.67	7.11	72.22	−57.78	−8.89
58.00	18.67	348.44	8.00	3.33	11.11	4.00	−0.67	0.44	62.22	−12.44	−2.22
46.00	6.67	44.44	8.00	3.33	11.11	8.00	3.33	11.11	22.22	22.22	11.11
39.00	−0.33	0.11	2.00	−2.67	7.11	2.00	−2.67	7.11	0.89	0.89	7.11
31.00	−8.33	69.44	2.00	−2.67	7.11	4.00	−0.67	0.44	22.22	5.56	1.78
8.00	−31.33	981.78	2.00	−2.67	7.11	8.00	3.33	11.11	83.56	−104.44	−8.89
52.00	12.67	160.44	4.00	−0.67	0.44	2.00	−2.67	7.11	−0.84	−33.83	1.79
42.00	2.67	7.11	4.00	−0.67	0.44	4.00	−0.67	0.44	−1.78	−1.78	0.44
26.00	−13.33	177.78	4.00	−0.67	0.44	8.00	3.33	11.11	8.89	−44.44	−2.22
73.00	33.67	1,133.44	8.00	3.33	11.11	2.00	−2.67	7.11	112.22	−89.78	−8.89
66.00	26.67	711.11	8.00	3.33	11.11	4.00	−0.67	0.44	88.89	−17.78	−2.22
52.00	12.67	160.44	8.00	3.33	11.11	8.00	3.33	11.11	42.22	42.22	11.11
Σ	708.00	0	6,214.00	84.00	0	112.00	84.00	112.00	672.00	−448.00	0
	SS _Y			SS _X			SS _T		SCP _{XY}	SCP _{TY}	SCP _{XT}

Table 5.3 The different quantities needed to compute the values of the parameters *a*, *b*, and *c* of a regression plane defined by the equation $\hat{Y} = a + bX + cT$. The following abbreviations are used to label the columns of the table: $t = (T - M_T)$, $x = (X - M_X)$, $y = (Y - M_Y)$.

- Slope for X :

$$\begin{aligned}
 b &= \frac{SS_T SCP_{XY} - SCP_{XT} SCP_{TY}}{SS_X SS_T - (SCP_{XT})^2} \\
 &= \frac{112 \times 672 - 0 \times (-448)}{112 \times 112 - 0^2} \\
 &= 6.00 .
 \end{aligned} \tag{5.8}$$

- Slope for T :

$$\begin{aligned}
 c &= \frac{SS_X SCP_{TY} - SCP_{XT} SCP_{XY}}{SS_X SS_T - (SCP_{XT})^2} \\
 &= \frac{112 \times (-448) - 0 \times 672}{112 \times 112 - 0^2} \\
 &= -4.00 .
 \end{aligned} \tag{5.9}$$

Did you notice that SCP_{XT} was nice enough to be equal to zero? This is always the case when X and T are orthogonal. This happens because, when X and T are orthogonal, their coefficient of correlation is zero, and therefore SCP_{XT} is zero.²

- Intercept³

$$\begin{aligned}
 a &= M_Y - bM_X - cM_T \\
 &= 39.33 - (6 \times 4.67) - (-4 \times 4.67) \\
 &= 30.00 .
 \end{aligned} \tag{5.10}$$

We can now write the equation predicting the values of the dependent variable (Y : number of words correctly recalled) from the independent variables (X : number of learning trials; and T : number of interpolated lists):

$$\begin{aligned}
 \hat{Y} &= a + bX + cT \\
 &= 30 + 6X - 4T
 \end{aligned} \tag{5.11}$$

or

$$\begin{aligned}
 \text{no. of words recalled} &= 30 + 6 \times (\text{no. of learning trials}) \\
 &\quad - 4 \times (\text{no. of interpolated lists}) .
 \end{aligned}$$

Figure 5.6 shows the original data as a 3D plot with the regression plane superimposed. From the values of the intercept and the slopes, we can see that adding one learning trial makes subjects remember 6 more words and that each interpolated list makes subjects forget 4 words (i.e. they remember ‘4 words less’).

² If this is not clear, re-read Chapter 2.

³ If you use a pocket calculator, you may find a value slightly different from this one (e.g. something like 29.99). We found the value of 30 because we keep more decimal places than we report in the text.

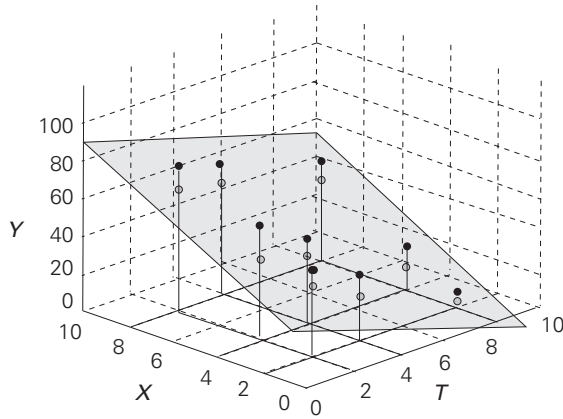


Figure 5.6 Plot of the results of a replication of the experiment of Slamecka's (1960) retroactive interference experiment. Each data point represents the number of words correctly recalled (Y , the dependent variable) by a subject having learned 2, 4, or 8 times a list (X , first independent variable) with 2, 4 or 8 interpolated lists learned before recall (T , the second independent variable). The regression plane to predict Y from X and T is given by the equation: $\hat{Y} = a + bX + cT = 30 + 6X - 4T$. Data from Table 5.2.

5.5 How to evaluate the quality of the prediction

By definition, we know that the prediction $\hat{Y}$ of the dependent variable computed from the two independent variables is the best possible one. The problem, however, remains to evaluate the quality of this prediction: is it very good or just mediocre?

Remember that in the previous chapter we saw that this can be done by first computing the coefficient of correlation between the predicted values $\hat{Y}$, and the actual values Y , and then evaluating the probability of obtaining this value of r by chance alone using an F ratio. The same approach can be used here. There are two possible notations. The first one expresses that we are correlating the predicted values $\hat{Y}$ with the actual values Y . This coefficient is noted $r_{\hat{Y}, Y}^2$. It is read as 'r square between Y -hat and Y '. The second notation is a variant of the notation used in the previous chapter, the only difference is that we will use the squared coefficient of correlation instead of the coefficient of correlation and that we will use the uppercase letter R instead of the lower case r used in the 1-independent variable case. This coefficient is denoted $R_{Y.XT}^2$ (read ' R squared of Y explained by X and T ', or R^2 for short). It is called a *multiple* correlation coefficient because it expresses the strength of the relationship between *several* independent variables and one dependent variable.

Table 5.4 gives the quantities needed to compute $R_{Y.XT}^2$ for our example. Using these values, we obtain:

$$\begin{aligned}
 R_{Y.XT}^2 &= \frac{(SCP_{\hat{Y}Y})^2}{SS_{\hat{Y}}SS_Y} \\
 &= \frac{5,824^2}{5,824 \times 6,214} \\
 &= .9372.
 \end{aligned} \tag{5.12}$$

To evaluate if the regression analysis gives a better prediction than chance alone would give, the procedure is the same as for the standard coefficient of correlation. The first step

Y	y	y^2	$\hat{Y}$	$\hat{y}$	$\hat{y}^2$	$\hat{y} \times y$
35.00	-4.33	18.78	34.00	-5.33	28.44	23.11
21.00	-18.33	336.11	26.00	-13.33	177.78	244.44
6.00	-33.33	1,111.11	10.00	-29.33	860.44	977.78
40.00	0.67	0.44	46.00	6.67	44.44	4.44
34.00	-5.33	28.44	38.00	-1.33	1.78	7.11
18.00	-21.33	455.11	22.00	-17.33	300.44	369.78
61.00	21.67	469.44	70.00	30.67	940.44	664.44
58.00	18.67	348.44	62.00	22.67	513.78	423.11
46.00	6.67	44.44	46.00	6.67	44.44	44.44
39.00	-0.33	0.11	34.00	-5.33	28.44	1.78
31.00	-8.33	69.44	26.00	-13.33	177.78	111.11
8.00	-31.33	981.78	10.00	-29.33	860.44	919.11
52.00	12.67	160.44	46.00	6.67	44.44	84.44
42.00	2.67	7.11	38.00	-1.33	1.78	-3.56
26.00	-13.33	177.78	22.00	-17.33	300.44	231.11
73.00	33.67	1,133.44	70.00	30.67	940.44	1,032.44
66.00	26.67	711.11	62.00	22.67	513.78	604.44
52.00	12.67	160.44	46.00	6.67	44.44	84.44
708.00	0	6,214.00			5,824.00	5,824.00
		SS_Y			$SS_{\hat{Y}}$	$SCP_{\hat{Y}}$

Table 5.4 Quantities needed to compute the coefficient of correlation between Y and $\hat{Y}$ (data from Table 5.3). The following abbreviations are used: $y = Y - M_Y$, $\hat{y} = \hat{Y} - M_{\hat{Y}}$. (Totals may not be exact due to rounding errors.)

is to compute an F ratio and then to check for its significance. To compute the F ratio we will use a generalization of the formula we used in the 1-independent variable case. This generalization takes into account the number of independent variables used to predict the dependent variable. It is given by this formula:

$$F_{Y.XT} = \frac{R_{Y.XT}^2}{1 - R_{Y.XT}^2} \times \frac{S - K - 1}{K} \quad (5.13)$$

where K is the number of independent variables used to predict the dependent variable. Under the null hypothesis, this F ratio follows a Fisher distribution with $\nu_1 = K$ and $\nu_2 = S - K - 1$ degrees of freedom.

For our example, $K = 2$ and hence, the $F_{Y.XT}$ ratio is obtained as

$$F_{Y.XT} = \frac{R_{Y.XT}^2}{1 - R_{Y.XT}^2} \times \frac{S - 3}{2}$$

which gives

$$\begin{aligned} F_{Y.XT} &= \frac{.9372}{1 - 0.9372} \times \frac{18 - 3}{2} \\ &= 111.93. \end{aligned}$$

We can then follow the standard test procedure to see if this value can be attributed to chance or if it reflects the existence of a linear relationship between the independent variables and the dependent variable. Suppose that we decided upon a value of $\alpha = .05$, the critical value

of F for $\nu_1 = 2$ and $\nu_2 = S - K - 1 = 15$ is equal to 3.68. Since the computed value of $F = 111.93$ is larger than the critical value, we can reject the null hypothesis and accept the alternative hypothesis stating that there is a linear relationship between the two independent variables (number of learning trials and number of interpolated lists) and the dependent variable (number of words recalled).

5.5.1 How to evaluate the importance of each independent variable in the prediction

In the previous section we have seen that, on the whole, 93.72% of the variability in the Y scores can be explained by the existence of a linear relationship between the dependent variable and the two independent variables. The next question is how much each independent variable contributes to this prediction. This can be found by computing a separate coefficient of correlation for each of the independent variables. Note that we are now using a lower case r to represent the coefficient of correlation because only one variable is correlated with $\hat{Y}$.

- The quality of prediction of variable X is estimated by computing the coefficient of correlation between $\hat{Y}$ and X :

$$r_{\hat{Y}.X}^2 = \frac{(SCP_{\hat{Y}X})^2}{SS_{\hat{Y}}SS_X} . \quad (5.14)$$

- The quality of prediction of variable T is estimated by computing the coefficient of correlation between $\hat{Y}$ and T :

$$r_{\hat{Y}.T}^2 = \frac{(SCP_{\hat{Y}T})^2}{SS_{\hat{Y}}SS_T} . \quad (5.15)$$

Table 5.5 gives the quantities needed to compute the coefficients of correlation $r_{\hat{Y}.X}^2$ and $r_{\hat{Y}.T}^2$ for our example. Using these values we find:

- quality of prediction of X

$$r_{\hat{Y}.X}^2 = \frac{672^2}{5824 \times 112} = .6923 , \quad (5.16)$$

which indicates that the variable X , number of learning trials, accounts for 69.23% of the prediction of the dependent variable.

- quality of prediction of T

$$r_{\hat{Y}.T}^2 = \frac{-448^2}{5824 \times 112} = .3077 , \quad (5.17)$$

which indicates that the variable T , number of interpolated lists, accounts for 30.77% of the prediction of the dependent variable.

Note in passing that

$$r_{\hat{Y}.X}^2 + r_{\hat{Y}.T}^2 = .6923 + .3077 = 1 .$$

This shows that the coefficients $r_{\hat{Y}.X}^2$ and $r_{\hat{Y}.T}^2$ can be interpreted as the proportion of the *prediction* of the dependent variable that can be accounted for by a given independent variable.

X	x	x^2	T	t	t^2	$\hat{Y}$	$\hat{y}$	$\hat{y}^2$	$x \times \hat{y}$	$t \times \hat{y}$
2	-2.67	7.11	2	-2.67	7.11	34	-5.33	28.44	14.23	14.23
2	-2.67	7.11	4	-0.67	0.44	26	-13.33	77.78	35.59	8.93
2	-2.67	7.11	8	3.33	11.11	10	-29.33	860.44	78.31	-97.66
4	-0.67	0.44	2	-2.67	7.11	46	6.67	44.44	-4.47	-17.80
4	-0.67	0.44	4	-0.67	0.44	38	-1.33	1.78	0.89	0.89
4	-0.67	0.44	8	3.33	11.11	22	-17.33	300.44	11.61	-57.70
8	3.33	11.11	2	-2.67	7.11	70	30.67	940.44	102.13	-81.89
8	3.33	11.11	4	-0.67	0.44	62	22.67	513.78	75.50	-15.18
8	3.33	11.11	8	3.33	11.11	46	6.67	44.44	22.21	22.21
2	-2.67	7.11	2	-2.67	7.11	34	-5.33	28.44	14.23	14.23
2	-2.67	7.11	4	-0.67	0.44	26	-13.33	177.78	35.59	8.93
2	-2.67	7.11	8	3.33	11.11	10	-29.33	860.44	78.31	-97.67
4	-0.67	0.44	2	-2.67	7.11	46	6.67	44.44	-4.47	-17.81
4	-0.67	0.44	4	-0.67	0.44	38	-1.33	1.78	0.89	0.89
4	-0.67	0.44	8	3.33	11.11	22	-17.33	300.44	11.61	57.11
8	3.33	11.11	2	-2.67	7.11	70	30.67	940.44	102.13	-81.89
8	3.33	11.11	4	-0.67	0.44	62	22.67	513.78	75.49	-15.19
8	3.33	11.11	8	3.33	11.11	46	6.67	44.44	22.21	22.21
		112.00			112.00			5824.00	672.00	-448.00
		SS_X			SS_T			$SS_{\hat{Y}}$	$SCP_{\hat{Y}X}$	$SCP_{\hat{Y}T}$

Table 5.5 Quantities needed to compute the coefficients of correlation between X and $\hat{Y}$ and between T and $\hat{Y}$. The following abbreviations are used: $y = Y - M_Y$, $\hat{y} = \hat{Y} - M_{\hat{Y}}$, $t = (T - M_T)$, $x = (X - M_X)$.

5.5.2 How to evaluate the importance of each independent variable for the dependent variable

In the previous section we have examined how each independent variable contributes to the *predicted value* $\hat{Y}$. We can also, alternatively, examine the contribution of each independent variable to the *dependent variable*. This is done by computing directly the square coefficients of correlation between each independent variable and the dependent variable.

- The quality of prediction of variable X is estimated by computing the coefficient of correlation between Y and X :

$$r_{Y \cdot X}^2 = \frac{(SCP_{YX})^2}{SS_Y SS_X}.$$

- The quality of prediction of variable T is estimated by computing the coefficient of correlation between Y and T :

$$r_{Y \cdot T}^2 = \frac{(SCP_{YT})^2}{SS_Y SS_T}.$$

Table 5.3 gives the quantities needed to compute the square coefficients of correlation $r_{Y \cdot X}^2$ and $r_{Y \cdot T}^2$ for our example. Using these values we find:

- Quality of prediction of X

$$r_{Y \cdot X}^2 = \frac{672^2}{6214 \times 112} = .6489,$$

which indicates that the variable X , number of learning trials, accounts for 64.89% of the dependent variable.

- Quality of prediction of T

$$r_{Y.T}^2 = \frac{-448^2}{6214 \times 112} = .2884 ,$$

which indicates that the variable T , number of interpolating lists, accounts for 28.84% of the dependent variable.

Note in passing that

$$r_{Y.X}^2 + r_{Y.T}^2 = .6489 + .2884 = .9372 = R_{Y.XT}^2 .$$

This shows that the coefficients $r_{Y.X}^2$ and $r_{Y.T}^2$ can be interpreted as the proportion of the variance of the dependent variable that can be accounted for by a given independent variable. It is important to stress that *this additive property (as well as the additive property reported in the previous section) is true only for orthogonal independent variables*. This is explored further in the following sections.

5.5.3 From the $r_{\hat{Y}}$ coefficients to the r_Y coefficients

It can be shown with some elementary algebraic manipulations⁴ that the computation of the coefficients of correlation of the independent variables with $\hat{Y}$ are related to the computation of the coefficients of correlation between the independent variables and Y .

Specifically,

$$r_{Y.X}^2 = r_{\hat{Y}.X}^2 \times R_{Y.XT}^2 \quad (5.18)$$

and

$$r_{Y.T}^2 = r_{\hat{Y}.T}^2 \times R_{Y.XT}^2 . \quad (5.19)$$

For example,

$$r_{Y.X}^2 = r_{\hat{Y}.X}^2 \times R_{Y.XT}^2 = .6923 \times .9372 \approx .6489 , \quad (5.20)$$

which corresponds (within rounding errors) to the value found previously.

5.6 *F* tests for the simple coefficients of correlation

To evaluate if the prediction of the dependent variable by a given independent variable is significantly different from zero, we can compute, as usual, an *F* test. Looking at formulas 5.18 and 5.19 shows that when one coefficient is zero, the other coefficient is also zero (because they are related by a multiplicative non-zero constant). Therefore, it suffices to test one of the coefficients. The tradition is to test the coefficients of the form r_Y .

We can now compute two *F* tests (one for each independent variable). They have a form already familiar (cf. Equation 5.13, page 97).

⁴ left to the audacious reader as an exercise...

For the independent variable X , we compute the ratio

$$F_{Y \cdot X} = \frac{r_{Y \cdot X}^2}{1 - R_{Y \cdot XT}^2} \times (S - K - 1), \quad (5.21)$$

where K is the number of independent variables used to predict the dependent variable (in our example $K = 2$). Under the null hypothesis, this F ratio follows a Fisher distribution with $\nu_1 = 1$ and $\nu_2 = S - K - 1$ degrees of freedom.

For our example, $K = 2$ and hence, the $F_{Y \cdot X}$ ratio is obtained as

$$F_{Y \cdot X} = \frac{r_{Y \cdot X}^2}{1 - R_{Y \cdot XT}^2} \times (S - 3) = \frac{.6489}{1 - .9372} \times 15 = 154.90.$$

We can now follow the standard test procedure to see if this value can be attributed to chance, or if it reflects the existence of a linear relationship between the independent variable X and the dependent variable. Suppose that we decided upon a value of $\alpha = .05$, the critical value of F for $\nu_1 = 1$ and $\nu_2 = S - K - 1 = 15$ is equal to 4.54. Since the computed value of $F = 154.90$ is larger than the critical value, we can reject the null hypothesis and accept the alternative hypothesis stating that there is a linear relationship between the independent variable X (number of learning trials) and the dependent variable (number of words recalled).

Likewise, for the independent variable T , we compute the ratio

$$F_{Y \cdot T} = \frac{r_{Y \cdot T}^2}{1 - R_{Y \cdot XT}^2} \times (S - K - 1), \quad (5.22)$$

where K is the number of independent variables used to predict the dependent variable. Under the null hypothesis, this F ratio follows a Fisher distribution with $\nu_1 = 1$ and $\nu_2 = S - K - 1$ degrees of freedom.

For our example, $K = 2$ and hence the $F_{Y \cdot T}$ ratio is obtained as

$$F_{Y \cdot T} = \frac{r_{Y \cdot T}^2}{1 - R_{Y \cdot XT}^2} \times (S - 3) = \frac{.2884}{1 - .9372} \times 15 = 68.88.$$

We can now follow the standard test procedure to see if this value can be attributed to chance, or if it reflects the existence of a linear relationship between the independent variable T and the dependent variable. Suppose that we decided upon a value of $\alpha = .05$, the critical value of F for $\nu_1 = 1$ and $\nu_2 = S - K - 1 = 15$ is equal to 4.54. Since the computed value of $F = 68.88$ is larger than the critical value, we can reject the null hypothesis and accept the alternative hypothesis stating that there is a linear relationship between the independent variable T (number of interpolated lists) and the dependent variable (number of words recalled).

5.7 Partitioning the sums of squares

In the 1-independent variable case we saw that the deviation of a Y score from the mean of the Y scores could be partitioned into

- $\hat{Y} - M_Y$: deviation of the predicted score from the mean of the Y scores
- $Y - \hat{Y}$: deviation of the score from the regression line (residual),

which we summarized by the formula

$$\begin{aligned}\sum (Y - M_Y)^2 &= \sum (\hat{Y} - M_Y)^2 + \sum (Y - \hat{Y})^2 \\ SS_Y &= SS_{\text{total}} = SS_{\text{regression}} + SS_{\text{residual}}.\end{aligned}\quad (5.23)$$

This partition of the total sum of squares into sum of squares regression and sum of squares residual holds true for the 2-independent variable case. The main difference, however, is that in the 2-independent variable case, when the two independent variables are orthogonal, the sum of squares regression can be further divided into two parts: the first part reflecting the variability explained by X and the second part the variability explained by T .

5.7.1 What is a score made of?

In the retroactive interference example, we have seen that the score of a given subject can be decomposed into

- the value predicted by the number of learning trials and by the number of interpolated lists (the independent variables), and
- the residual or prediction error.

More formally we would say that a given score can be expressed as

$$\begin{aligned}Y &= \hat{Y} + (Y - \hat{Y}) \\ &= \hat{Y} + \text{error}.\end{aligned}\quad (5.24)$$

This simply says that Y is made of a predicted value plus an error term as we saw in the 1-independent variable case.

Now, if we rewrite the regression plane equation we can see that the predicted value, itself, is composed of three quantities. Let us start with the basic equation

$$\hat{Y} = a + bX + cT,$$

in which a represents the intercept and is equal to

$$a = M_Y - bM_X - cM_T.$$

If we replace a by its value in the regression plane equation, we obtain

$$\hat{Y} = M_Y - bM_X - cM_T + bX + cT.$$

Rearranging gives

$$\hat{Y} = M_Y + b(X - M_X) + c(T - M_T). \quad (5.25)$$

which shows that the predicted value $\hat{Y}$ is made of the mean of Y , the deviation of X from its mean weighted by b , and the deviation of T from its mean weighted by c .

5.7.2 The score model

By combining Equations 5.24 and 5.25, we obtain a new equation called the *score model* of Y . The score model gives the composition of a score according to the regression model. Specifically, it expresses Y as the sum of four terms:

$$Y = M_Y + b(X - M_X) + c(T - M_T) + (Y - \hat{Y}) \quad (5.26)$$

where:

- M_Y represents the mean of the scores Y ,
- $(X - M_X)$ represents the deviation of X from the mean of X ,
- $(T - M_T)$ represents the deviation of T from the mean of T ,
- $(Y - \hat{Y})$ represents the error of prediction or residual.

5.7.3 Life is simple when X and T are orthogonal: partitioning the sum of squares regression

In this section we show that, when the two independent variables X and T are orthogonal, the sum of squares of regression can be obtained as the sum of their respective sums of squares of regression, and we also show that, as a consequence, their r^2 add up to $R_{Y.XT}^2$. Let us start from Equation 5.25 with a predicted score:

$$\hat{Y} = M_Y + b(X - M_X) + c(T - M_T)$$

which we can rewrite as

$$\hat{Y} - M_Y = b(X - M_X) + c(T - M_T)$$

to obtain the equation of the deviation of a predicted score from the mean of the actual scores Y . If each of these deviations is squared and summed, we obtain

$$\begin{aligned} \sum (\hat{Y} - M_Y)^2 &= SS_{\text{regression}} \\ &= \sum [b(X - M_X) + c(T - M_T)]^2 \\ &= \sum [b^2(X - M_X)^2 + c^2(T - M_T)^2 + 2bc(X - M_X)(T - M_T)] \\ &= b^2 \sum (X - M_X)^2 + c^2 \sum (T - M_T)^2 + 2bc \sum (X - M_X)(T - M_T) \\ &= b^2 SS_X + c^2 SS_T + 2bc SCP_{XT} \end{aligned} \quad (5.27)$$

but, because X and T are uncorrelated, $SCP_{XT} = 0$ and so:

$$2bc SCP_{XT} = 0.$$

Hence, Equation 5.27 reduces to

$$SS_{\text{regression}} = b^2 SS_X + c^2 SS_T.$$

If we replace b^2 and c^2 by their expression, when X and T are orthogonal we find:

$$\begin{aligned} SS_{\text{regression}} &= \frac{(SCP_{XY})^2}{SS_X^2} SS_X + \frac{(SCP_{TY})^2}{SS_T^2} SS_T \\ &= \frac{(SCP_{XY})^2}{SS_X} + \frac{(SCP_{TY})^2}{SS_T}. \end{aligned} \quad (5.28)$$

If we denote by $r_{Y.X}^2$ (respectively $r_{Y.T}^2$) the square coefficient of correlation of Y and X (respectively Y and T), and recall that

$$r_{Y.X}^2 = \frac{(SCP_{XY})^2}{SS_X SS_Y} \quad \text{and} \quad r_{Y.T}^2 = \frac{(SCP_{TY})^2}{SS_T SS_Y},$$

we find that

$$\begin{aligned} SS_{\text{regression}} &= r_{Y \cdot X}^2 SS_Y + r_{Y \cdot T}^2 SS_Y \\ &= SS_{\text{regression of } Y \text{ on } X} + SS_{\text{regression of } Y \text{ on } T} \end{aligned} \quad (5.29)$$

The terms $SS_{\text{regression of } Y \text{ on } X}$ and $SS_{\text{regression of } Y \text{ on } T}$ are the sums of squares of the regression of Y on X and T respectively. To make reading some of the following equations smoother, we will use the abbreviations $SS_{Y \cdot X}$ and $SS_{Y \cdot T}$ for these two sums of squares.

Also, dividing Equation 5.29 by SS_Y shows that when X and T are orthogonal:

$$R_{Y \cdot XT}^2 = r_{Y \cdot X}^2 + r_{Y \cdot T}^2 .$$

Et voilà!

5.7.4 Degrees of freedom

Because we know that the point with coordinates $[M_X, M_T, M_Y]$ belongs to the regression plane, and because we need to know 3 points to draw a plane, the number of degrees of freedom for the sum of squares regression is 2:

$$df_{\text{regression}} = 2 .$$

These two degrees of freedom can be divided into

$$df_X = 1 \quad \text{and} \quad df_T = 1 .$$

For the residual sum of squares, the number of degrees of freedom is equal to the number of subjects minus the number of independent variables minus 1: $df_{\text{residual}} = S - 2 - 1 = S - 3$.

5.7.5 Mean squares

We can compute the mean squares corresponding to each of the sums of squares in the regression model. To do that, we divide the sums of squares by their respective degrees of freedom:

$$\begin{aligned} MS_{\text{residual}} &= \frac{SS_{\text{residual}}}{df_{\text{residual}}} = \frac{SS_{\text{residual}}}{S - 3} \\ MS_{\text{regression}} &= \frac{SS_{\text{regression}}}{df_{\text{regression}}} = \frac{SS_{\text{regression}}}{2} \\ MS_{Y \cdot X} &= \frac{SS_{Y \cdot X}}{df_X} = SS_{Y \cdot X} \\ MS_{Y \cdot T} &= \frac{SS_{Y \cdot T}}{df_T} = SS_{Y \cdot T} . \end{aligned}$$

Note, however, that unlike the sum of squares and the degrees of freedom, the mean squares do not add up.

5.7.6 The return of F

To evaluate the overall quality of prediction of our model we can now compute the following F ratio:

$$F_{Y \cdot XT} = \frac{MS_{\text{regression}}}{MS_{\text{residual}}}$$

If we want to evaluate the quality of prediction of each variable we can also compute

$$F_{Y \cdot X} = \frac{MS_{Y \cdot X}}{MS_{\text{residual}}}$$

and

$$F_{Y \cdot T} = \frac{MS_{Y \cdot T}}{MS_{\text{residual}}}.$$

5.7.7 Back to the example

In order to compute the F ratios, we need first to find the different sums of squares. The quantities needed to compute these values are presented in Table 5.6, and in Equations 5.14 and 5.17 (page 98). Using these quantities we find:

- $SS_{\text{regression}} = 5,824.00$
- $SS_{\text{residual}} = 390.00$
- $SS_{\text{total}} = SS_Y = 6,214.00$
- $r_{Y \cdot X}^2 = .6489$
- $r_{Y \cdot T}^2 = .2884$
- $SS_{Y \cdot X} = r_{Y \cdot X}^2 \times SS_Y = .6489 \times 6,214.00 \approx 4,032.00$
- $SS_{Y \cdot T} = r_{Y \cdot T}^2 \times SS_Y = .2884 \times 6,214.00 \approx 1,792.00$.

Note in passing that

$$SS_{\text{total}} = SS_{\text{regression}} + SS_{\text{residual}} = 5,824.00 + 390.00 = 6,214.00$$

and that

$$SS_{\text{regression}} = SS_{Y \cdot X} + SS_{Y \cdot T} = 4,032.00 + 1,792.00 = 5,824.00.$$

The following mean squares are then obtained

$$\begin{aligned} MS_{Y \cdot X} &= \frac{SS_{Y \cdot X}}{df_X} = \frac{4,032.00}{1} = 4,032.00 \\ MS_{Y \cdot T} &= \frac{SS_{Y \cdot T}}{df_T} = \frac{1,792.00}{1} = 1,792.00 \\ MS_{\text{regression}} &= \frac{SS_{\text{regression}}}{df_{\text{regression}}} = \frac{5,824.00}{2} = 2,912.00 \\ MS_{\text{residual}} &= \frac{SS_{\text{residual}}}{df_{\text{residual}}} = \frac{390.00}{15} = 26.00 \end{aligned} \quad (5.30)$$

and finally the three F ratios:

$$\begin{aligned} F_{Y \cdot X} &= \frac{MS_{Y \cdot X}}{MS_{\text{residual}}} = \frac{4,032.00}{26.00} \approx 155.08 \\ F_{Y \cdot T} &= \frac{MS_{Y \cdot T}}{MS_{\text{residual}}} = \frac{1,792.00}{26.00} = 68.92 \\ F_{Y \cdot XT} &= \frac{MS_{\text{regression}}}{MS_{\text{residual}}} = \frac{2,912.00}{26.00} = 112.00 \end{aligned} \quad (5.31)$$

which are equivalent (within rounding errors, as usual) to what we found in the previous section!

X	x	x^2	T	t	t^2	Y	y	y^2	$\hat{Y}$	$\hat{y}$	$\hat{y}^2$	$Y - \hat{Y}$	$(Y - \hat{Y})^2$
2	-2.67	7.11	2	-2.67	7.11	35	-4.33	18.78	34	-5.33	28.44	1	1
2	-2.67	7.11	4	-0.67	0.44	21	-18.33	336.11	26	-13.33	77.78	-5	25
2	-2.67	7.11	8	3.33	11.11	6	33.33	1,111.11	10	-29.33	860.44	-4	16
4	-0.67	0.44	2	-2.67	7.11	40	0.67	0.44	46	6.67	44.44	-6	36
4	-0.67	0.44	4	-0.67	0.44	34	-5.33	28.44	38	-1.33	1.78	-4	16
4	-0.67	0.44	8	3.33	11.11	18	-21.33	455.11	22	-17.33	300.44	-4	16
8	3.33	11.11	2	-2.67	7.11	61	21.67	469.44	70	30.67	940.44	-9	81
8	3.33	11.11	4	-0.67	0.44	58	18.67	348.44	62	22.67	513.78	-4	16
8	3.33	11.11	8	3.33	11.11	46	6.67	44.44	46	6.67	44.44	0	0
2	-2.67	7.11	2	-2.67	7.11	39	-0.33	0.11	34	-5.33	28.44	5	25
2	-2.67	7.11	4	-0.67	0.44	31	-8.33	69.44	26	-13.33	177.78	5	25
2	-2.67	7.11	8	3.33	11.11	8	-31.33	981.78	10	-29.33	860.44	-2	4
4	-0.67	0.44	2	-2.67	7.11	52	12.67	160.44	46	6.67	44.44	6	36
4	-0.67	0.44	4	-0.67	0.44	42	2.67	7.11	38	-1.33	1.78	4	16
4	-0.67	0.44	8	3.33	11.11	26	-13.33	177.78	22	-17.33	300.44	4	16
8	3.33	11.11	2	-2.67	7.11	73	33.67	1,133.44	70	30.67	940.44	3	9
8	3.33	11.11	4	-0.67	0.44	66	26.67	711.11	62	22.67	513.78	4	16
8	3.33	11.11	8	3.33	11.11	52	12.67	160.44	46	6.67	44.44	6	36
		112.00			112.00			6,214.00			5,824.00		390.00
		SS_X			SS_T			SS_{total}			$SS_{\text{regression}}$		SS_{residual}

Table 5.6 Some of the quantities needed to compute the coefficient of correlation between Y and $\hat{Y}$ (data from Table 5.3). The following abbreviations are used: $y = Y - M_Y$, $\hat{y} = \hat{Y} - M_{\hat{Y}}$, $t = (T - M_T)$, $x = (X - M_X)$. (Totals may not be exact due to rounding errors.)

5.8 Mathematical digressions

In the text, we mentioned several results that justified the approach taken. If you want to know how these derivations are obtained, we give them below.

5.8.1 Digression 1: finding the values of a , b , and c

In this section, we show that the values of a , b , and c that satisfy the equation

$$\min = \sum (Y - \hat{Y})^2 = \sum [Y - (a + bX + cT)]^2$$

are as follows.

- For the slope parameters:

$$b = \frac{SS_T SCP_{XY} - SCP_{XT} SCP_{TY}}{SS_X SS_T - (SCP_{XT})^2}.$$

$$c = \frac{SS_X SCP_{TY} - SCP_{XT} SCP_{XY}}{SS_X SS_T - (SCP_{XT})^2}.$$

- For the intercept:

$$a = M_Y - bM_X - cM_T.$$

with

$$SS_X = \sum (X - M_X)^2 = \sum X^2 - SM_X^2$$

$$SS_T = \sum (T - M_T)^2 = \sum T^2 - SM_T^2$$

$$SCP_{XY} = \sum (X - M_X)(Y - M_Y) = \sum XY - SM_X M_Y$$

$$SCP_{XT} = \sum (X - M_X)(T - M_T) = \sum XT - SM_X M_T$$

$$SCP_{TY} = \sum (T - M_T)(Y - M_Y) = \sum TY - SM_T M_Y.$$

The goal of this section is to give you an insight into the techniques used to derive the statistical tools. If you feel comfortable accepting this result at face value you can skip this section.

The general idea is to use a basic result from calculus stating that the minimum of a quadratic function is reached when its derivative is zero.

The first step is to develop the quantity we want to minimize. Let us denote by $\mathcal{E}$ (like Error) the expression to minimize:

$$\mathcal{E} = \sum (Y - \hat{Y})^2. \quad (5.32)$$

Developing $\mathcal{E}$ gives:

$$\begin{aligned} \mathcal{E} &= \sum [Y - (a + bX + cT)]^2 \\ &= \sum [Y^2 + (a + bX + cT)^2 - 2Y(a + bX + cT)] \\ &= \sum Y^2 + \sum (a + bX + cT)^2 - 2 \sum Y(a + bX + cT) \end{aligned}$$

$$\begin{aligned}
&= \sum Y^2 + \sum (a + bX + cT)^2 - 2 \sum (aY + bXY + cTY) \\
&= \sum Y^2 + \sum a^2 + \sum b^2 X^2 + \sum c^2 T^2 + 2 \sum abX + 2 \sum acT \\
&\quad + 2 \sum bcTX - 2 \sum aY - 2 \sum bXY - 2 \sum cTY \\
&= \sum Y^2 + Sa^2 + b^2 \sum X^2 + c^2 \sum T^2 + 2ab \sum X + 2ac \sum T \\
&\quad + 2bc \sum TX - 2a \sum Y - 2b \sum XY - 2c \sum TY . \tag{5.33}
\end{aligned}$$

As we have done in Chapter 4, we use the ‘magical formula from calculus’. A quadratic function reaches its minimum when its derivatives are equal to zero.

So the second step is to take the derivative of $\mathcal{E}$ with respect to a , b and c . This gives

$$\frac{\partial \mathcal{E}}{\partial a} = 2Sa + 2b \sum X + 2c \sum T - 2 \sum Y , \tag{5.34}$$

and

$$\frac{\partial \mathcal{E}}{\partial b} = 2b \sum X^2 + 2a \sum X + 2c \sum XT - 2 \sum XY , \tag{5.35}$$

and

$$\frac{\partial \mathcal{E}}{\partial c} = 2c \sum T^2 + 2a \sum T + 2b \sum XT - 2 \sum TY . \tag{5.36}$$

We are now trying to find the values of a , b and c for which the derivatives vanish. This gives the following set of equations (called the ‘normal equations’). From Equation 5.34, we obtain

$$\frac{\partial \mathcal{E}}{\partial a} = 0 \iff Sa + b \sum X + c \sum T - \sum Y = 0 . \tag{5.37}$$

From Equation 5.35, we obtain

$$\frac{\partial \mathcal{E}}{\partial b} = 0 \iff b \sum X^2 + a \sum X + c \sum XT - \sum XY = 0 . \tag{5.38}$$

From Equation 5.36, we obtain

$$\frac{\partial \mathcal{E}}{\partial c} = 0 \iff c \sum T^2 + a \sum T + b \sum XT - \sum TY = 0 . \tag{5.39}$$

The next (easy) step is to find the value of a (the intercept). From Equation 5.37, it is found as

$$Sa = \sum Y - b \sum X - c \sum T$$

which simplifies as

$$\begin{aligned}
a &= \frac{1}{S} (\sum Y - b \sum X - c \sum T) \\
&= M_Y - bM_X - cM_T
\end{aligned} \tag{5.40}$$

as stated.

Plugging in the results of Equation 5.40 into 5.38 and developing gives

$$\begin{aligned}
b \sum X^2 + a \sum X + c \sum XT - \sum XY &= 0 \\
b \sum X^2 + (M_Y - bM_X - cM_T) \sum X + c \sum XT - \sum XY &= 0
\end{aligned}$$

$$b \sum X^2 + M_Y \sum X - bM_X \sum X - cM_T \sum X + c \sum XT - \sum XY = 0 \quad (5.41)$$

Because

$$M_X = \frac{1}{S} \sum X, \sum X = SM_X.$$

Using this equivalence in Equation 5.41 gives

$$b \sum X^2 + SM_Y M_X - bSM_X^2 - cSM_X M_T + c \sum XT - \sum XY = 0.$$

Rearranging and simplifying gives

$$\begin{aligned} b \sum X^2 - bSM_X^2 + c \sum XT - cSM_X M_T - \sum XY + SM_Y M_X &= 0 \\ b \left(\sum X^2 - SM_X^2 \right) + c \left(\sum XT - SM_X M_T \right) - \left(\sum XY - SM_Y M_X \right) &= 0 \end{aligned} \quad (5.42)$$

which is equivalent to

$$bSS_X + cSCP_{XT} - SCP_{XY} = 0. \quad (5.43)$$

The same technique that we have just used gives the following result when applied to Equation 5.39:

$$cSS_T + bSCP_{XT} - SCP_{TY} = 0. \quad (5.44)$$

So we end up with two equations and two unknowns (i.e. b and c). We will use the standard method of elimination. The first step is to solve the second equation (i.e. Equation 5.44) for c . We find

$$\begin{aligned} cSS_T &= SCP_{TY} - bSCP_{XT} \\ c &= \frac{SCP_{TY} - bSCP_{XT}}{SS_T}. \end{aligned} \quad (5.45)$$

We then substitute the value of c in Equation 5.43. This gives

$$bSS_X + \left[\frac{SCP_{TY} - bSCP_{XT}}{SS_T} SCP_{XT} \right] - SCP_{XY} = 0.$$

Solving this equation for b , we obtain

$$\begin{aligned} bSS_X + \frac{SCP_{XT}SCP_{TY} - b(SCP_{XT})^2}{SS_T} - SCP_{XY} &= 0 \\ b \left[SS_X - \frac{(SCP_{XT})^2}{SS_T} \right] + \frac{SCP_{XT}SCP_{TY}}{SS_T} - SCP_{XY} &= 0 \end{aligned} \quad (5.46)$$

which is equivalent to

$$b \left[\frac{SS_X SS_T - (SCP_{XT})^2}{SS_T} \right] = SCP_{XY} - \frac{SCP_{XT}SCP_{TY}}{SS_T}$$

and hence,

$$\begin{aligned} b &= \left[\frac{SS_T SCP_{XY} - SCP_{XT}SCP_{TY}}{SS_T} \right] \times \left[\frac{SS_T}{SS_X SS_T - (SCP_{XT})^2} \right] \\ &= \frac{SS_T SCP_{XY} - SCP_{XT}SCP_{TY}}{SS_X SS_T - (SCP_{XT})^2}, \end{aligned} \quad (5.47)$$

which is the result we wanted to prove.

We can now find the value of c by using a technique similar to the one used for finding the value of b (this is left as an exercise for the courageous reader). Doing so will give

$$c = \frac{SCP_X SCP_{TY} - SCP_{XT} SCP_{XY}}{SS_X SS_T - (SCP_{XT})^2}.$$

Et voilà!

Chapter summary

5.9 Key notions of the chapter

Below are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

3D scatter plot or 3D plot	Orthogonality of the independent variables
Equation of a plane	Partition of the sum of squares of regression
Regression plane	
Multiple correlation coefficient	
Score model	

5.10 New notations

Below are the new notations introduced in this chapter. Test yourself on their meaning.

R^2	$R^2_{Y \cdot XT}$
$r^2_{\hat{Y} \cdot Y}$	$SS_{Y \cdot X} \quad MS_{Y \cdot X} \quad SS_{Y \cdot T} \quad MS_{Y \cdot T}$
$r^2_{Y \cdot X} \quad r^2_{Y \cdot T}$	

5.11 Key formulas of the chapter

Below are the main formulas introduced in this chapter: try to go through them and understand what they mean.

$\hat{Y} = a + bX + cT$
with

- b being the slope for X

$$b = \frac{SS_T SCP_{XY} - SCP_{XT} SCP_{TY}}{SS_X SS_T - (SCP_{XT})^2}$$

When X and T are orthogonal:

$$b = \frac{SCP_{XY}}{SS_X}$$

- c being the slope for T

$$c = \frac{SS_X SCP_{TY} - SCP_{XT} SCP_{XY}}{SS_X SS_T - (SCP_{XT})^2}.$$

When X and T are orthogonal:

$$b = \frac{SCP_{XY}}{SS_T}$$

- a being the intercept

$$a = M_Y - bM_X - cM_T$$

$$F_{Y \cdot XT} = \frac{R^2}{1 - R^2} \times \frac{S - K - 1}{K}$$

where S is the number of subjects and K the number of independent variables used to predict the dependent variable.

$$Y = M_Y + b(X - M_X) + c(T - M_T) + (Y - \hat{Y})$$

$$\begin{aligned} SS_{\text{regression}} &= r_{Y \cdot X}^2 SS_X + r_{Y \cdot T}^2 SS_T \\ &= SS_{\text{regression of } Y \text{ on } X} + SS_{\text{regression of } Y \text{ on } T} \\ &= SS_{Y \cdot X} + SS_{Y \cdot T} \end{aligned}$$

$$R_{Y \cdot XT}^2 = r_{Y \cdot X}^2 + r_{Y \cdot T}^2$$

$$df_{\text{regression}} = 2 \text{ and } df_{\text{residual}} = S - K - 1.$$

5.12 Key questions of the chapter

Below are some questions about the content of this chapter. All the answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- * In which case can we apply multiple regression analysis?
- * What is the score of a given subject made of?
- * Why is it important that the independent variables be orthogonal?

6

Non-orthogonal multiple regression

6.1 Introduction

We have seen in Chapter 5 that multiple regression analysis is easy (both computationally and methodologically) when the independent variables are orthogonal. Therefore experimenters prefer to use orthogonal variables, but it is not always possible to design experiments with orthogonal independent variables. This is the case, in particular, if one or both of the independent variables are ‘tag variables’ (i.e. variables that describe a subject but that cannot be manipulated *per se* like age, sex, nationality, etc. See Chapter 1 for a refresher about the notion of tag variable).

We will see that the general equations derived previously can still be used. However, the problem of evaluating the *specific* contribution of each independent variable to the explanation of the variance of the dependent variable becomes more delicate. The main new notion that we will use to that effect is called the *semi-partial correlation* of an independent variable with the dependent variable. This coefficient expresses the portion of the dependent variable explained *specifically* by this independent variable.

6.2 Example: age, speech rate and memory span

To illustrate an experiment with two quantitative independent variables, suppose that we performed the following replication of an experiment originally designed by Hulme, Thomson, Muir, and Lawrence (1984, as reported by Baddeley, 1990, p. 78ff.). In this study, children aged 4, 7, or 10 years (hence ‘age’ is the first independent variable in this experiment, we will denote it X) were tested for 10 series of immediate serial recall of 15 items. The dependent variable is the total number of words correctly recalled (i.e. in the correct order). In addition to age the speech rate of each child was obtained by asking the child to read aloud a list of words. Dividing the number of words read by the time needed to read them gave the *speech rate* (expressed in words per second) of the child. Speech rate is the second independent variable in this experiment (we will denote it T).

The research hypothesis (which the authors wanted to evaluate) states that the age and the speech rate of the children determine their memory performance. Because the authors cannot manipulate the independent variables speech rate and age, they cannot design their

The independent variables		The dependent variable
X Age (in years)	T Speech rate (words per second)	Y Memory span (number of words recalled)
4	1	14
4	2	23
7	2	30
7	4	50
10	3	39
10	6	67

Table 6.1 Data from a (fictitious) replication of an experiment of Hulme *et al.* (1984). The dependent variable is the total number of words recalled in 10 series of immediate recall of items: it is a measure of the *memory span*. The first independent variable is the *age* of the child, the second independent variable is the *speech rate* of the child.

experiment in such a way as to have the two independent variables orthogonal to each other. In other words, one can expect speech rate to be partly correlated with age (on the average, older children tend to speak faster than younger children). In addition, the authors believe that speech rate should be the major determinant of performance and that the effect of age reflects more the confounded effect of speech rate rather than age *per se*. The data obtained from a sample of 6 subjects are given in Table 6.1.

6.3 Computation of the regression plane

Table 6.2 gives the quantities needed to compute the values of the intercept and the slopes of the regression plane for the memory span example. If we use these quantities to determine the equation plane, we obtain the following values:

- slope for X:

$$\begin{aligned}
 b &= \frac{SS_T SCP_{XY} - SCP_{XT} SCP_{TY}}{SS_X SS_T - (SCP_{XT})^2} \\
 &= \frac{16 \times 207 - 18 \times 170}{36 \times 16 - 18^2} \\
 &= 1.00 .
 \end{aligned} \tag{6.1}$$

- slope for T:

$$\begin{aligned}
 c &= \frac{SS_X SCP_{TY} - SCP_{XT} SCP_{XY}}{SS_X SS_T - (SCP_{XT})^2} \\
 &= \frac{36 \times 170 - 18 \times 207}{36 \times 16 - 18^2} \\
 &= 9.50 .
 \end{aligned} \tag{6.2}$$

<i>Y</i>	<i>y</i>	<i>y</i> ²	<i>X</i>	<i>x</i>	<i>x</i> ²	<i>T</i>	<i>t</i>	<i>t</i> ²	<i>xy</i>	<i>ty</i>	<i>xt</i>
14.00	−23.17	536.69	4.00	−3.00	9.00	1.00	−2.00	4.00	69.50	46.33	6.00
23.00	−14.17	200.70	4.00	−3.00	9.00	2.00	−1.00	1.00	42.50	14.17	3.00
30.00	−7.17	51.36	7.00	0.00	0.00	2.00	−1.00	1.00	0.00	7.17	0.00
50.00	12.83	164.69	7.00	0.00	0.00	4.00	1.00	1.00	0.00	12.83	0.00
39.00	1.83	3.36	10.00	3.00	9.00	3.00	0.00	0.00	5.50	0.00	0.00
67.00	29.83	890.03	10.00	3.00	9.00	6.00	3.00	9.00	89.50	89.50	9.00
Σ		1,846.83			36.00			16.00	207.00	170.00	18.00
		<i>SS_Y</i>			<i>SS_X</i>			<i>SS_T</i>	<i>SCP_{XY}</i>	<i>SCP_{TY}</i>	<i>SCP_{XT}</i>

Table 6.2 The different quantities needed to compute the values of the parameters *a*, *b*, and *c* of a regression plane defined by the equation $\hat{Y} = a + bX + cT$. The following abbreviations are used to label the columns of the table: $t = (T - M_T)$, $x = (X - M_X)$, $y = (Y - M_Y)$.

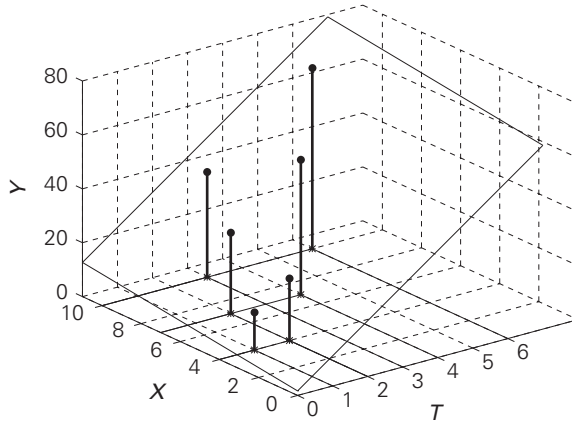


Figure 6.1 Plot of the results of a replication of the experiment of Hulme *et al.* (1984). Each data point represents the number of words correctly recalled (Y , the dependent variable) by a child aged 4, 7, or 10 years (X , the first independent variable) with a speech rate of 1, 2, 3, 4, or 6 words per second (T , the second independent variable). The regression plane to predict Y from X and T is given by the equation: $\hat{Y} = a + bX + cT = 1.67 + 1.00X + 9.50T$. Data from Table 6.1.

- intercept

$$\begin{aligned}
 a &= M_Y - bM_X - cM_T \\
 &= 37.17 - (1 \times 7) - (9.5 \times 3) \\
 &= 1.67 .
 \end{aligned} \tag{6.3}$$

Substituting these values in the general equation of a plane, we obtain the equation predicting the values of the dependent variable (memory span) from the independent variables (age and speech rate):

$$\hat{Y} = a + bX + cT \tag{6.4}$$

$$= 1.67 + 1.00X + 9.50T \tag{6.5}$$

or

$$\text{no. of words recalled} = 1.67 + (1.00 \times \text{age}) + (9.50 \times \text{speech rate}) .$$

Figure 6.1 shows the original data plotted as a 3D plot with the regression plane superimposed. From the values of the intercept and the slopes, we can see that being 1 year older enables children to remember 1 more word and that each increment of 1 word per second in speech rate enables children to remember 9.5 more words.

	Y	y	y^2	$\hat{Y}$	$\hat{y}$	$\hat{y}^2$	$\hat{y} \times y$
	14.00	−23.17	536.69	15.17	−22.00	484.00	509.67
	23.00	−14.17	200.69	24.67	−12.50	156.25	177.08
	30.00	−7.17	51.36	27.67	−9.50	90.25	68.08
	50.00	12.83	164.69	46.67	9.50	90.25	121.92
	39.00	1.83	3.36	40.17	3.00	9.00	5.50
	67.00	29.82	890.03	68.67	31.50	992.25	939.75
Σ			1,846.83			1,822.00	1,822.00
			SS_Y			$SS_{\hat{Y}}$	$SCP_{\hat{Y}Y}$

Table 6.3 Quantities needed to compute the coefficient of correlation between Y and $\hat{Y}$ (data from Table 6.1). The following abbreviations are used: $y = Y - M_Y$, $\hat{y} = \hat{Y} - M_{\hat{Y}}$.

6.4 How to evaluate the quality of the prediction

Table 6.3 gives the quantities to compute $R_{Y.XT}^2$ for our example. Using these values, we obtain:

$$\begin{aligned}
 R_{Y.XT}^2 &= \frac{(SCP_{\hat{Y}Y})^2}{SS_{\hat{Y}}SS_Y} \\
 &= \frac{1,822.00^2}{1,822.00 \times 1,846.83} \\
 &= .9866, \tag{6.6}
 \end{aligned}$$

which indicates that, taken together, age and speech rate account for 98.66% of the variance of the number of words correctly recalled.

To test the significance of $R_{Y.XT}^2$ we use the formula of $F_{Y.XT}$ for multiple regression presented in the previous chapter:

$$F_{Y.XT} = \frac{R_{Y.XT}^2}{1 - R_{Y.XT}^2} \times \frac{S - K - 1}{K},$$

where K is the number of independent variables used to predict the dependent variable. For our example, $S = 6$ and $K = 2$, thus the F ratio is obtained as

$$F_{Y.XT} = \frac{R_{Y.XT}^2}{1 - R_{Y.XT}^2} \times \frac{S - 2 - 1}{2}$$

which gives

$$\begin{aligned}
 F_{Y.XT} &= \frac{.9866}{1 - .9866} \times \frac{6 - 3}{2} \\
 &= 110.50.
 \end{aligned}$$

We can now compare this value with the critical value given by the Fisher distribution table (see Table 2 in the Appendix on page 499) for $\nu_1 = 2$ and $\nu_2 = S - K - 1 = 3$. If we decide upon an α value of .05, the critical value of F is equal to 9.55. Since the computed value of $F_{Y.XT} = 110.50$ is larger than the critical value, we can reject the null hypothesis and accept the alternative hypothesis stating that there is a linear relationship between the two independent variables (age and speech rate) and the dependent variable (number of words recalled).

	X	x	x^2	T	t	t^2	$\hat{Y}$	$\hat{y}$	$\hat{y}^2$	$\hat{y} \times x$	$\hat{y} \times t$
	4	-3	9	1	-2	4	15.17	-22.00	484.00	66.00	44.00
	4	-3	9	2	-1	1	24.67	-12.50	156.25	37.50	12.50
	7	0	0	2	-1	1	27.67	-9.50	90.25	0.00	9.50
	7	0	0	4	1	1	46.67	9.50	90.25	0.00	9.50
	10	3	9	3	0	0	40.17	3.00	9.00	9.00	0.00
	10	3	9	6	3	9	68.67	31.50	992.25	94.50	94.50
Σ			36.00			16.00			1,822.00	207.00	170.00
			SS_X			SS_T			$SS_{\hat{Y}}$	$SCP_{\hat{Y}X}$	$SCP_{\hat{Y}T}$

Table 6.4 Quantities needed to compute the coefficients of correlation between X and $\hat{Y}$ and between T and $\hat{Y}$. The following abbreviations are used: $y = Y - M_Y$, $\hat{y} = \hat{Y} - M_{\hat{Y}}$, $t = (T - M_T)$, $x = (X - M_X)$.

6.4.1 How to evaluate the importance of each independent variable in the prediction

To evaluate the quality of prediction of each independent variable separately we can compute a separate coefficient of correlation for each of the independent variables and the predicted values of the dependent variable. Each of these coefficients gives the importance of a given independent variable for the predicted values of the dependent variable. Table 6.4 gives the quantities needed for these computations.

- The quality of prediction of variable X is estimated by computing the coefficient of correlation between $\hat{Y}$ and X :

$$\begin{aligned}
 r_{\hat{Y}.X}^2 &= \frac{(SCP_{\hat{Y}X})^2}{SS_{\hat{Y}}SS_X} \\
 &= \frac{207^2}{1,822 \times 36} \\
 &= .6533 .
 \end{aligned}$$

- The quality of prediction of variable T is estimated by computing the coefficient of correlation between $\hat{Y}$ and T :

$$\begin{aligned}
 r_{\hat{Y}.T}^2 &= \frac{(SCP_{\hat{Y}T})^2}{SS_{\hat{Y}}SS_T} \\
 &= \frac{170^2}{1,822 \times 16} \\
 &= .9914
 \end{aligned}$$

which indicates that the variable T (speech rate) is a better predictor of the number of words correctly recalled than the variable X (age).

As we can see, when the independent variables are non-orthogonal, the square coefficients of correlation between the independent variables and the predicted value do not add up to 1 any more:

$$r_{\hat{Y}.X}^2 + r_{\hat{Y}.T}^2 \neq 1 \quad (6.7)$$

$$.6533 + .9914 = 1.6447 \neq 1 ! \quad (6.8)$$

In order to assess the importance of each independent variable we can also, as previously, compute the coefficient of correlation of each independent variable with the dependent variable. This coefficient gives the importance of a given independent variable for the predicted values of the dependent variable. Table 6.2 gives the quantities needed to compute these coefficients of correlation.

- The quality of prediction of the variable X is estimated by computing the coefficient of correlation between Y and X :

$$\begin{aligned} r_{Y.X}^2 &= \frac{(SCP_{XY})^2}{SS_Y SS_X} \\ &= \frac{207^2}{1,846.83 \times 36} \\ &= .6445 . \end{aligned}$$

- The quality of prediction of variable T is estimated by computing the coefficient of correlation between Y and T :

$$\begin{aligned} r_{Y.T}^2 &= \frac{(SCP_{TY})^2}{SS_Y SS_T} \\ &= \frac{170^2}{1,846.83 \times 16} \\ &= .9780 . \end{aligned}$$

As we can see, when the independent variables are non-orthogonal, the square coefficients of correlation between the independent variables and the dependent variable no longer add up to $R_{Y.XT}^2$ (the square multiple coefficient of correlation):

$$\begin{aligned} r_{Y.X}^2 + r_{Y.T}^2 &\neq R_{Y.XT}^2 & (6.9) \\ .6445 + .9780 &= 1.6225 \neq .9866 ! \end{aligned}$$

This is clearly a problem, because we can no longer interpret the square coefficients of correlation between one independent variable and the dependent variable as the proportion of the explained variance of the dependent variable by this independent variable. In other words, because the independent variables are correlated, they share some variance (actually the square coefficient of correlation between X and T is equal to .56, hence they share 56% of their variance), and this common variance explains some of the variance of the dependent variable. As a consequence, we are now *overestimating* the variance explained by adding up the squared coefficients of correlation. In the present example, we will find later that the variance common to X and T explains a proportion of .6360 of the variance of the dependent variable. When we add $r_{Y.X}^2$ and $r_{Y.T}^2$ this common variance is added twice.

6.4.2 The specific contribution of each independent variable: the semi-partial contribution

There are several ways to try to overcome the problem of the variance of the dependent variable explained by the common variance of X and T being counted twice. One possible way of doing this could be to use a *hierarchical* approach. This can make sense if we have a *theoretical* reason to impose an order on the independent variables. In this case, we would

	X	x	x^2	T	t	t^2	$x \times t$
	4	-3	9	1	-2	4	6
	4	-3	9	2	-1	1	3
	7	0	0	2	-1	1	0
	7	0	0	4	1	1	0
	10	3	9	3	0	0	0
	10	3	9	6	3	9	9
Σ	42	0	36	18	0	16	18
	SS_X			SS_T		SCP_{XT}	

Table 6.5 The different quantities needed to compute the values of the parameters $a_{X \cdot T}$, $b_{X \cdot T}$, $a_{T \cdot X}$, and $b_{T \cdot X}$. The following abbreviations are used: $x = (X - M_X)$, $t = (T - M_T)$.

start by evaluating the proportion of variance of the dependent variable explained by the first independent variable. Then we will try to determine what the second independent variable can specifically explain. Even though this approach can be meaningful in some domains of psychology, it is rarely relevant in an experimental context. What most experimenters want to know when dealing with a design with several independent variables is the *specific* contribution of each independent variable. In order to do so, the first step is to isolate the specific part of each independent variable. This is done by first predicting a given independent variable from the other independent variable.¹ The residual of the prediction is by definition uncorrelated with the predictor, hence it represents the specific part of the independent variable under consideration. To illustrate this procedure, let us denote by $\hat{X}_T$ the prediction of X from T and by $\hat{T}_X$ the prediction of T from X .

The equation for predicting X from T is given by

$$\hat{X}_T = a_{X \cdot T} + b_{X \cdot T} T, \quad (6.10)$$

where $a_{X \cdot T}$ and $b_{X \cdot T}$ denote the intercept and slope of the regression line of the prediction of X from T .

The equation for predicting T from X is given by

$$\hat{T}_X = a_{T \cdot X} + b_{T \cdot X} X. \quad (6.11)$$

where $a_{T \cdot X}$ and $b_{T \cdot X}$ denote the intercept and slope of the regression line of the prediction of T from X .

Table 6.5 gives the values of the sums of squares and sum of cross-products needed to compute the prediction of X from T and the prediction of T from X .

- We find the following values for predicting X from T :

$$b_{X \cdot T} = \frac{SCP_{XT}}{SS_T} = \frac{18}{16} = 1.125; \quad (6.12)$$

$$a_{X \cdot T} = M_X - b_{X \cdot T} \times M_T = 7 - 1.125 \times 3 = 3.625. \quad (6.13)$$

¹ Or by all the other independent variables if we are dealing with the general case of more than two independent variables.

	Y	y	y ²	X	$\widehat{X}_T$	$e_{X \cdot T}$	$e_{X \cdot T}^2$	$y \times e_{X \cdot T}$
	14	-23.1667	536.69	4	4.7500	-0.7500	0.5625	17.3750
	23	-14.1667	200.69	4	5.8750	-1.8750	3.5156	26.5625
	30	-7.1667	51.36	7	5.8750	1.1250	1.2656	-8.0625
	50	12.8333	164.69	7	8.1250	-1.1250	1.2656	-14.4375
	39	1.8333	3.36	10	7.0000	3.0000	9.0000	5.5000
	67	29.8333	890.03	10	10.3750	-0.3750	0.1406	-11.1875
Σ	223	0	1,846.83	42	42.0000	0	15.7500	15.7500
	SS_Y						$SS_{e_{X \cdot T}}$	$SCP_{Ye_{X \cdot T}}$

Table 6.6 The different quantities needed to compute the semi-partial coefficient of correlation between Y and X after the effects of T have been partialled out of X. The following abbreviations are used: $y = Y - M_Y$, $e_{X \cdot T} = X - \widehat{X}_T$.

	Y	y	y ²	T	$\widehat{T}_X$	$e_{T \cdot X}$	$e_{T \cdot X}^2$	$y \times e_{T \cdot X}$
	14	-23.1667	536.69	1	1.50	-0.50	0.25	11.5833
	23	-14.1667	2,200.69	2	1.50	0.50	0.25	-7.0833
	30	-7.1667	51.36	2	3.00	-1.00	1.00	7.1667
	50	12.8333	164.94	4	3.00	1.00	1.00	12.8333
	39	1.8333	3.36	3	4.50	-1.50	2.25	-2.7500
	67	29.8333	890.03	6	4.50	1.50	2.25	44.7500
Σ	223	0	1,846.83	18	18.00	0	7.00	66.5000
	SS_Y						$SS_{e_{T \cdot X}}$	$SCP_{Ye_{T \cdot X}}$

Table 6.7 The different quantities needed to compute the semi-partial coefficient of correlation between Y and T after the effects of X have been partialled out of T. The following abbreviations are used: $y = Y - M_Y$, $e_{T \cdot X} = T - \widehat{T}_X$.

- We find the following values for predicting T from X:

$$b_{T \cdot X} = \frac{SCP_{TX}}{SS_X} = \frac{18}{36} = .5; \quad (6.14)$$

$$a_{T \cdot X} = M_T - b_{T \cdot X} \times M_X = 3 - .5 \times 7 = -0.5. \quad (6.15)$$

So, the first step is to predict one independent variable from the other one. Then, by subtracting the predicted value of the independent variable from its actual value, we will obtain the residual of the prediction of this independent variable. This residual constitutes the *specific* component of this independent variable. We need a new notation to denote this. The specific part of X is denoted $e_{X \cdot T}$, computed as

$$e_{X \cdot T} = X - \widehat{X}_T. \quad (6.16)$$

The specific part of T is denoted $e_{T \cdot X}$, computed as

$$e_{T \cdot X} = T - \widehat{T}_X. \quad (6.17)$$

Tables 6.6 and 6.7 give the quantities needed to compute the correlation between the dependent variable and the specific component of X and T. This new coefficient of correlation computed between the residual of an independent variable and the dependent variable is called

the *semi-partial correlation* between the independent variable and the dependent variable. It is called semi-partial because we correlate *part* of an independent variable with the *whole* dependent variable. A synonym of semi-partial correlation is *part correlation*. The square semi-partial correlation between the dependent variable and the independent variable X is noted $r_{Y \cdot X|T}^2$ read (' r square of Y and X after T has been partialled out'). It is obtained as

$$r_{Y \cdot X|T}^2 = r_{Y \cdot e_{X \cdot T}}^2 = \frac{(SCP_{Ye_{X \cdot T}})^2}{SS_Y SS_{e_{X \cdot T}}} . \quad (6.18)$$

In our example, we find

$$r_{Y \cdot X|T}^2 = \frac{15.75^2}{1,846.83 \times 15.75} = .0085 .$$

The square semi-partial correlation between the dependent variable and the independent variable T is noted $r_{Y \cdot T|X}^2$. It is obtained as

$$r_{Y \cdot T|X}^2 = r_{Y \cdot e_{T \cdot X}}^2 = \frac{(SCP_{Ye_{T \cdot X}})^2}{SS_Y SS_{e_{T \cdot X}}} . \quad (6.19)$$

In our example, we find

$$r_{Y \cdot T|X}^2 = \frac{66.50^2}{1,846.83 \times 7} = .3421 .$$

As the values of the semi-partial correlation coefficients show, the specific explanation of each independent variable is very different from the total explanation given by each of them.

6.5 Semi-partial correlation as increment in explanation

Another way to evaluate the specific contribution of each independent variable to the explanation of the dependent variable is to evaluate it as the increase in the amount of explained variance when the independent variable is introduced in addition to the other independent variables in the regression. In other words, we consider that the importance of an independent variable corresponds to what this independent variable can explain *in addition* to what the other independent variables already explain. A nice property of the square semi-partial coefficient of correlation between a given independent variable and the dependent variable is to give the increment in the proportion of variance of the dependent variable obtained when the independent variable under consideration is used in addition to all the other independent variables. Therefore, the square semi-partial coefficient of correlation between a given independent variable and the dependent variable gives the part of variance that is explained by this independent variable only (i.e. beyond and above what the other independent variables already explain). This is why we say that the square semi-partial coefficient of correlation between a given independent variable and the dependent variable is the *specific* or *unique* contribution of this independent variable to explaining the dependent variable.

For example, with two independent variables, the square semi-partial coefficient of correlation between Y and X corresponds to the increment of explained variance that is obtained beyond and above the explanation of Y by T . The amount of explained variance of Y by T is given by the square coefficient of correlation between Y and T . The variance explained by T and X is $R_{Y \cdot XT}^2$. Therefore, the specific or unique contribution of X corresponds to the

increase in explained variance when we go from a model with T being the only independent variable to a model with T and X as the independent variables. With a formula, this amounts to saying that the square semi-partial coefficient of correlation can be obtained by subtraction as

$$r_{Y \cdot X|T}^2 = R_{Y \cdot XT}^2 - r_{Y \cdot T}^2 . \quad (6.20)$$

In other words, the semi-partial correlation of X and the dependent variable is what is left of the square correlation between the dependent variable and the independent variables after subtraction of what can be explained by something else other than X . For our example, we find that²

$$r_{Y \cdot X|T}^2 = R_{Y \cdot XT}^2 - r_{Y \cdot T}^2 = .9866 - .9780 \approx .0085 ,$$

and

$$r_{Y \cdot T|X}^2 = R_{Y \cdot XT}^2 - r_{Y \cdot X}^2 = .9866 - .6445 = .3421 . \quad (6.21)$$

From the interpretation of the semi-partial correlation as the proportion of the variance of the dependent variable explained by the *specific* portion of an independent variable, we can also deduce the proportion of the variance of the dependent variable explained by the common portion of the independent variable by subtracting the square semi-partial correlation coefficients from $R_{Y \cdot XT}^2$. With a formula:

$$\text{Part of } Y \text{ explained by both } X \text{ and } T = R_{Y \cdot XT}^2 - (r_{Y \cdot X|T}^2 + r_{Y \cdot T|X}^2) . \quad (6.22)$$

For our example, we find that

$$R_{Y \cdot XT}^2 - (r_{Y \cdot X|T}^2 + r_{Y \cdot T|X}^2) = .9866 - (.0085 + .3421) = .6360$$

or 63.6% of the variance of Y is explained by the common portion of X and T .

Hence we can decompose the variance of Y as

$$\begin{aligned} \text{variance of } Y &= \text{variance explained specifically by } X \\ &\quad + \text{variance explained specifically by } T \\ &\quad + \text{variance explained by the common portion of } X \text{ and } T \\ &\quad + \text{variance left unexplained by } X \text{ and } T . \end{aligned}$$

For our example, we find that the variance of Y is decomposed proportionally as

$$1 = \underbrace{\frac{r_{Y \cdot X|T}^2}{R_{Y \cdot XT}^2} + \frac{r_{Y \cdot T|X}^2}{R_{Y \cdot XT}^2} + \frac{.6360}{R_{Y \cdot XT}^2}}_{R_{Y \cdot XT}^2} + \underbrace{.0134}_{\text{unexplained variance}} .$$

Because the square coefficient of correlation between X and T is equal to $r_{X \cdot T}^2 = .5625$, X and T have 56.25% of their variance in common, or each independent variable has a specific variance of

$$1 - r_{X \cdot T}^2 = 1 - .5625 = .4375 .$$

We can try to picture the relationship of the independent variables with the dependent variable by showing what part of the independent variables explains what proportion of the dependent variable. This is done in Figure 6.2 which shows these relationships as a Venn diagram for the independent variables and as proportions for the dependent variable.

² The values are not exactly the same because of rounding errors, but if you keep enough decimal values you'll find them.

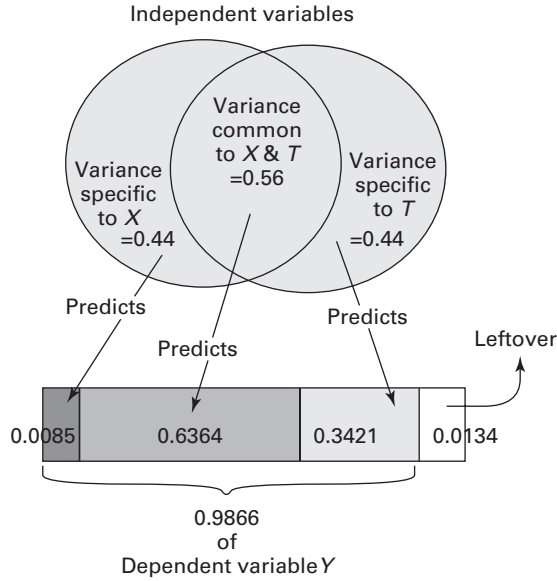


Figure 6.2 Illustration of the relationship of the independent variables with the dependent variable showing what part of the independent variables explains what proportion of the dependent variable. The independent variables are represented by a Venn diagram, the dependent variable is represented by a bar.

6.5.1 Alternative formulas for the semi-partial correlation coefficients

The semi-partial coefficient of correlation can also be computed directly from the different coefficients of correlation of the independent variable and the dependent variable. Specifically, they can be obtained as

- for the semi-partial correlation of Y and X :

$$r_{Y \cdot X|T}^2 = \frac{(r_{Y \cdot X} - r_{Y \cdot T} r_{X \cdot T})^2}{1 - r_{X \cdot T}^2}. \quad (6.23)$$

- for the semi-partial correlation of Y and T :

$$r_{Y \cdot T|X}^2 = \frac{(r_{Y \cdot T} - r_{Y \cdot X} r_{X \cdot T})^2}{1 - r_{X \cdot T}^2}. \quad (6.24)$$

For our example, taking into account that

- $r_{X \cdot T} = .7500$
- $r_{Y \cdot X} = .8028$
- $r_{Y \cdot T} = .9890$

we find the following values:

- for the semi-partial correlation of Y and X :

$$r_{Y \cdot X|T}^2 = \frac{(r_{Y \cdot X} - r_{Y \cdot T} r_{X \cdot T})^2}{1 - r_{X \cdot T}^2} = \frac{(.8028 - .9890 \times .7500)^2}{1 - .7500^2} \approx .0085 \quad (6.25)$$

- for the semi-partial correlation of Y and T :

$$r_{Y \cdot T|X}^2 = \frac{(r_{Y \cdot T} - r_{Y \cdot X} r_{X \cdot T})^2}{1 - r_{X \cdot T}^2} = \frac{(.9890 - .8028 \times .7500)^2}{1 - .7500^2} \approx .3421. \quad (6.26)$$

These values agree (within rounding errors) with the values previously computed.

6.6 F tests for the semi-partial correlation coefficients

To evaluate if the specific prediction of the dependent variable by a given independent variable is significantly different from zero, we can compute (once again) an F test.

In order to do so, we compute two F tests (one for each independent variable). They have a form already familiar (cf. Equations 5.21 and 5.22, page 101). In fact they are almost identical to the formula seen in Chapter 5. The only change is that we use now the square semi-partial coefficient of correlation instead of the square coefficient of correlation.

For the independent variable X , we compute the ratio

$$F_{Y \cdot X|T} = \frac{r_{Y \cdot X|T}^2}{1 - R_{Y \cdot XT}^2} \times (S - K - 1) \quad (6.27)$$

where K is the number of independent variables used to predict the dependent variable. Under the null hypothesis, this F ratio follows a Fisher distribution with $\nu_1 = 1$ and $\nu_2 = S - K - 1$ degrees of freedom.

For our example, $K = 2$ and hence the $F_{Y \cdot X|T}$ ratio is obtained as

$$F_{Y \cdot X|T} = \frac{r_{Y \cdot X|T}^2}{1 - R_{Y \cdot XT}^2} \times (S - 3) = \frac{.0085}{1 - .9866} \times 3 = 1.91.$$

We can now follow the standard test procedure to see if this value can be attributed to chance, or if it reflects the existence of a linear relationship between the independent variable X and the dependent variable. Suppose that we decided upon a value of $\alpha = .05$: the critical value of F for $\nu_1 = 1$ and $\nu_2 = S - K - 1 = 3$ is equal to 10.13. Since the computed value of $F_{Y \cdot X|T} = 1.91$ is smaller than the critical value, we fail to reject the null hypothesis and we cannot accept the alternative hypothesis stating that there is a linear relationship between the independent variable X (age) and the dependent variable (memory span). In other words we cannot say that age *per se* affects short term memory span (which was, remember, one research hypothesis that the authors wanted to investigate).

Likewise, for the independent variable T , we compute the ratio

$$F_{Y \cdot T|X} = \frac{r_{Y \cdot T|X}^2}{1 - R_{Y \cdot XT}^2} \times (S - K - 1) \quad (6.28)$$

where K is the number of independent variables used to predict the dependent variable. Under the null hypothesis, this F ratio follows a Fisher distribution with $\nu_1 = 1$ and $\nu_2 = S - K - 1$ degrees of freedom.

For our example, $K = 2$ and hence the $F_{Y \cdot T|X}$ ratio is obtained as

$$F_{Y \cdot T|X} = \frac{r_{Y \cdot T|X}^2}{1 - R_{Y \cdot XT}^2} \times (S - 3) = \frac{.3421}{1 - .9866} \times 3 = 76.59.$$

We can now follow the standard test procedure to see if this value can be attributed to chance or if it reflects the existence of a linear relationship between the independent variable T and the dependent variable. Suppose that we decided upon a value of $\alpha = .05$: the critical value of F for $\nu_1 = 1$ and $\nu_2 = S - K - 1 = 3$ is equal to 10.13. Since the computed value of $F_{Y \cdot T|X} = 76.59$ is larger than the critical value, we can reject the null hypothesis and accept the alternative hypothesis stating that there is a linear relationship between the independent variable T (speech rate) and the dependent variable (memory span).

6.7 What to do with more than two independent variables

It should be clear by now that dealing with orthogonal independent variables is much easier than dealing with non-orthogonal independent variables. Let us examine the differences between these cases and what this implies when the regression involves more than two independent variables.

When the independent variables are orthogonal, each of them can be analyzed separately. Because the slope coefficients (i.e. b and c) are the same for the whole regression or for the regression line computed from only one independent variable, they can be computed one at a time with the simple formula of Chapter 4. The square multiple coefficient of correlation is obtained as the sum of the square correlation coefficients of the dependent variable with each independent variable. The *specific* contribution of each independent variable is nicely given by its square coefficient of correlation with the dependent variable. To summarize: *multiple regression with several orthogonal independent variables is easy and reduces to a series of simple regression analyses.*

When the independent variables are not orthogonal, then life is much more complex. First, the computation of the slope becomes much more complex as the number of variables grows. In fact, it is almost out of the question to be able to compute the slopes by hand with more than three independent variables.³ The same problem arises for the computation of the semi-partial correlations. However, the *interpretation* of the square semi-partial coefficient of correlation as the part of the variance explained by the *specific* part of an independent variable is still valid, as well as its interpretation as the increment of explained variance when the independent variable is entered last in the regression equation. Because of the computational complexity of computing slopes and semi-partial correlations, it is best left to computers!

6.7.1 Computing semi-partial correlation with more than two independent variables

When there are more than two independent variables, the semi-partial coefficient of correlation can be computed by subtraction as previously. For example, if we are dealing

³ Actually, even the formula for three independent variables are too complex to print them here using only $\sum$ signs. We would need *matrix* algebra to do so because it is particularly well suited for this type of problem.

with three independent variables named X , T and Q , the semi-partial coefficient of correlation between Y and X after having partialled out the effect of T and Q from X is noted $r_{Y \cdot X|TQ}$. It corresponds to the specific explanatory power of X . It is computed from the multiple coefficients of correlation as (cf. Equation 6.20, page 122)

$$r_{Y \cdot X|TQ}^2 = R_{Y \cdot XTQ}^2 - R_{Y \cdot TQ}^2, \quad (6.29)$$

where $R_{Y \cdot XTQ}^2$ is the multiple coefficient of correlation between Y with X , T , and Q , and $R_{Y \cdot TQ}^2$ is the multiple coefficient of correlation between Y with T , and Q .

The semi-partial coefficient of correlation of one independent variable with the dependent variable with the effect of the other two independent variables being partialled out can also be computed directly from the different semi-partial coefficients of correlation of the independent variable and the dependent variable with one independent variable being partialled out. Specifically, the semi-partial correlation of Y and X , with the effect of T and Q partialled out, can be obtained as (cf. Equation 6.23, page 123)

$$r_{Y \cdot X|TQ}^2 = \frac{(r_{Y \cdot X|Q} - r_{Y \cdot T|Q}r_{X \cdot T|Q})^2}{1 - r_{X \cdot T|Q}^2}. \quad (6.30)$$

This recursive procedure can be extended to any number of independent variables (see, e.g. Nunally, 1978, pp. 168–175, or Pedhazur, 1982, pp. 105–110, for more details).

6.7.2 Multicollinearity: a specific problem with non-orthogonal independent variables

A specific problem called *multicollinearity* can occur when using non-orthogonal independent variables. This happens when one independent variable is completely explained by the other independent variables. In other words this independent variable has *no specific* component. When this is the case, the set of independent variables is called *multicollinear*. To see why this is a problem, consider the particular case of a two independent variable design, with X and T being perfectly correlated. Consider the equation of the semi-partial correlation (from Section 6.5.1) of Y and X :

$$r_{Y \cdot X|T}^2 = \frac{(r_{Y \cdot X} - r_{Y \cdot T}r_{X \cdot T})^2}{1 - r_{X \cdot T}^2}. \quad (6.31)$$

The equation of the semi-partial correlation uses the correlation between X and T in the denominator of a fraction. When X and T are perfectly correlated, $r_{X \cdot T}^2$ is equal to 1, and the denominator of Equation 6.31 is equal to zero, and this is where the problem lies.⁴ A similar ‘division by zero’ error will occur for the computation of the slopes. It will also occur with more than two independent variables when at least two of them are perfectly correlated (in this case the quantity being zero is called the ‘determinant’ of the set of independent variables). A subtle version of multicollinearity occurs when one independent variable is *almost* perfectly explained by the other independent variables. Then instead of dividing by zero, we divide by a very small number. Therefore we are likely to reach the limits of the

⁴ Remember, it is a very bad idea to try to divide by zero. Any calculator will give you short shrift!

precision of the computer that we are using, and so the results obtained are likely to be meaningless.

6.8 Bonus: partial correlation

The semi-partial coefficient of correlation can be interpreted as the correlation between the dependent variable and an independent variable after the effect of a second independent variable has been removed (i.e. partialled out) from the first independent variable. The semi-partial coefficient of correlation can be obtained as the correlation between the dependent variable and the residual of the prediction of the first independent variable by the second independent variable. It can also be computed directly using a formula (cf. Section 6.5.1, Equation 6.23, page 123) involving only the coefficients of correlation between pairs of variables. This formula is reprinted below as a refresher:

$$r_{Y \cdot X|T}^2 = \frac{(r_{Y \cdot X} - r_{Y \cdot T} r_{X \cdot T})^2}{1 - r_{X \cdot T}^2}.$$

When dealing with a set of dependent variables, we sometimes want to evaluate the correlation between two dependent variables after the effect of a third dependent variable has been removed from *both* dependent variables. This can be obtained by computing the coefficient of correlation between the residuals of the prediction of each of the first two dependent variables by the third dependent variable (i.e. if you want to eliminate the effect of say variable Q from variables Y and W , you first predict Y from Q and W from Q , and then you compute the residuals and correlate them). This coefficient of correlation is called a *partial coefficient of correlation*.⁵ It can also be computed directly using a formula involving only the coefficients of correlation between pairs of variables. As an illustration, suppose that we want to compute the square partial coefficient of correlation between Y and X after having eliminated the effect of T from both of them (this is done only for illustrative purposes because X and T are independent variables, not dependent variables). This coefficient is noted $r_{(Y \cdot X)|T}^2$ (read ‘ r square of Y and X after T has been partialled out from Y and X) and is computed as

$$r_{(Y \cdot X)|T}^2 = \frac{(r_{Y \cdot X} - r_{Y \cdot T} r_{X \cdot T})^2}{(1 - r_{Y \cdot T}^2)(1 - r_{X \cdot T}^2)}. \quad (6.32)$$

For our example, taking into account that

- $r_{X \cdot T} = .7500$,
- $r_{Y \cdot X} = .8028$, and
- $r_{Y \cdot T} = .9890$,

we find the following values for the partial correlation of Y and X :

$$r_{(Y \cdot X)|T}^2 = \frac{(r_{Y \cdot X} - r_{Y \cdot T} r_{X \cdot T})^2}{(1 - r_{Y \cdot T}^2)(1 - r_{X \cdot T}^2)} = \frac{(.8028 - .9890 \times .7500)^2}{(1 - .9890^2)(1 - .7500^2)} \approx .3894. \quad (6.33)$$

⁵ Do you see the reason for the ‘semi’ in semi-partial correlation now?

Chapter summary

6.9 Key notions of the chapter

Below are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

Non-additivity of the coefficients of correlation when the independent variables are not orthogonal.	Increment of explanation when a variable is entered last in the regression equation.
Specific or unique contribution of an independent variable.	Partial correlation.
The residual of the prediction of one variable by another one gives the <i>specific</i> component of the first variable (because it cannot be explained by the second variable).	Multiple regression with orthogonal independent variables is easy. Multiple regression with non-orthogonal independent variables is a mess (but we cannot always choose to have orthogonal independent variables!)
Semi-partial correlation (synonym: part correlation).	Multicollinearity.

6.10 New notations

Below are the new notations introduced in this chapter. Test yourself on their meaning.

$$\hat{X}_T, \hat{T}_X, a_{X.T} \text{ and } b_{X.T}, a_{T.X} \text{ and } b_{T.X}, e_{T.X}, e_{X.T}$$

$$\hat{X}_T = a_{X.T} + b_{X.T}T, \quad (6.34)$$

$$\hat{T}_X = a_{T.X} + b_{T.X}X. \quad (6.35)$$

$$r_{Y.X|T}^2, r_{Y.T|X}^2, r_{Y.X|TQ}^2, F_{Y.X|T}, F_{Y.T|X}.$$

6.11 Key formulas of the chapter

Below are the main formulas introduced in this chapter: try to go through them and understand what they mean.

When X are not orthogonal:

$$r_{Y.X}^2 + r_{Y.T}^2 \neq 1$$

$$r_{Y.X}^2 + r_{Y.T}^2 \neq R_{Y.XT}^2$$

$$\hat{X}_T = a_{X \cdot T} + b_{X \cdot T} T ,$$

$$\hat{T}_X = a_{T \cdot X} + b_{T \cdot X} X .$$

$$r_{Y \cdot X|T}^2 = r_{Y \cdot e_{X \cdot T}}^2 = \frac{(SCP_{Ye_{X \cdot T}})^2}{SS_Y SS_{e_{X \cdot T}}} .$$

$$r_{Y \cdot X|T}^2 = R_{Y \cdot XT}^2 - r_{Y \cdot T}^2 .$$

$$\text{Part of } Y \text{ explained by both } X \text{ and } T = R_{Y \cdot XT}^2 - (r_{Y \cdot X|T}^2 + r_{Y \cdot T|X}^2) .$$

Variance of Y = variance explained specifically by X

+ variance explained specifically by T

+ variance explained by the common portion of X and T

+ variance left unexplained by X and T .

$$r_{Y \cdot X|T}^2 = \frac{(r_{Y \cdot X} - r_{Y \cdot T} r_{X \cdot T})^2}{1 - r_{X \cdot T}^2}$$

$$r_{Y \cdot T|X}^2 = \frac{(r_{Y \cdot T} - r_{Y \cdot X} r_{X \cdot T})^2}{1 - r_{X \cdot T}^2}$$

$$F_{Y \cdot X|T} = \frac{r_{Y \cdot X|T}^2}{1 - R_{Y \cdot XT}^2} \times (S - K - 1) \quad (6.36)$$

$$F_{Y \cdot T|X} = \frac{r_{Y \cdot T|X}^2}{1 - R_{Y \cdot XT}^2} \times (S - K - 1)$$

$$r_{Y \cdot X|TQ}^2 = R_{Y \cdot XTQ}^2 - R_{Y \cdot TQ}^2 ,$$

$$r_{Y \cdot X|TQ}^2 = \frac{(r_{Y \cdot X|Q} - r_{Y \cdot T|Q} r_{X \cdot T|Q})^2}{1 - r_{X \cdot T|Q}^2} .$$

$$r_{(Y \cdot X)|T}^2 = \frac{(r_{Y \cdot X} - r_{Y \cdot T} r_{X \cdot T})^2}{(1 - r_{Y \cdot T}^2)(1 - r_{X \cdot T}^2)} .$$

6.12 Key questions of the chapter

Below are some questions about the content of this chapter. All the answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- * Why is it important that the independent variables be orthogonal?
- * How do we define the specific or unique contribution of an independent variable to the explanation of a dependent variable?
- * What are the consequences of non-orthogonality of the independent variables?
- * What is the difference between a standard correlation, a semi-partial correlation, and a partial correlation? In what case are they equal to each other?
- * Why is multicollinearity a problem?

7

ANOVA, one factor: intuitive approach and computation of F

7.1 Introduction

In Chapters 4 to 6 we have explored regression. This technique is used to assess the effect of one or several quantitative independent variables on one dependent variable. Very often, psychologists design experiments with nominal independent variables. For example, any experiment which involves *different* groups manipulates a nominal independent variable. The levels of the nominal independent variable correspond to the different groups (i.e. each level 'gives the name' of the group). The technique used to analyze these designs is called *analysis of variance* and is often abbreviated by the acronym ANOVA. Even though the name is analysis of *variance*, the goal of this method is to compare the *means* of different groups and to evaluate the null hypothesis stating that these means come from the same population (i.e. all the population means are equal and the differences that we observe are due to chance). The essential idea behind the technique is to compare the variability of the experimental means to the error variability (hence the appellation analysis of *variance*). We present first the classical way of introducing analysis of variance. We show later on that analysis of variance can also be seen as a special case of regression analysis with some astute coding.

7.2 Intuitive approach

7.2.1 An example: mental imagery

Suppose that we are interested in the effect of mental imagery on memory. Specifically, we think that it is easier to remember concrete rather than abstract information and therefore we predict that it should be easier to remember something if we can transform it into an image. In order to test this idea, we decide to run an experiment.

Because we want to measure memory, we decide to use a task of memorization of pairs of words (we could have used other tasks). In these tasks, subjects are asked to learn pairs

of words (e.g. ‘beauty–carrots’). Then, after some delay, they are asked to give the second word (e.g. ‘carrot’) as the answer to the first word in the pair (e.g. ‘beauty’). In this task, it is common to count the number of pairs correctly recalled by each subject and to use that number as the dependent variable. We decide to have two groups in this experiment: the *experimental* group (in which subjects learn using imagery), and the *control* group (in which subjects learn without using imagery). Subjects in the experimental group will be given the following instructions:

A good way to memorize pairs of words is to try to picture them and then to make the pictures corresponding to how the words interact together. For example, if you are asked to learn the pair ‘piano–cigar’, you can try to picture a piano whose keys are teeth that allow it to bite a cigar and smoke it.

Subjects in the control group will be asked to ‘learn these pairs of words the best they can’. (What is the independent variable? What are its levels?) The performance of subjects is measured by testing their memory for 20 word pairs, 24 hours after learning.

The main objective of this experiment is to demonstrate an effect of the independent variable (learning with imagery *vs* learning without) on the dependent variable (number of word pairs correctly recalled 24 hours after learning). Suppose that we happen to have ten subjects available. We randomly assign five subjects to the experimental condition (these subjects constitute the experimental group), and we randomly assign the other five subjects to the control condition (these subjects constitute the control group). Suppose, now, that we have run the experiment and that we have obtained the results given in Table 7.1 (for each subject of each group, the number of word pairs correctly recalled is given).

A good way of making sense of the results is to make a picture. Figure 7.1 shows the results as a histogram, which is easier to read than a table of numbers. The control group scores can be seen to vary from 1 to 6 with a mean of 4, while the experimental scores vary from 8 to 14 with a mean of 10. When all scores are considered together, the range is broader, from 1 to 14, and the grand mean is 7. If you were told nothing more than the group in which a subject had been placed and were asked to guess, or estimate, the subject’s score, the best estimate would be the *group mean*. If the subject group assignment was unknown, the best estimate of a subject’s score would be the *grand mean*.

7.2.2 An index of effect of the independent variable

By looking at Figure 7.1, we can see that the independent variable seems, indeed, to affect the dependent variable. But how can we be assured that this opinion is well founded? Could the pattern of results seen here be attributed to chance? The problem comes from the fact that

Control group	Experimental group
1	8
2	8
5	9
6	11
6	14

Table 7.1 (Fictitious and nice) results of the experiment on the effects of imagery on memorization.

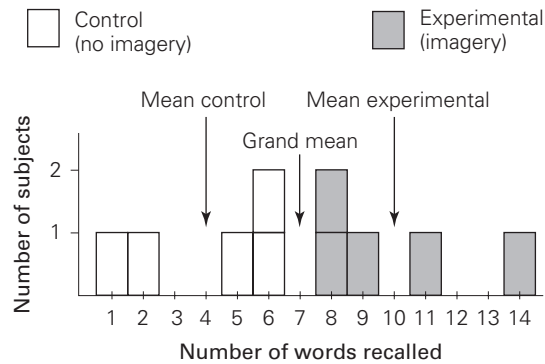


Figure 7.1 Results of the experiment on the effects of imagery on memorization.

random fluctuations always exist and so it is always possible that the observed differences are due to chance variations rather than to the independent variable. So if you try to convince yourself or (and this is much harder) a colleague, you need to have a way of showing that the pattern of results you have obtained cannot be attributed only to chance. In other words, we need to have a way to assess the effect of the independent variable taking into account the effect that can be attributed to random fluctuations. To do that, we will try to construct an index reflecting the strength of the experimental effect.

The first step is to look at the differences between the means of the two groups (the mean of the control group is 4 and the mean of the experimental group is 10). Here, there is only one difference, namely $10 - 4$. However, with more than two groups we would have several differences. To keep the discussion general, the term differences instead of difference is used in what follows. Those differences express the variability between groups, and can be attributed to two possible sources:

- On the one hand, the effect of the random fluctuations (i.e. errors, sampling fluctuations, individual differences of subjects, etc.). This is called the *experimental error*.
- On the other hand, the (possible) effect of the independent variable.

So,

$$\begin{aligned} \text{between-group variability} &= [(\text{possible}) \text{ effect of independent variable}] \\ &\quad + [\text{experimental error}]. \end{aligned}$$

Let us now examine another source of variability in our experiment. The subjects within one group do not necessarily all produce the same performance. Because the subjects in one group are in the same experimental condition, the differences observed between subjects within a given group cannot reflect the effect of the independent variable. They can reflect only the experimental error. Therefore, the within-group variability gives an estimation of the experimental error, and

$$\text{within-group variability} = \text{experimental error}.$$

These two measures of variability (i.e. within-group vs between-group) express different aspects of the same data: differences between group means on the one hand; differences between subjects in the same experimental condition on the other hand. From these two different measures of variability it is possible to build an index reflecting the importance of the

(possible) effect of the independent variable relative to the error. A first possible idea would be to subtract one score (the within-group variability) from the other (the between-group variability). But this score would depend on the unit of measurement used (e.g. it will not be the same if, say, we measure the dependent variable in inches or in feet). As we have seen in Chapter 2 (Section 2.3.6, page 26), a standard way of getting rid of the unit of measurement is to divide a score in one unit by a measure using the same unit of measurement. Taking that into account, the index should be a ratio. It is called F (this is the same Fisher's F as we have seen before under a third disguise) and it is computed as

$$F = \frac{\text{variability between groups}}{\text{variability within groups}} \\ = \frac{\text{effect of independent variable} + \text{error}}{\text{error}}$$

The denominator and the numerator of F show the effect of the independent variable on the dependent variable. If there is no effect of the independent variable, then the F ratio will be equivalent to dividing one estimation of the error (based on the between-group variability) by another estimation of the error (based on the within-group variability). Note that we have two different estimations of the same error. So, when there is no effect, F becomes

$$F = \frac{\text{estimation of experimental error (between groups)}}{\text{estimation of experimental error (within groups)}}$$

Because the denominator and the numerator of this ratio estimate the same quantity (i.e. the error), we can expect to obtain an F close to 1. However, it is still possible for the F ratio to be different from the expected value of 1. To help you understand this problem, take a coin and toss it 50 times. The number of heads gives the probability of obtaining a head with that coin. Repeat that experiment. You obtain, then, another estimation of the probability of obtaining heads with that coin. Both values will be close to each other, but, in general, they will be different. Hence two independent estimations of the same (unknown) parameter are likely to be different (but they will be close to each other most of the time).

On the other hand, if the independent variable has an effect on the dependent variable the F ratio becomes

$$F = \frac{\text{Effect of independent variable} + \text{experimental error}}{\text{experimental error}}$$

Then, we can expect to obtain a value greater than 1, and the greater the effect of the independent variable the larger F will be. In order to compute this F ratio, we need to have a workable and precise definition of the notion of variability. This is done in the next section.

7.3 Computation of the F ratio

After the intuitive presentation, we need now to have a more formal look at the analysis of variance. First, let us start with some notations.

7.3.1 Notation, etc.

We have already seen that an independent variable was denoted by an upper-case cursive letter. In general, alphabetical order will be followed in choosing the letters. So, if the design

has one independent variable, this variable will be called $\mathcal{A}$. The number of levels of $\mathcal{A}$ will be denoted A (same name, but with a different font). If we want to talk about any given level of $\mathcal{A}$, it will be denoted by the lower case letter a .

In most examples in this book, the different experimental groups will have the same number of subjects. In that case, the experimental design is said to be *balanced*. The number of subjects per group will be denoted S (note that S is the number of subjects in a group; the total number of subjects in the experiment is given by $A \times S$). The subject factor will be denoted by $\mathcal{S}$. If we want to talk about any given subject without being specific we will use the letter s .

Using this notation, the score of subject s in group a is

$$Y_{as}$$

(note that the subscripts are written following alphabetical order).

A *condensed notation* will be used to denote different sums. The sum of scores of all the subjects in group a will be

$$\sum_{s=1}^S Y_{as} = Y_{a\cdot}$$

With this notation, a dot replaces the subscript denoting the variable over which the sum is computed.¹ Incidentally, some texts use a slightly different convention and put a plus sign instead of a dot in the summation subscript. (This convention can make things more readable on the blackboard too.) With that convention, the sum of the scores of all the subjects in group a will be

$$\sum_{s=1}^S Y_{as} = Y_{a\cdot} = Y_{a+}$$

This text uses the dot notation rather than the $+$ notation.

The sum of the scores of all the subjects for all the groups is given by

$$\sum_{a=1}^A \sum_{s=1}^S Y_{as} = Y_{..}$$

The same notation serves for denoting the different means. A mean is symbolized by the letter M . For example, the mean of group a will be denoted

$$M_{a\cdot} = \frac{Y_{a\cdot}}{S} = \frac{1}{S} Y_{a\cdot}$$

(note that the ‘dot’ pattern is the same for the mean $M_{a\cdot}$ as for the $Y_{a\cdot}$).

The ‘grand mean’ (i.e. the mean of all the scores) is

$$M_{..} = \frac{Y_{..}}{A \times S} = \frac{1}{AS} Y_{..}$$

To illustrate this notation, Table 7.2 gives the results of the experiment about ‘the effects of imagery on memory’ described earlier in the chapter. There are two groups and ten subjects

¹ If you need a culprit, this notation is attributed to Einstein!

Control: a_1	Experimental: a_2
$Y_{1,1} = 1$	$Y_{2,1} = 8$
$Y_{1,2} = 2$	$Y_{2,2} = 8$
$Y_{1,3} = 5$	$Y_{2,3} = 9$
$Y_{1,4} = 6$	$Y_{2,4} = 11$
$Y_{1,5} = 6$	$Y_{2,5} = 14$
$Y_{1.} = 20$	$Y_{2.} = 50$
$M_{1.} = \frac{Y_{1.}}{S} = \frac{20}{5} = 4$	$M_{2.} = \frac{Y_{2.}}{S} = \frac{50}{5} = 10$
$Y_{..} = 70$	
$M_{..} = \frac{Y_{..}}{A \times S} = \frac{70}{10} = 7$	

Table 7.2 Results from an experiment on ‘the effect of imagery on memorization’ (see text for explanation).

participating in the experiment (five subjects per group). Therefore $A = 2$ and $S = 5$. $\mathcal{A}$ is the independent variable and it has two levels:

- use of imagery to learn word pairs;
- no use of imagery.

7.3.2 Distances from the mean

Figure 7.2 reproduces these data with an illustration of the relationship between a single score ($Y_{a,s}$) and the group and grand means ($M_{a.}$ and $M_{..}$, respectively). This figure shows an essential point for the analysis of variance that we detail below. We start by looking at the grand mean ($M_{..}$) which is equal to 7. And let us take any score, for example, subject 1 in group 1 who has a score of 1. The distance from that subject to the grand mean is:

$$Y_{1,1} - M_{..} = 1 - 7 = -6.$$

Figure 7.2 shows that the distance to the grand mean ($Y_{1,1} - M_{..}$) is made of two parts:

- The distance of subject 1 to the mean of group 1 ($Y_{1,1} - M_{1.}$).
- The distance from the mean of group 1 to the grand mean ($M_{1.} - M_{..}$).

Or, with a formula:

$$(Y_{1,1} - M_{..}) = (Y_{1,1} - M_{1.}) + (M_{1.} - M_{..}) = (1 - 4) + (4 - 7) = -3 + (-3) = -6.$$

Using the formula for the general case, this is written as (in the above example $a = 1$, $s = 1$)

$$(Y_{as} - M_{..}) = (Y_{as} - M_{a.}) + (M_{a.} - M_{..}).$$

Note that this relation can also be proved using elementary computations:

$$(Y_{as} - M_{a.}) + (M_{a.} - M_{..}) = Y_{as} - M_{a.} + M_{a.} - M_{..} = Y_{as} - M_{..}.$$

The first term of this partition,

$$Y_{as} - M_{a.},$$

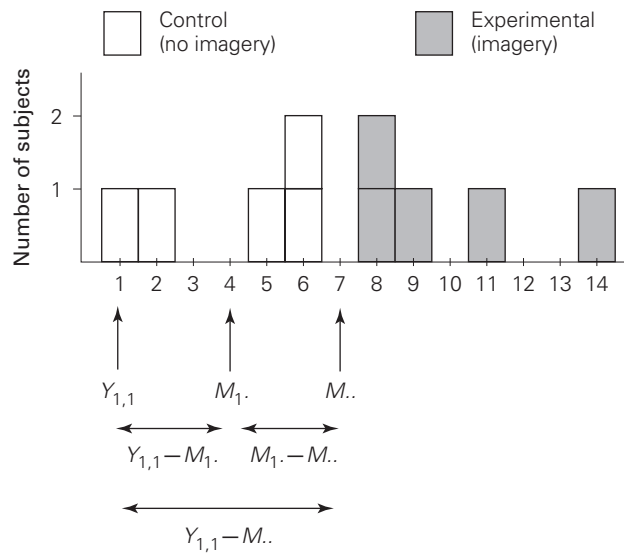


Figure 7.2 Results of the experiment on the effects of imagery on memorization: decomposition of the deviation from the grand mean into the within group and between group deviations.

is the difference between the score of a given subject and the mean of the group she or he belongs to. It is called the *distance* or *deviation within groups*.

The second term of this partition,

$$M_{a.} - M_{..} ,$$

is the difference between the subject’s group means and the grand mean. It is called the *distance* or *deviation between groups*.

Let us come back now to the index that we are trying to design to ‘reflect’ the effect of the independent variable on the dependent variable. That index was expressed as the ratio of the between-group variability to the within-group variability. The partition of the deviation from the grand mean into two components reflecting these two sources of variability for each score is a first step toward quantification. The problem now is to take into account the variability of all the scores. An obvious way is to take the sum, but the sum of deviations from the mean will always be zero. Intuitively, this is because the mean is at the ‘center’ of the scores, so there are as many scores greater than the mean as there are scores smaller. A classic way of avoiding this problem is to square each of the deviations before summing them. This approach is, indeed, very similar to the first step in the computation of the variance of a sample. A brief reminder about this important notion follows—you can skip it if it is still vivid in your memory.

7.3.3 A variance refresher

Recall that in elementary statistics, the classic measure of dispersion of the scores in a sample is called the *variance*. It is obtained by computing, first, the sum of squares of the deviations (i.e. distances) of the scores from the mean of the sample, and second, by dividing that sum by the number of independent scores of the sum (i.e. the ‘degrees of freedom’). If the number of scores is N , and their mean is M_Y , then the variance, noted $\hat{\sigma}^2$ (some other notations are

used, the most common are: $\hat{s}^2$, s^2) is given by

$$\hat{\sigma}^2 = \frac{\sum (Y - M_Y)^2}{N - 1}.$$

Remember that we use N to represent the number of ‘something’. However, when the ‘something’ is subjects we will use S .

The numerator of the previous equation is called the sum of the squared deviations or *sum of squares*. It is denoted by SS . The denominator is called the number of *degrees of freedom* of the sum of squares. It is denoted by df . In this case, it is equal to $N - 1$, because the value of the mean is needed in order to compute the sum of squares. Knowing the mean of a set of scores is equivalent to knowing its sum (since the mean is the sum divided by the number of scores). As a consequence, only the first $N - 1$ values can be free to vary (because the last value is constrained to make the sum equal to its known value). For example, suppose that we have 6 scores whose sum is 30. Knowing that the first 5 scores are 1, 1, 1, 1, 1 implies that the last one must be 25 [$25 = 30 - (1 + 1 + 1 + 1 + 1)$]. The notion of degrees of freedom will be detailed later on (See Section 7.3.6, page 140).

With this notation, the previous equation can be rewritten as

$$\hat{\sigma}^2 = \frac{SS}{df}.$$

This formula shows that the variance is a sort of average. Specifically, it is the mean value of the squared deviations of scores from the mean of the sample.

The sum of squares expresses the variability of the scores around the mean. Dividing by the degrees of freedom has the function of scaling the sum of squares, so that different sums of squares (i.e. computed with samples of different sizes) can be compared.

7.3.4 Back to the analysis of variance

From the example of the variance, it seems straightforward to generalize, and measure the variability by computing sums of squared deviations, and then divide them by their degrees of freedom (in order to compare them).

From the previous derivation, we have three deviations for each score:

- deviation of a score from the grand mean ($Y_{as} - M_{..}$),
- deviation of a score from the mean of its group ($Y_{as} - M_{a.}$),
- deviation of the mean of the group from the grand mean ($M_{a.} - M_{..}$),

and we can now put them into a single equation:

$$(Y_{as} - M_{..}) = (Y_{as} - M_{a.}) + (M_{a.} - M_{..}).$$

If each of these deviations is squared and summed that will give three sums of squares.

We start with the deviations of a score from the grand mean. The sum of these squared deviations is the *total sum of squares*. It is noted SS_{total} and computed as

$$SS_{\text{total}} = \sum_a \sum_s (Y_{as} - M_{..})^2.$$

The deviations of a score from the mean of its group will give the *sum of squares within groups*, noted SS_{within} , and computed as

$$SS_{\text{within}} = \sum_a \sum_s (Y_{as} - M_{a.})^2.$$

The deviations of the mean of the group to which a score belongs from the grand mean gives the *sum of squares between groups*, noted SS_{between} and computed as

$$SS_{\text{between}} = \sum_a \sum_s (M_{a\cdot} - M_{..})^2.$$

Now, since the deviation of the group means $M_{a\cdot}$ from the grand mean $M_{..}$ is the same for all subjects in a given group, we can just multiply the between-group deviation by the number of subjects S in order to obtain the total for this group, thus we can bring S outside the summation sign. We obtain, then,

$$SS_{\text{between}} = \sum_a \sum_s (M_{a\cdot} - M_{..})^2 = S \sum_a (M_{a\cdot} - M_{..})^2.$$

The next paragraph derives the fundamental equation of the analysis of variance, based on the notion that the total sum of squares is equal to the sum of squares within groups plus the sum of squares between groups.

7.3.5 Partition of the total sum of squares

The total sum of squares is computed as

$$SS_{\text{total}} = \sum_a \sum_s (Y_{as} - M_{..})^2.$$

But as seen previously, the deviation of a score from the grand mean can be partitioned into two parts as

$$(Y_{as} - M_{..}) = (Y_{as} - M_{a\cdot}) + (M_{a\cdot} - M_{..}).$$

Hence,

$$SS_{\text{total}} = \sum_a \sum_s (Y_{as} - M_{..})^2 = \sum_a \sum_s [(Y_{as} - M_{a\cdot}) + (M_{a\cdot} - M_{..})]^2.$$

By expanding the square of the terms within brackets on the right-hand side, the equation becomes

$$SS_{\text{total}} = \sum_a \sum_s [(Y_{as} - M_{a\cdot})^2 + (M_{a\cdot} - M_{..})^2 + 2(Y_{as} - M_{a\cdot})(M_{a\cdot} - M_{..})].$$

We can then distribute the summations to each term (for revision of the sum sign $\sum$ see Appendix B) in order to obtain

$$\begin{aligned} SS_{\text{total}} &= \sum_a \sum_s (Y_{as} - M_{a\cdot})^2 + \sum_a \sum_s (M_{a\cdot} - M_{..})^2 \\ &\quad + 2 \sum_a \sum_s (Y_{as} - M_{a\cdot})(M_{a\cdot} - M_{..}). \end{aligned}$$

A very useful feature for simplifying this equation is that we can show (the proof is given in the next paragraph):

$$2 \sum_a \sum_s (Y_{as} - M_{a\cdot})(M_{a\cdot} - M_{..}) = 0.$$

On the basis of that result, the total sum of squares is equal to

$$\begin{aligned}
 SS_{\text{total}} &= \sum_a \sum_s (Y_{as} - M_{a\cdot})^2 + \sum_a \sum_s (M_{a\cdot} - M_{..})^2 \\
 &= \sum_a \sum_s (Y_{as} - M_{a\cdot})^2 + S \sum_a (M_{a\cdot} - M_{..})^2 \\
 &= \begin{array}{cc} \text{Sum of squares} & + \text{Sum of squares} \\ \text{within groups} & \text{between groups} \end{array} \\
 &= SS_{\text{within}} + SS_{\text{between}} . \tag{7.1}
 \end{aligned}$$

7.3.5.1 Proof of the additivity of the sum of squares

In this paragraph it is proved that

$$2 \sum_a \sum_s (Y_{as} - M_{a\cdot})(M_{a\cdot} - M_{..}) = 0 .$$

You can skip the following proof for the first reading and come back to it later. The first step of the proof is to develop the term

$$\sum_a \sum_s (Y_{as} - M_{a\cdot})(M_{a\cdot} - M_{..}) .$$

It becomes

$$\sum_a \sum_s (Y_{as}M_{a\cdot} - Y_{as}M_{..} - M_{a\cdot}^2 + M_{a\cdot}M_{..}) .$$

We can ‘distribute’ the sum sign $\sum$, which gives

$$(\sum_a \sum_s Y_{as}M_{a\cdot}) - (\sum_a \sum_s Y_{as}M_{..}) - (\sum_a \sum_s M_{a\cdot}^2) + (\sum_a \sum_s M_{a\cdot}M_{..}) .$$

And after simplification the expression is

$$(\sum_a M_{a\cdot} \sum_s Y_{as}) - (M_{..} \sum_a \sum_s Y_{as}) - (S \sum_a M_{a\cdot}^2) + (SM_{..} \sum_a M_{a\cdot}) .$$

But,

$$\sum_s Y_{as} = SM_{a\cdot} .$$

If we make this substitution and carry out the additions and subtractions, we find that the result is zero.

$$(S \sum_a M_{a\cdot}^2) - (SM_{..} \sum_a M_{a\cdot}) - (S \sum_a M_{a\cdot}^2) + (SM_{..} \sum_a M_{a\cdot}) = 0 ,$$

which was what we wanted to prove.

It is worth noting that the term

$$\sum_a \sum_s (Y_{as} - M_{a\cdot})(M_{a\cdot} - M_{..}) ,$$

corresponds to a *cross-product* term (remember Chapter 2 on correlation?). In fact, this term would be the numerator of the coefficient of correlation between the within-group and between-group deviations. Because its value is zero, the coefficient of correlation would also be zero. This shows that these two types of deviations are uncorrelated. In other words,

the within-group and the between-group deviations constitute two *independent* sources of variation.

7.3.5.2 Back to the sum of squares

So far we have seen that the total sum of squares can be partitioned into two components:

1. The sum of squares within groups.
2. The sum of squares between groups.

This can be expressed as a formula:

$$SS_{\text{total}} = SS_{\text{within}} + SS_{\text{between}} .$$

The formula for the sums of squares shows that their values are dependent upon the number of elements of which they are composed. In order to be able to compare them, they should be scaled. As with the variance, the scaling factors to use are the degrees of freedom of the sums of squares. Because a sum of squares divided by its degrees of freedom is a kind of average, the result of the division is called a *mean square*. It is denoted MS . There are two sources of variability to assess, hence two mean squares to compute: the mean square within groups and the mean square between groups. Here is their formula:

$$MS_{\text{within}} = \frac{SS_{\text{within}}}{df_{\text{within}}}$$

$$MS_{\text{between}} = \frac{SS_{\text{between}}}{df_{\text{between}}} .$$

We need now to find the number of degrees of freedom of these sums of squares.

7.3.6 Degrees of freedom

There are different ways of explaining degrees of freedom. This is a concept borrowed from physics (like so many other concepts in statistics). In physics it refers to the number of free parameters (hence degrees of *freedom*) of a system, or the number of parameters a physicist can ‘play with’ when dealing with a system. Physicists and statisticians often talk about the ‘dimensionality of a system’ in this context.

In statistics, the number of degrees of freedom of an estimation is the number of elements that can vary freely. For example, suppose we want to estimate the variance of a population from a sample of S observations. Because the variance is computed by squaring the deviations from the scores to the mean, we need first to compute the sample mean (as an estimation of the population mean). Knowing the mean of a set of numbers is equivalent to knowing their sum (because the mean is the sum divided by the number of scores). Now, as the sum of the scores is known, only $S - 1$ scores can freely vary. The value of the last score is constrained by the value of the sum. For example, suppose that we have 5 scores whose sum is 30. Knowing that the first 4 scores are 5, 6, 2, 10 implies that the last one must be 7 [because $7 = 30 - (5 + 6 + 2 + 10)$].

How this notion can be linked to the ‘dimensions’ of a system can be seen from an example. Suppose you draw a graph of the points (20,80), (0,100), (50,50), (99,1), and (100,0). Because of the two coordinates, these points are points on the plane, but if you plot them, they all fall

on the line going from the point (0,100) to the point (100,0). Therefore, even though these points seem to lie in two dimensions, because the sum of their two components is fixed they lie on one line. So, the constraint on the sum of the components leaves the points free to vary in only one dimension instead of two.

Another way of looking at the degrees of freedom of a sum is to realize that the estimation of a parameter of the population is equivalent to losing a degree of freedom. In the previous example, the estimation of the mean of the population necessary to compute the deviations 'costs' 1 degree of freedom. From that, it is possible to derive a general formula for the number of degrees of freedom of a sum:

$$\begin{aligned} df &= \text{number of independent observations} \\ &\quad - \text{number of estimated parameters of the population.} \end{aligned}$$

For example, this formula gives the correct value of $S - 1$ for the degrees of freedom of the sum of squares for the estimation of the variance of a population from a sample.

When applied to the problem of finding the degrees of freedom of the sums of squares for an analysis of variance, this formula will give the number of degrees of freedom for the between-group, within-group, and total sum of squares. The derivations are given in the following paragraphs.

7.3.6.1 Between-group degrees of freedom

The sum of squares between groups is computed with the deviations of the group means from the grand mean. There are A groups, and hence, A means. But the grand mean is the mean of the group means. In other words, in order to compute that sum of squares an estimate of the population grand mean is needed first. That estimation 'costs' one degree of freedom. Hence, the number of degrees of freedom of the sum of squares between is

$$df_{\text{between}} = A - 1 .$$

The mean square between groups is then

$$MS_{\text{between}} = \frac{SS_{\text{between}}}{df_{\text{between}}} = \frac{SS_{\text{between}}}{A - 1} .$$

7.3.6.2 Within-group degrees of freedom

The sum of squares within groups is computed with the deviations of each score from its group mean. There are S observations per group. For each group, in order to compute the sum of squares an estimate of the group mean is needed first. Because there are A groups, there are A parameters to estimate. There are $A \times S$ independent scores. Hence the number of degrees of freedom of the sum of squares within groups is

$$df_{\text{within}} = AS - A = A(S - 1) .$$

The mean square within groups is then

$$MS_{\text{within}} = \frac{SS_{\text{within}}}{df_{\text{within}}} = \frac{SS_{\text{within}}}{A(S - 1)} .$$

7.3.6.3 Total number of degrees of freedom

The total sum of squares is computed with the deviations of each score from the grand mean. There are $A \times S$ independent scores. Estimation of the grand mean ‘costs’ one degree of freedom. Therefore the number of degrees of freedom of the total sum of squares is:

$$df_{\text{total}} = (A \times S) - 1 = N - 1.$$

Remember that

$$SS_{\text{total}} = SS_{\text{within}} + SS_{\text{between}}.$$

The same relationship holds for the degrees of freedom:

$$\begin{aligned} df_{\text{total}} &= (A \times S) - 1 = A \times S - A + A - 1 = A(S - 1) + A - 1 \\ &= df_{\text{within}} + df_{\text{between}}. \end{aligned} \quad (7.2)$$

These two equations are essential for the analysis of variance.

Warning: The sums of squares are additive, which means

$$SS_{\text{total}} = SS_{\text{within}} + SS_{\text{between}}.$$

The degrees of freedom are additive:

$$df_{\text{total}} = df_{\text{within}} + df_{\text{between}}.$$

However, the mean squares *are not additive*. Although the relationship described above holds for the sums of squares and the degrees of freedom, it is not true for mean squares.

$$\frac{SS_{\text{total}}}{df_{\text{total}}} \neq \frac{SS_{\text{within}}}{df_{\text{within}}} + \frac{SS_{\text{between}}}{df_{\text{between}}}.$$

7.3.7 Index F

We are now ready to compute the index F . Recall that it is obtained by dividing the variability between groups by the variability within groups. We just showed that the variability between groups is estimated using the mean square between groups (MS_{between}) and that the variability within groups is estimated using the mean square within groups (MS_{within}). Therefore, the index F , or Fisher’s F is computed as

$$\begin{aligned} F &= \frac{\text{variability between groups}}{\text{variability within groups}} \\ &= \frac{MS_{\text{between}}}{MS_{\text{within}}}. \end{aligned}$$

7.4 A bit of computation: mental imagery

As a way of checking your understanding of the formulas, the computational steps are detailed for the example used in the beginning of the chapter. The data are reprinted here for convenience.

Control: a_1	Experimental: a_2
$Y_{1,1} = 1$	$Y_{2,1} = 8$
$Y_{1,2} = 2$	$Y_{2,2} = 8$
$Y_{1,3} = 5$	$Y_{2,3} = 9$
$Y_{1,4} = 6$	$Y_{2,4} = 11$
$Y_{1,5} = 6$	$Y_{2,5} = 14$
$Y_{1.} = 20$	$Y_{2.} = 50$
$M_{1.} = \frac{Y_{1.}}{S} = \frac{20}{5} = 4$	$M_{2.} = \frac{Y_{2.}}{S} = \frac{50}{5} = 10$
$Y_{..} = 70$	
$M_{..} = \frac{Y_{..}}{A \times S} = \frac{70}{10} = 7$	

Table 7.3 Results from the experiment on 'the effect of imagery on memory' (data reprinted from Table 7.2)

Here are the detailed steps. Try to redo the computations to make sure you understand them.

$$SS_{\text{total}} = (1 - 7)^2 + (2 - 7)^2 + \cdots + (14 - 7)^2 = 138$$

$$SS_{\text{between}} = 5 \times (4 - 7)^2 + 5 \times (10 - 7)^2 = 90$$

$$SS_{\text{within}} = (1 - 4)^2 + \cdots + (14 - 10)^2 = 48.$$

Note: $SS_{\text{total}} = SS_{\text{within}} + SS_{\text{between}} = 48 + 90 = 138.$

$$df_{\text{total}} = A \times S - 1 = 2 \times 5 - 1 = 10 - 1 = 9$$

$$df_{\text{between}} = A - 1 = 2 - 1 = 1$$

$$df_{\text{within}} = A(S - 1) = 2(5 - 1) = 2 \times 4 = 8.$$

Note: $df_{\text{total}} = df_{\text{within}} + df_{\text{between}} = 8 + 1 = 9.$

$$MS_{\text{between}} = \frac{90}{1} = 90$$

$$MS_{\text{within}} = \frac{48}{8} = 6.$$

We can now (at last!) compute the index of effect of the independent variable on the dependent variable.

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}} = \frac{90}{6} = 15.00.$$

The main results of the computations can be condensed in a table (called the *analysis of variance table*).

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Between	1	90.00	90.00	15.00
Within <i>S</i>	8	48.00	6.00	
Total	9	138.00		

Intuitively, it seems clear that the larger the size of the *F* ratio, the stronger the evidence of an effect of the independent variable. The next chapter will try to make that intuition more precise.

Chapter summary

7.5 Key notions of the chapter

Below are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

Mean of a group and grand mean	The different sums of squares (total, between-group, and within-group), and how to compute them
Variability between groups, and variability within groups	
How to obtain an index of effect	The different degrees of freedom (total, between-group, and within-group), and how to compute them
Composition of the <i>F</i> ratio when the independent variable does not have an effect	The different mean squares (between-group and within-group), and how to compute them
Composition of the <i>F</i> ratio when the independent variable has an effect	Variance
Criterion	Parameters and estimation of parameters
Balanced design	
Within-group and between groups deviations	

7.6 New notations

Below are the new notations introduced in this chapter. Test yourself on their meaning.

$\mathcal{A}$:	Name of the independent variable
A :	Number of levels of $\mathcal{A}$
a :	a^{th} level of the independent variable $\mathcal{A}$

$S :$	Name of the factor subject
$S :$	Number of subjects per group
$s :$	s^{th} subject
$Y_{as} :$	Score of subject s of group a
$Y_{a.} = \sum_{s=1}^S Y_{as} :$	Sum of the scores for all the subjects of group a
$Y_{..} = \sum_{a=1}^A \sum_{s=1}^S Y_{as} :$	Sum of the scores for all the subjects for all the groups
$M_{a.} = \frac{Y_{a.}}{S} :$	Mean of group a
$M_{..} = \frac{Y_{..}}{A \times S} :$	Grand mean
$SS_{\text{total}} :$	Total sum of squares
$SS_{\text{between}} :$	Between-group sum of squares
$SS_{\text{within}} :$	Within-group sum of squares
$df_{\text{total}} :$	Total number of degrees of freedom
$df_{\text{between}} :$	Between-group degrees of freedom
$df_{\text{within}} :$	Within-group degrees of freedom
$MS_{\text{between}} = \frac{SS_{\text{between}}}{df_{\text{between}}} :$	Mean square between groups
$MS_{\text{within}} = \frac{SS_{\text{within}}}{df_{\text{within}}} :$	Mean square within groups

7.7 Key formulas of the chapter

Below are the main formulas introduced in this chapter: try to go through them and understand what they mean.

$$SS_{\text{total}} = \sum_a \sum_s (Y_{as} - M_{..})^2 .$$

$$SS_{\text{within}} = \sum_a \sum_s (Y_{as} - M_{a.})^2 .$$

$$SS_{\text{between}} = \sum_a \sum_s (M_{a.} - M_{..})^2 = S \sum_a (M_{a.} - M_{..})^2 .$$

$$SS_{\text{total}} = \sum_a \sum_s (Y_{as} - M_{..})^2 = \sum_a \sum_s (Y_{as} - M_{a.})^2 + S \sum_a (M_{a.} - M_{..})^2 .$$

$$SS_{\text{total}} = SS_{\text{within}} + SS_{\text{between}}.$$

$$df_{\text{between}} = A - 1.$$

$$df_{\text{within}} = AS - A = A(S - 1).$$

$$MS_{\text{between}} = \frac{SS_{\text{between}}}{df_{\text{between}}} = \frac{SS_{\text{between}}}{A - 1}.$$

$$MS_{\text{within}} = \frac{SS_{\text{within}}}{df_{\text{within}}} = \frac{SS_{\text{within}}}{A(S - 1)}.$$

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}}.$$

7.8 Key questions of the chapter

Below are some questions about the content of this chapter. All the answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- * What is the purpose of analysis of variance?
- * What are the different components of a given subject score?
- * How do you show that the total sum of squares is composed of two parts (the sum of squares between groups and the sum of squares within groups)?
- * Why is it preferable to compare mean squares rather than sums of squares?
- * Why is it meaningless to add the mean squares, even though we can add sums of squares and degrees of freedom?

8

ANOVA, one factor: test, computation, and effect size

8.1 Introduction

In this first section we adapt for analysis of variance the notion of test previously introduced for correlation in Chapter 3.

8.2 Statistical test: a refresher

8.2.1 General considerations

The value of the criterion F reflects the effect of the independent variable on the dependent variable. The larger F , the clearer the indication of an effect of the independent variable. But how should we decide, from this indication, if an effect really exists? The problem here is not simply to compute the value of F for the sample, but rather to decide if there is an effect of the independent variable for the whole population of subjects from which our sample was drawn. In other words, if we find a value of F greater than 1 in an experiment, this may be taken as an indication of the effect of the independent variable on the dependent variable. Would this result occur again if the experiment were repeated with another sample of subjects? If yes, then the indication of the effect will appear to be trustworthy and solid. If not, then it can be attributed to the effects of chance.

This idea of repeating an experiment and of predicting the results constitutes the basis of the statistical argument. This amounts to saying: ‘The results of my experiment are solid, trustworthy, robust, etc. If I run this experiment again, using different subjects, the results are likely to be the same’. To sustain this argument, you could actually repeat the original experiment a large number of times (without modifying the experimental conditions) and check that the results always go in the same direction. This may take a lot of time and risk becoming tiresome or running out of available subjects. In order to show that the results of an experiment are replicable, we could *equivalently* show that the results we obtained could not be attributed to chance, and ought therefore to be attributed to the effects of the independent variable.

8.2.2 The null hypothesis and the alternative hypothesis

In statistical terms, this amounts to inferring, from the sample value of F , the existence of an effect of the independent variable on the dependent variable in the whole population of subjects. Or, equivalently, to decide between the two following alternatives:

the observed value of F reflects the effects of chance

or

the observed value of F reflects the effect of the independent variable

In other words, we must choose between two hypotheses describing the population from which our sample of subjects is drawn, in order to ‘explain’ the observed value of F :

- The *null hypothesis* (H_0): the independent variable has no effect on the dependent variable for the population of subjects as a whole. Recall that this is an *exact statistical hypothesis*. It gives a precise value to the intensity of the effect: the effect is null, its intensity is zero.
- The *alternative hypothesis* (H_1): the independent variable has an effect on the dependent variable for the population of subjects as a whole. Recall that this is an *inexact statistical hypothesis*: it does not provide a precise value for the intensity of the effect of the independent variable.

8.2.3 A decision rule: reject H_0 when it is ‘unlikely’

Remember that the general logic behind a statistical test is, in a sense, backward. In order to show that there is an effect of the independent variable on the dependent variable we try to show that it is improbable that there is no effect. More formally, because we cannot estimate the probability of obtaining a given value of F if the alternative hypothesis were true (because H_1 is inexact), we estimate instead the probability of observing the results of the experiment if the null hypothesis were true. If the null hypothesis appears unlikely in terms of the results of the experiment then we decide to reject the null hypothesis. Since the null hypothesis and the alternative hypothesis are contradictory, rejecting the null hypothesis leads to accepting the alternative hypothesis.

8.2.4 Sampling distributions: the distributions of Fisher’s F

Recall that the first step in carrying out a statistical test is to state the null and alternative hypotheses explicitly. These statistical hypotheses always concern the population from which the sample is drawn, not the sample itself.

The next step is to compute the probability of observing the results of the experiment *if the null hypothesis were true*. We say that we specify the *sampling distribution of the test statistic under the null hypothesis*. Precisely, when H_0 is true, the distribution gives the probability associated with each possible value of the test statistic.

As a general rule, and especially here, the null hypothesis is not sufficient for the calculation of the sampling distribution of the test statistic. It is necessary to add ‘*technical assumptions*’ describing the population from which the sample is drawn, called ‘normality’ and ‘homoscedasticity’ (see Chapter 11 for more). These terms mean that the population distributions involved in the experiment are normal and have the same variance.

When these technical assumptions are satisfied, *and* when the null hypothesis is true, then we can derive the sampling distribution of the test statistic F . This work, initiated by Student and developed by Fisher, was completed by Fisher's student Snedecor who named the statistic F_{Fisher} in honor of his mentor. Recall that there is not just one distribution but a family of distributions of Fisher's F . The number of degrees of freedom of the numerator (the between-groups mean square) and of the denominator (the within-groups mean square) give the *parameters* of the sampling distribution, ν_1 and ν_2 .

When the probability associated with F is small, we decide to reject the (unlikely) null hypothesis, and we accept instead the alternative hypothesis. We decide that the independent variable has an effect on the dependent variable.

8.2.5 Region of rejection, region of suspension of judgment, critical value

Let us summarize the rule for statistical decisions. We decide to reject H_0 when the probability associated with F is less than the fixed threshold α . We thus separate the values of F into two regions:

1. Rejection region: values of F associated with a probability less than or equal to the α level. If the calculated F falls in this region, we reject H_0 and consequently accept H_1 .
2. Region of suspension of judgment: value of F associated with a probability greater than the α level. If the calculated F falls in this region we cannot reject H_0 and suspend judgment.

The value of F that divides these two regions is called the *critical value*.

- If the F calculated for the experimental sample is greater than or equal to F_{critical} , its associated probability is less than α . We therefore decide to reject the null hypothesis.
- If the F calculated for the experimental sample is less than F_{critical} , its associated probability is greater than α . We therefore suspend judgment; we neither accept nor reject the null hypothesis. Strictly speaking, we decide not to decide.

8.3 Example: back to mental imagery

We now return to the example used in the previous chapter in which the problem was to explore the relationship between imagery and memory. Ten subjects were divided into two groups of five subjects each: the control group and the experimental group. The dependent variable was the number of pairs of words correctly recalled by the subjects.

For this example, we have already computed the different components of the ANOVA in the previous chapter. The six steps of the statistical test described in Chapter 3 are detailed below:

1. Statistical hypothesis. There are several (equivalent) ways to express the null hypothesis, H_0 . Some of them are detailed below:

- The independent variable has no effect on the dependent variable.
- The experimental manipulation has no effect on the subjects' behavior.
- Imagery has no influence on memory.

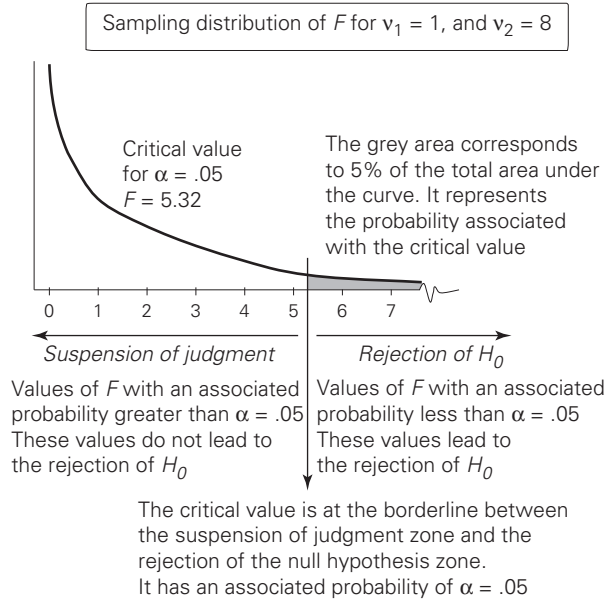


Figure 8.1 The Fisher F distribution for $v_1 = 1$ and $v_2 = 8$ degrees of freedom. Detailed test procedure for $\alpha = .05$.

- The experimental group and the control group do not really differ in the number of pairs of words correctly recalled. The apparent differences are only due to chance.
- For the whole population of subjects, the average number of syllables remembered in the condition ‘learning with image’ is the same as the average number of syllables remembered in the condition ‘learning without image’.

There are several (equivalent) ways to express the alternative hypothesis, H_1 . Some of them are detailed below:

- The independent variable has an effect on the dependent variable.
- The experimental manipulation does change the subjects’ behavior.
- Imagery improves memory.
- The experimental group and the control group do differ for the number of pairs of words correctly recalled.
- For the whole population of subjects, the average number of syllables remembered in the condition ‘learning with image’ is larger than the average number of syllables remembered in the condition ‘learning without image’.

2. Criterion. As explained previously, F reflects the effect of the independent variable on the dependent variable.

3. Alpha level. In this example, because we have no compelling reason to chose a specific value and because this will be a good opportunity to use the Table of Fisher’s F we will just follow tradition and use the classic α levels of $\alpha = .01$ and $\alpha = .05$.

4. Sampling distribution. If the null hypothesis is true, and if the ‘technical assumptions’ of *normality* and *homoscedasticity* are valid then the sampling

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Between	1	90.00	90.00	15.00
Within S	8	48.00	6.00	
Total	9	138.00		

Table 8.1 The ANOVA table for the experiment 'effect of imagery on memory' (see text for explanation).

distribution of F is a Fisher distribution with

$$\nu_1 = A - 1 = 1 \text{ df}$$

and

$$\nu_2 = A(S - 1) = 2 \times 4 = 8 \text{ df}.$$

5. Decision rule, region of rejection and critical values. If we can compute directly (i.e. with computer program) the probability of F , the decision rule is straightforward: reject the null hypothesis when the probability associated with F is smaller than the chosen alpha level.

If we do not have access to a computer, we can use the region of rejection and critical value approach. The region of rejection corresponds to the values of the F ratio larger than the critical values obtained from the table for:

$$\nu_1 = A - 1 = 1 \text{ df}$$

and

$$\nu_2 = A(S - 1) = 8 \text{ df}.$$

- For $\alpha = .05$, $F_{\text{critical}} = 5.32$, and we reject the null hypothesis and accept the alternative hypothesis when F is larger than or equal to the critical value $F_{\text{critical}} = 5.32$.
- For $\alpha = .01$, $F_{\text{critical}} = 11.26$, and we reject the null hypothesis and accept the alternative hypothesis when F is larger than or equal to the critical value $F_{\text{critical}} = 11.26$.

This decision rule is illustrated by Figure 8.1 for $\alpha = .05$. Figure 8.2 illustrates the fact that all Fisher sampling distributions do not have the same shape.

6. Results and decision. The results are summarized in Table 8.1. The details of the computation were presented in Chapter 7.

A statistical computer program gave a value of .00048 for $\Pr(F)$ which is the probability associated with $F = 15.00$. This probability being less than the significance level, $\alpha = .01$, we can reject the null hypothesis because it is unlikely to be true taking into account the experimental results.

Or, if you prefer to use the critical value approach,

- for $\alpha = .05$: $F = 15.00$ larger than $F_{\text{critical}} = 5.32$;
- for $\alpha = .01$: $F = 15.00$ larger than $F_{\text{critical}} = 11.26$;

Hence, we reject H_0 and accept H_1 for both α levels (.05 and .01).

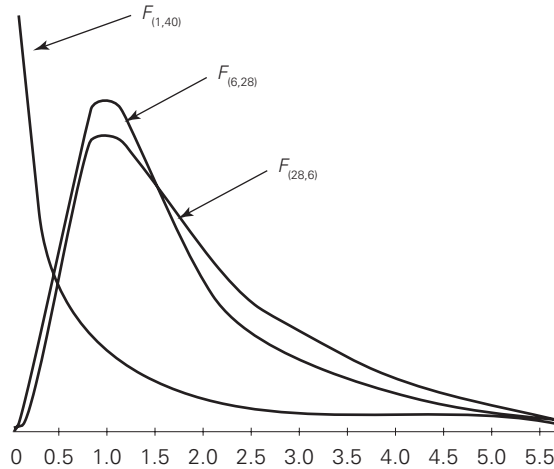


Figure 8.2 Examples of Fisher sampling distributions.

8.4 Another more general notation: $\mathcal{A}$ and $\mathcal{S}(\mathcal{A})$

As the ANOVA table indicates, the analysis of variance of a one-factor between-subject design isolates two sources of variability (or sources of variation) out of the total variability: variability between and within groups. Up to now we have abbreviated these sources by the subscripts ‘between’ and ‘within’ to designate sums of squares, degrees of freedom, and mean squares. This notation, which is clear and easy to remember for designs with one factor, becomes unwieldy as soon as we take up the study of more complex factorial designs with more than one independent variable. Because of this, now that you are familiar with the basic procedures for analysis of variance, we can use a more general notation which will facilitate the transition from simple to complex designs.

The *between*-group variability will correspond to the effects of the independent variable symbolized by $\mathcal{A}$. We shall use the letter $\mathcal{A}$ as an index instead of the term ‘between’.

Similarly, the *within*-group variability will correspond in some sense to the effect of the ‘subjects’ factor $\mathcal{S}$. But there is a twist here and we cannot use the notation $\mathcal{S}$ instead of ‘within’. We need to use the so-called ‘nested’ notation.

Specifically, recall that each subject appears only in one level of $\mathcal{A}$, and so we will say that $\mathcal{S}$ is *nested* in $\mathcal{A}$. In general, a factor is said to be nested in a second factor when each level of the first factor appears only in a single level of the second. The relationship ‘ $\mathcal{S}$ is nested in $\mathcal{A}$ ’ is denoted $\mathcal{S}(\mathcal{A})$.¹ Due to the fact that subjects in one-factor between-subject design are assigned to (nested in) different levels of the independent variable $\mathcal{A}$, we call this type of design an $\mathcal{S}(\mathcal{A})$ design (read ‘ $\mathcal{S}$ in $\mathcal{A}$ design’).

The within-group variability, symbolized $\mathcal{S}(\mathcal{A})$, corresponds to the difference between subjects assigned to a particular condition. Although $\mathcal{S}(\mathcal{A})$ is used to reflect the experimental

¹ A word of caution: the notation is somewhat counterintuitive. A common error is to confuse the roles of the factors in the notation $\mathcal{S}(\mathcal{A})$. The factor $\mathcal{S}$ is the *nested* factor even though ($\mathcal{A}$) is enclosed or ‘nested’ within parentheses.

error, note that it is impossible to isolate the experimental error pertaining to the interaction between subjects and the independent variable.

In summary:

$$\begin{aligned} SS_{\text{between}} &\text{ becomes } SS_A, & df_{\text{between}} &\text{ becomes } df_A. \\ SS_{\text{within}} &\text{ becomes } SS_{S(A)}, & df_{\text{within}} &\text{ becomes } df_{S(A)}. \end{aligned} \quad (8.1)$$

$$\begin{aligned} MS_{\text{between}} &\text{ becomes } MS_A. \\ MS_{\text{within}} &\text{ becomes } MS_{S(A)}. \end{aligned} \quad (8.2)$$

With this new notation, the ANOVA table becomes:

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
<i>A</i>	1	90.00	90.00	15.00
<i>S(A)</i>	8	48.00	6.00	
Total	9	138.00		

8.5 Presentation of the ANOVA results

While it may be a good method to present the details of the statistical tests in six points as we have been doing, it would be tiresome to present the test in that way in a research report, where, among other things, space is generally at a premium, just as in psychology journal articles and books. In effect, these six points constitute above all a healthy regimen in our work, and ought to be specified before carrying out an experiment. That way we can avoid finding ourselves at the end of a study that we cannot use, because we do not know what to do with the data we have collected, or because its methodological errors prevent any sensible interpretation of the results.

For the presentation of results, a long tradition demands that we present essentially the ANOVA table, indicating whether the value of *F* permits us to reject the null hypothesis at the given α level. We indicate suspension of judgment by the term *ns* (non-significant), and often indicate rejection of H_0 at the .05 level by * and rejection at the .01 level by **. (We need to specify the meanings of * and ** in a legend or note appended to the table). We usually indicate the smallest α level possible for each *F*. That way it is understood that if we reject H_0 at the .05 level then we could not reject it at the .01 level. For example, here is how we would present the ANOVA table for 'imagery and memory', following these conventions:

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
<i>A</i>	1	90.00	90.00	15.00**
<i>S(A)</i>	8	48.00	6.00	
Total	9	138.00		

** : $p < .01$

In such a table, some authors prefer to indicate the actual name of the variable rather than the impersonal designation $\mathcal{A}$. In general, these authors prefer to call the source of variation $S(\mathcal{A})$ —within the groups—the *error* term. They would present the following table:

Source	df	SS	MS	F
Imagery	1	90.00	90.00	15.00**
Error	8	48.00	6.00	
Total	9	138.00		

** : $p < .01$

Another presentation detached from the two traditional α levels .05 and .01 is gaining popularity. Instead of rejecting H_0 at a given α level, the table simply indicates the probability associated with F . Thus, readers can choose their own level and see whether the probability associated with F permits the rejection of H_0 . Such a table would look like this:

Source	df	SS	MS	F	$p(F)$
$\mathcal{A}$	1	90.00	90.00	15.00**	.00048
$S(\mathcal{A})$	8	48.00	6.00		
Total	9	138.00			

8.5.1 Writing the results in an article

In a scientific publication, space is at a premium, and so publishing the complete ANOVA table would be considered superfluous. In general, the authors of a paper will give only the indispensable information. This amounts to giving the value of F in the text, indicating the degrees of freedom in parentheses, along with the value of the mean square of error followed by the probability associated with the value of F , or by the indication of the α level chosen. Thus, for our example we would write:

As our theory predicted, the imagery group performed better than the control group, $F(1, 8) = 15.00$, $MS_e = 6.00$, $p < .01$ (cf. American Psychological Association manual, 2001, p. 22ff.).

8.6 ANOVA with two groups: F and t

The analysis of variance compares the means of different groups and evaluates whether the difference between these means reflects real differences as opposed to random fluctuations. In the particular case of comparing two means, we can use another popular technique known as Student's t . So when dealing with two groups, two different techniques could be used. This is, in general, judged by most users of statistics as being an unpleasant situation, because it would imply the possibility of having to chose between two ways of proceeding. If this idea creates some dissonance, you can rest assured, because the choice here is purely illusory. These two techniques are strictly equivalent. It is due only to some quirk of history that they are

presented as different techniques. Precisely the value of F can be derived directly from the value of t by the relation

$$F = t^2 ,$$

and the sampling distribution of t^2 with ν degrees of freedom (and hence of t) is exactly the sampling distribution of F with $\nu_1 = 1$ and $\nu_2 = \nu$ degrees of freedom. It is relatively easy (but somewhat tedious) to show that F and t (precisely t^2) are equivalent. The following digression shows the details of the proof. If you feel comfortable with accepting it, feel free to skip it.

8.6.1 For experts: Fisher and Student ... Proof

In this section, we show that F ratio and Student's t^2 are equal. The main idea is to start with the formula of Student's t , then to square and develop it till we end up with Fisher's F . In what follows, we assume that the two groups being compared have the same size (this makes the proof easier to follow, but this is not essential).

The first thing to remark is that when we square the critical values of the Student distribution for the parameter ν , we find exactly the same values as for a Fisher distribution with $\nu_1 = 1$ and $\nu_2 = \nu$. So the sampling distributions are equivalent. What we now need to show is that the criteria themselves are equivalent.

We start with Student formula:

$$t = \frac{M_{1.} - M_{2.}}{\widehat{\sigma}_{M_{1.} - M_{2.}}} ,$$

with

$$\widehat{\sigma}_{M_{1.} - M_{2.}} = \sqrt{\widehat{\sigma}_{M_{1.} - M_{2.}}^2}$$

being the standard deviation of the means sampling distribution (see Section A.5.4). It is computed as

$$\begin{aligned} \widehat{\sigma}_{M_{1.} - M_{2.}}^2 &= \frac{\sum (Y_{1s} - M_{1.})^2}{S(S-1)} + \frac{\sum (Y_{2s} - M_{2.})^2}{S(S-1)} \\ &= \frac{SS_1}{S(S-1)} + \frac{SS_2}{S(S-1)} \\ &= \frac{SS_{S(A)}}{S(S-1)} . \end{aligned} \tag{8.3}$$

We return to Student's t formula and square it,

$$t^2 = \frac{(M_{1.} - M_{2.})^2}{SS_{S(A)}/S(S-1)} = \frac{S(S-1)(M_{1.} - M_{2.})^2}{SS_{S(A)}} .$$

Let us now go back to F , and re-write:

$$F = \frac{MS_A}{MS_{S(A)}} = \frac{SS_A/(A-1)}{SS_{S(A)}/A(S-1)} .$$

When dealing with two groups, $A = 2$, and F becomes

$$F = \frac{SS_A}{SS_{S(A)}/2(S-1)} = \frac{2(S-1)SS_A}{SS_{S(A)}} .$$

But,

$$SS_A = S \sum (M_a - M_{..})^2 .$$

And therefore we obtain the following expression:

$$\begin{aligned}
 \sum (M_{a.} - M_{..})^2 &= (M_{1.} - M_{..})^2 + (M_{2.} - M_{..})^2 \\
 &= M_{1.}^2 + M_{..}^2 - 2M_{..}M_{1.} + M_{2.}^2 + M_{..}^2 - 2M_{..}M_{2.} \\
 &= M_{1.}^2 + M_{2.}^2 + 2M_{..}^2 - 2M_{..}M_{1.} - 2M_{..}M_{2.} \\
 &= M_{1.}^2 + M_{2.}^2 - M_{..}(2M_{1.} + 2M_{2.} - 2M_{..}) .
 \end{aligned} \tag{8.4}$$

Taking into account that

$$\frac{(M_{1.} + M_{2.})}{2} = M_{..} ,$$

we obtain

$$\begin{aligned}
 \sum (M_{a.} - M_{..})^2 &= M_{1.}^2 + M_{2.}^2 - M_{..}(M_{1.} + M_{2.}) \\
 &= M_{1.}^2 + M_{2.}^2 - (M_{1.} + M_{2.})(M_{1.} + M_{2.})/2 \\
 &= \frac{(2M_{1.}^2 + 2M_{2.}^2 - M_{1.}^2 - M_{2.}^2 - 2M_{1.}M_{2.})}{2} \\
 &= \frac{(M_{1.} - M_{2.})^2}{2} .
 \end{aligned} \tag{8.5}$$

Therefore,

$$SS_A = \frac{S(M_{1.} - M_{2.})^2}{2} .$$

Plugging in the formula for F , this last equation gives

$$F = \frac{2(S-1)SS_A}{SS_{S(A)}} = \frac{S(S-1)(M_{1.} - M_{2.})^2}{SS_{S(A)}} = t^2 .$$

8.6.2 Another digression: F is an average

In this additional digression, we indicate that F can be seen as the average of all the differences between pairs of experimental means. In fact, the relation that we have used in the previous digression, namely

$$SS_A = \frac{S(M_{1.} - M_{2.})^2}{2}$$

is a particular case of a more general one which is

$$SS_A = \frac{S \sum_{a < a'} (M_{a.} - M_{a'.})^2}{A} .$$

For example, when $A = 3$

$$SS_A = \frac{S(M_{1.} - M_{2.})^2 + S(M_{1.} - M_{3.})^2 + S(M_{2.} - M_{3.})^2}{3} .$$

The proof is left as an exercise.²

² In case you have insomnia!

8.7 Another example: Romeo and Juliet

So far we have seen how to deal with the case of two groups. The following example shows that this procedure can easily be extended to several groups. In an experiment on the effect of context on memory, Bransford and Johnson (1972) read the following passage to their subjects:

If the balloons popped, the sound would not be able to carry since everything would be too far away from the correct floor. A closed window would also prevent the sound from carrying since most buildings tend to be well insulated. Since the whole operation depends on a steady flow of electricity, a break in the middle of the wire would also cause problems. Of course the fellow could shout, but the human voice is not loud enough to carry that far. An additional problem is that a string could break on the instrument. Then there could be no accompaniment to the message. It is clear that the best situation would involve less distance. Then there would be fewer potential problems. With face to face contact, the least number of things could go wrong.

To show the importance of the context on the memorization of texts, the authors used four experimental conditions which correspond to four experimental groups:

1. 'No context' condition: subjects listened to the passage and tried to remember it.
2. 'Appropriate context before' condition: subjects were provided with an appropriate context in the form of a picture (see Figure 8.3b) and then listened to the passage.
3. 'Appropriate context after' condition: subjects first listened to the passage and then were provided with an appropriate context in the form of a picture (see Figure 8.3b).
4. 'Partial context' condition: subjects were provided with a context (i.e. the picture shown in Figure 8.3a) that did not allow them to make sense of the text when they listened to the passage.

Strictly speaking this experiment involves one experimental group (group 2: 'appropriate context before'), and three control groups (groups 1, 3, and 4). The function of the control groups is to eliminate rival theoretical hypotheses (i.e. rival theories that would give the same experimental predictions as the theory advocated by the authors). For example, the authors want to make sure that the superior performance of group 2 over group 1 (which is the difference of importance) cannot be explained by the effect of imagery. Hence they introduce the control given by group 4. The imagery theory would predict that group 4 should perform as well as group 2 whereas the authors' theory would predict that group 4 should perform at the level of group 1. To make sure that the facilitating effect of the picture comes from the integration of long-term and short-term memory material, the authors added the experimental condition of group 3. If the integration effect does not require an 'online' process, then group 2 and group 3 should have the same level of performance. If, on the contrary, integration does require some 'online' process (which is the authors' contention), then group 2 should do better than group 3. What the authors were doing when they were planning their experiment was to make sure that the experimental condition would *dissociate* the different theories and hence would allow them to decide which one (if any) of the current theories would be in agreement with the data they would generate.

Even though we know that the authors will measure memory, we do not know yet how they are going to *operationalize* it. For this study, the authors decided to use two dependent variables: a score of comprehension given by the subjects on a rating scale from 0 to 7 (with 0

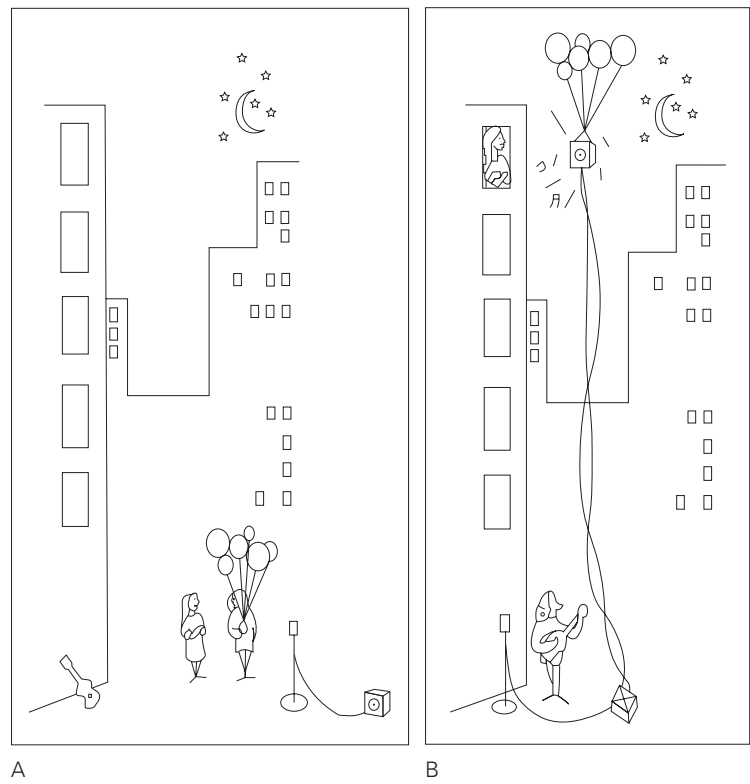


Figure 8.3 (A) Partial context, and (B) appropriate context for Bransford and Johnson’s (1972) experiment.

meaning no comprehension and 7 meaning perfect comprehension), and the number of ideas of the text correctly recalled. In order to do so, Bransford *et al.* considered that the text is made up of 14 ideas (try to find them as an exercise) and counted how many of them are recalled by the subjects. Even though this dependent variable can lead to some delicate coding problems (i.e. of reliability) because it is not always easy to find a clear way of making sure that an idea is present or not, it clearly reflects the theoretical problem at stake in this experiment.

For the (fictitious) replication of this experiment, we have chosen to include 20 subjects assigned randomly to 4 groups. Hence there are $S = 5$ subjects *per* group. The dependent variable we will focus on is the ‘number of ideas’ recalled (of a maximum of 14). The results are presented in Table 8.2. Before inspecting it, try to do what any good experimenter should do, namely specifying the first 5 steps of the statistical decision test.

Here are the detailed steps to compute the F ratio. Try to do the computation on your own first, then check that you obtain the right results.

First, we need to find the different degrees of freedom:

$$df_A = (A - 1) = 4 - 1 = 3 \tag{8.6}$$

$$df_{S(A)} = A(S - 1) = 4(5 - 1) = 16$$

$$df_{\text{total}} = AS - 1 = 4 \times 5 - 1 = 19 .$$

	No context	Context before	Context after	Partial context
	3	5	2	5
	3	9	4	4
	2	8	5	3
	4	4	4	5
	3	9	1	4
$Y_{a\cdot}$	15	35	16	21
$M_{a\cdot}$	3	7	3.2	4.2
$M_{..} = 4.35$				

Table 8.2 Scores for 20 subjects in the fictitious replication of the 'Romeo and Juliet' experiment.

We then need to compute the different sums of squares:

$$\begin{aligned}
 SS_{\text{between}} &= SS_A = S \sum_a (M_{a\cdot} - M_{..})^2 \\
 &= 5[(3 - 4.35)^2 + (7 - 4.35)^2 + (3.2 - 4.35)^2 + (4.2 - 4.35)^2] \\
 &= 5[(-1.35)^2 + (2.65)^2 + (-1.15)^2 + (-.15)^2] \\
 &= 5[1.82 + 7.02 + 1.32 + .02] \\
 &= 5 \times 10.18 \\
 &= 50.90
 \end{aligned}$$

$$\begin{aligned}
 SS_{\text{within}} &= SS_{S(A)} = \sum_a \sum_s (Y_{as} - M_{a\cdot})^2 \\
 &= (3 - 3)^2 + (3 - 3)^2 + (2 - 3)^2 + (4 - 3)^2 + (3 - 3)^2 \\
 &\quad + (5 - 7)^2 + (9 - 7)^2 + (8 - 7)^2 + (4 - 7)^2 + (9 - 7)^2 \\
 &\quad + (2 - 3.2)^2 + (4 - 3.2)^2 + (5 - 3.2)^2 + (4 - 3.2)^2 + (1 - 3.2)^2 \\
 &\quad + (5 - 4.2)^2 + (4 - 4.2)^2 + (3 - 4.2)^2 + (5 - 4.2)^2 + (4 - 4.2)^2 \\
 &= (0)^2 + (0)^2 + (-1)^2 + (1)^2 + (0)^2 \\
 &\quad + (-2)^2 + (-2)^2 + (1)^2 + (-3)^2 + (2)^2 \\
 &\quad + (-1.2)^2 + (.8)^2 + (1.8)^2 + (.8)^2 + (-2.2)^2 \\
 &\quad + (.8)^2 + (-.2)^2 + (-1.2)^2 + (.8)^2 + (-.2)^2 \\
 &= 0 + 0 + 1 + 1 + 0 \\
 &\quad + 4 + 4 + 1 + 9 + 4 \\
 &\quad + 1.44 + .64 + 3.24 + .64 + 4.84 \\
 &\quad + .64 + .04 + 1.44 + .64 + .04 \\
 &= 37.60
 \end{aligned}$$

$$\begin{aligned}
SS_{\text{total}} &= \sum_a \sum_s (Y_{as} - M_{..})^2 \\
&= (3 - 4.35)^2 + (3 - 4.35)^2 + (2 - 4.35)^2 + (4 - 4.35)^2 + (3 - 4.35)^2 \\
&\quad + (5 - 4.35)^2 + (9 - 4.35)^2 + (8 - 4.35)^2 + (4 - 4.35)^2 + (9 - 4.35)^2 \\
&\quad + (2 - 4.35)^2 + (4 - 4.35)^2 + (5 - 4.35)^2 + (4 - 4.35)^2 + (1 - 4.35)^2 \\
&\quad + (5 - 4.35)^2 + (4 - 4.35)^2 + (3 - 4.35)^2 + (5 - 4.35)^2 + (4 - 4.35)^2 \\
&= (-1.35)^2 + (-1.35)^2 + (-2.35)^2 + (-.35)^2 + (-1.35)^2 \\
&\quad + (.65)^2 + (4.65)^2 + (3.65)^2 + (-.35)^2 + (4.65)^2 \\
&\quad + (-2.35)^2 + (-.35)^2 + (.65)^2 + (-.35)^2 + (-3.35)^2 \\
&\quad + (.65)^2 + (-.35)^2 + (-1.35)^2 + (.65)^2 + (-.35)^2 \\
&= (1.82) + (1.82) + (5.52) + (.120) + (1.82) \\
&\quad + (.42) + (21.62) + (13.32) + (.12) + (21.62) \\
&\quad + (5.52) + (.12) + (.42) + (.12) + (11.22) \\
&\quad + (.42) + (.12) + (1.82) + (.42) + (.12) \\
&= 88.50
\end{aligned}$$

Next, we find the mean squares.

$$MS_A = \frac{SS_A}{df_A} = \frac{50.90}{3} = 16.97 \quad (8.7)$$

$$MS_{S(A)} = \frac{SS_{S(A)}}{df_{S(A)}} = \frac{37.60}{16} = 2.35$$

And, finally, we can compute our F value.

$$F = \frac{MS_A}{MS_{S(A)}} = \frac{16.97}{2.35} = 7.22 .$$

We can now fill in the following table:

Source	df	SS	MS	F
$\mathcal{A}$	$A - 1$	SS_A	$\frac{SS_A}{(A - 1)}$	$\frac{MS_A}{MS_{S(A)}}$
$S(\mathcal{A})$	$A(S - 1)$	$SS_{S(A)}$	$\frac{SS_{S(A)}}{A(S - 1)}$	
Total	$AS - 1$	SS_{total}		

which gives:

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p</i> (<i>F</i>)
<i>A</i>	3	50.90	16.97	7.22**	.00288
<i>S</i> (<i>A</i>)	16	37.60	2.35		
Total	19	88.50			

If we decide to use the critical values procedure, we will find that for the α level of .01, $F_{\text{critical}} = 5.29$. Because F is larger than the critical value we decide to reject the null hypothesis and to accept the alternative hypothesis stating that there is an effect of the experimental manipulations on the number of ideas recalled by the subjects.

Hence, the type of context or of activity during encoding or reading the text clearly influences the memory of subjects, $F(3, 16) = 7.22$, $MS_e = 2.35$, $p < .01$. Figure 8.4 shows the details of the experimental effect: it seems that the authors' theory is supported by the data (more about that in Chapter 12 about contrast analysis).

It is worth noting that the analysis of variance leads to the conclusion that there is an effect of the independent variable on the dependent variable but that it cannot give a detailed or specific account of this effect. The ANOVA reveals only a *global* effect of the independent variable. In order to make that clear, the F ratio is often called an *omnibus* test (from the latin *omnibus* meaning 'for all' or 'for everything'). In particular, the results of the ANOVA as they are stated cannot be used to decide if group 2 differs from all the other groups significantly (which is the crucial feature of the authors' prediction). This specific type of question will be handled later on when we will deal with the problem of contrast analysis.

8.8 How to estimate the effect size

8.8.1 Motivation

Rejecting the null hypothesis allows us to draw the conclusion that the independent variable has an effect on the dependent variable but it *does not give an estimation of the intensity of the effect*. In fact, the only thing that we can say for sure is that the effect is not equal to zero! A common mistake is to confuse the size of the experimental effect with the probability level. The propensity to make this mistake was illustrated in a (beautiful) study by Rosenthal and Gaito (1963). They asked their subjects (who were not only students but also active researchers) a question, which we would express with our notations, as:

Consider two experiments using an $S(A)$ design, the first with a number of subjects $S = 5$ and the second with a number of subjects $S = 50$. Suppose that for both experiments the null hypothesis is rejected for $\alpha = .05$ but not for $\alpha = .01$. Which one of these two experiments indicates the stronger (i.e. clearer) effect of the independent variable: the experiment with $S = 5$ subjects or the experiment with $S = 50$ subjects?

If you gave the (wrong) answer, $S = 50$, don't feel too bad—you were not alone in doing so! Quite a large number of subjects in this experiment did it too and were confusing the intensity of an effect with the number of subjects as well as confusing the α level with the intensity

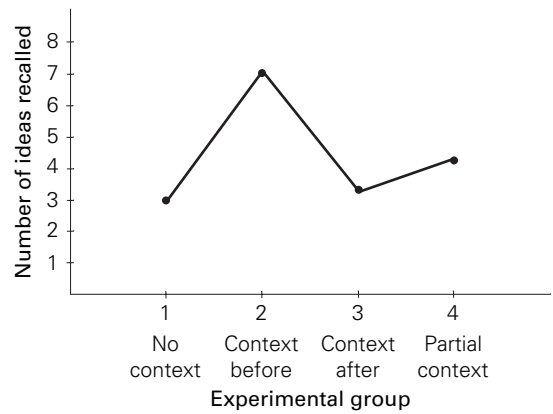


Figure 8.4 Results of a (fictitious) replication of Bransford and Johnson’s (1972) experiment. The Y axis represents the dependent variable which is the number of ideas recalled.

of an effect. The best thing to do with a mistake is to avoid making it twice (or too often). Using coefficients designed to estimate the intensity of the effect helps to avoid making these mistakes, especially because they make them easier to spot.

So, in general, an ANOVA is followed by some estimation of the size of the experimental effect. Unfortunately there are several ways to define this estimation. The common idea in the different approaches is to express the amount of variance of the dependent variable that is explained by the independent variable as a proportion of the total variance of the dependent variable. Another way of saying this is that the coefficients of effect size try to express the effect of the independent variable as a reduction of the *uncertainty* (and hence variance) of the dependent variable when the knowledge of the independent variable is taken into account (sounds like correlation, no?).

The coefficients of effect can be described as belonging to two main families.

- The first family corresponds to descriptive statistics. It defines the effect of the independent variable as the ratio of the sum of squares of effect (SS_A) to the total sum of squares. The coefficients known as $R^2_{Y \cdot A}$ and η^2 belong to this family. Actually these two coefficients correspond to a case of double identity (i.e. they are the same coefficient under two different names), but were defined in different contexts and are known under these two names in the literature.
- The second family corresponds to inferential statistics. Its goal is to estimate a parameter of the population (i.e. the intensity of effect in the population). Specifically, in this case, the goal is to estimate the ratio of the variance of A to the error variance. The statistics known as the ‘corrected’ or ‘shrunk’ $R^2_{Y \cdot A}$ (see Chapter 3, page 56) belong to this category. Two cousins of these coefficients are known as ω and ρ . These coefficients will be introduced after we have defined the ANOVA score model on which these coefficients rely (see Section 10.5, page 201).

8.8.2 **R and η**

The first coefficient of effect exists under two (or even three) different identities: η^2 and $R^2_{Y \cdot A}$. Historically, η comes from the analysis of variance tradition. The coefficient $R^2_{Y \cdot A}$ comes from the regression/correlation tradition. The current trend is to favor the label $R^2_{Y \cdot A}$ in any

context, but the old notation can still be found frequently. Here we will follow the current trend and use the denomination $R^2_{Y \cdot A}$ from now on.

This coefficient expresses the intensity of the effect of the independent variable as the ratio of the sum of squares of A to the total sum of squares:

$$R^2_{Y \cdot A} = \frac{SS_A}{SS_A + SS_{S(A)}} = \frac{SS_A}{SS_{\text{total}}} \quad (8.8)$$

Because the total sum of squares is always positive and larger than or equal to the sum of squares of A (SS_A is a proportion of SS_{total}), the value of $R^2_{Y \cdot A}$ will always stay between the values 0 and 1. When multiplied by 100, $R^2_{Y \cdot A}$ gives the part of variance of the dependent variable explained by the independent variable.

For example, with the experiment ‘Romeo and Juliet’ used previously, we had:

$$SS_A = 50.90, \quad SS_{S(A)} = 37.60, \quad SS_{\text{total}} = 88.50$$

and

$$R^2_{Y \cdot A} = \frac{50.90}{88.50} = .57514,$$

which means that 57.514% of the variance of the dependent variable is explained by the independent variable.

8.8.2.1 Digression: $R^2_{Y \cdot A}$ is a coefficient of correlation

Equation 8.8 looks suspiciously similar to the equation of the squared coefficient of correlation used in regression analysis. Remember that the equation we used was

$$r^2_{Y \cdot X} = \frac{SS_{\text{regression}}}{SS_Y} = \frac{SS_{\text{regression}}}{SS_{\text{total}}}.$$

It suffices to equate the name of the independent variable with the regression effect for the equations to be equivalent.

In the following paragraph, we will prove that the $R^2_{Y \cdot A}$ and $r^2_{Y \cdot X}$ are in fact the *same* coefficient under two different disguises. (You can skip this section if you feel comfortable accepting this assertion at face value.) Actually, the effect of the independent variable is expressed through the experimental means. This amounts to trying to predict the values Y_{as} of the subjects’ scores by the mean of their group; or which is more general, to try to predict Y_{as} from the deviation of the group mean from the grand mean ($M_a - M_{..}$). Let us denote this squared coefficient of correlation by $r^2_{Y \cdot X}$. The general formula for a squared correlation coefficient between the variables X and Y is:

$$r^2_{Y \cdot X} = \frac{[\sum (X_s - M_X)(Y_s - M_Y)]^2}{\sum (X_s - M_X)^2 \sum (Y_s - M_Y)^2}$$

with M_Y and M_X representing the means of Y and X respectively. Adapting this formula to the present context and simplifying, we find that:

$$\begin{aligned} R^2_{Y \cdot A} &= \frac{\left[\sum (Y_{as} - M_{..})(M_a - M_{..}) \right]^2}{\sum (Y_{as} - M_{..})^2 \sum (M_a - M_{..})^2} \\ &= \frac{\left[\sum (Y_{as} M_a - M_{..} Y_{as} - M_{..} M_a + M_{..}^2) \right]^2}{SS_{\text{total}} SS_A} \end{aligned}$$

$$\begin{aligned}
&= \frac{\left(\sum Y_{as} M_{a.} - M_{..} \sum Y_{as} - M_{..} \sum M_{a.} + ASM_{..}^2 \right)^2}{SS_{\text{total}} SS_A} \\
&= \frac{\left(\sum_a M_{a.} \sum_s Y_{as} - ASM_{..}^2 - ASM_{..}^2 + ASM_{..}^2 \right)^2}{SS_{\text{total}} SS_A} \\
&= \frac{\left(S \sum M_{a.}^2 - ASM_{..}^2 \right)^2}{SS_{\text{total}} SS_A} \\
&= \frac{\left[S \sum (M_{a.} - M_{..})^2 \right]^2}{SS_{\text{total}} SS_A} \\
&= \frac{SS_A^2}{SS_{\text{total}} SS_A} \\
&= \frac{SS_A}{SS_{\text{total}}} \\
&= R_{Y \cdot A}^2,
\end{aligned}$$

which proves that $R_{Y \cdot A}^2$ is actually the square of a coefficient of correlation.

This makes clear that $R_{Y \cdot A}^2$ is the proportion of the variance of the dependent variable explained by the independent variable $\mathcal{A}$. The value $1 - R^2$ gives the proportion of the variance of the dependent variable left unexplained by the independent variable. It is also sometimes called the *coefficient of alienation*.

8.8.2.2 F and $R_{Y \cdot A}^2$

It helps if we rewrite the formula for F in order to make its relations with $R_{Y \cdot A}^2$ clearer.

$$\begin{aligned}
F &= \frac{MS_A}{MS_{S(A)}} \\
&= \frac{SS_A / (A - 1)}{SS_{S(A)} / A(S - 1)} \\
&= \frac{SS_A}{SS_{S(A)}} \times \frac{A(S - 1)}{(A - 1)} \\
&= \frac{SS_A / SS_{\text{total}}}{SS_{S(A)} / SS_{\text{total}}} \times \frac{A(S - 1)}{A - 1} \\
&= \frac{R_{Y \cdot A}^2}{(1 - R_{Y \cdot A}^2)} \times \frac{A(S - 1)}{A - 1}
\end{aligned}$$

So F and $R_{Y \cdot A}^2$ are related to each other by the formula:

$$F = \frac{R_{Y \cdot A}^2}{(1 - R_{Y \cdot A}^2)} \times \frac{A(S - 1)}{A - 1} = \frac{R_{Y \cdot A}^2}{(1 - R_{Y \cdot A}^2)} \times \frac{df_{S(A)}}{df_A}. \quad (8.9)$$

This formula shows that F is made up of two terms:

- *The first term:*

$$\frac{R_{Y \cdot A}^2}{(1 - R_{Y \cdot A}^2)}$$

depends only upon the intensity of the experimental effect (assuming as a first approximation that $R_{Y \cdot A}^2$ corresponds to the intensity of the effect in the population). This term is often referred to as a 'signal-to-noise' ratio.

- *The second term:*

$$\frac{A(S - 1)}{A - 1}$$

depends only upon the number of subjects (the number of experimental conditions A is supposed to be fixed, because it is dictated by the theoretical reasons leading to the design of the experiment). Another way to refer to this term is to call it a coefficient of amplification: the more subjects the more amplification for the effect of the independent variable.

The fact that F is made of these two components makes it clear that for any given level of intensity of effect, as long as *it is not strictly zero*, it will always be possible to find a number S which will be large enough to allow for the rejection of the null hypothesis. In other words, the smaller the effect is, the larger the number of subjects needed to detect an experimental effect.

8.8.3 How many subjects? Quick and dirty power analysis

Another potential use of Equation 8.9 is to give a first rough approximation of the number of subjects needed to conduct an experiment. This is a very common problem, because the number of subjects available is in general limited, and to run subjects in an experiment can be costly.

In some cases, the experimenter is able to decide *a priori* of the intensity of the effect that is of interest. In order to find the number of subjects needed we then solve the previous equation after plugging in the appropriate values of F_{critical} . For example, suppose that we plan a new experiment with $A = 5$ experimental conditions, and we want to detect some effect only if it is $\geq .25$ (i.e. we want to be able to explain at least 25% of the variability of the dependent variable with our model), and that we will be using the α level of .01.

We start by making a first trial with an arbitrary number of subjects $S = 20$, and then solve the equation for F .

- The first step is to find the critical value for $\nu_1 = A - 1 = 4$ and for $\nu_2 = A(S - 1) = 5 \times 19 = 95$. The table gives a value of $F_{\text{critical}} = 3.51$.
- The second step is to plug in the values of $R_{Y \cdot A}^2$ and of S in the equation and to see if the F obtained leads to the rejection of H_0 . Here we find:

$$F = \frac{R_{Y \cdot A}^2}{(1 - R_{Y \cdot A}^2)} \times \frac{A(S - 1)}{A - 1} = \frac{.25}{.75} \times \frac{95}{4} = 7.92 > 3.51,$$

which shows that with $S = 20$ we can expect to be able to detect an experimental effect explaining at least 25% of the dependent variable variance.

If we try to use $S = 10$ subjects:

- The first step is to find the critical value for $\nu_1 = A - 1 = 4$ and for $\nu_2 = A(S - 1) = 5 \times 9 = 45$. $F_{\text{critical}} = 3.77$.
- The second step is to plug in the values of $R^2_{Y \cdot A}$ and of S in the equation and to see if the F obtained leads to the rejection of null hypothesis. Here we find,

$$\frac{.25}{.75} \times \frac{45}{4} = 3.75 < 3.77 ,$$

which shows that with $S = 10$ we cannot expect to be able to detect an experimental effect explaining at least 25% of the dependent variable variance.

If we use $S = 11$, on the other hand, we find that we can reject the null hypothesis:

$$\frac{.25}{.75} \times \frac{50}{4} = 4.17 > 3.72 = F_{\text{critical}} \quad \text{for } \nu_1 = 4 \quad \text{and } \nu_2 = 50 .$$

From this we can see that $S = 11$ is the minimum number of subjects that we can use in order to find an experimental effect explaining at least 25% percent of the variance of the dependent variable.

The procedure described here should be considered as being only a quick and dirty procedure because of the fact that $R^2_{Y \cdot A}$ is actually a poor estimate of the intensity of effect in the population. It can be shown, in particular, that $R^2_{Y \cdot A}$ always overestimates the parameter in the population. As a consequence this quick and dirty procedure *underestimates* the number of subjects needed. One way of improving it is to use a better approximation of the parameter of the population. Several other more complex procedures exist, but the present one gives the gist of most of them.

8.8.4 How much explained variance? More quick and dirty power analysis

Another use for Equation 8.9 is to find, from a given number of available subjects, what is the smallest possible effect that can be detected. This is obtained by following the same procedure as in Section 8.8.3, but now we want to find how large a value of $R^2_{Y \cdot A}$ can be when the number of groups and the number of subjects is given.

The first step of this procedure is to compute a quantity which we will call W_{critical} :

$$W_{\text{critical}} = \frac{A - 1}{A(S - 1)} F_{\text{critical}} ,$$

with F_{critical} corresponding to the α level chosen and with $\nu_1 = (A - 1)$ and $\nu_2 = A(S - 1)$ degrees of freedom.

Then the minimum $R^2_{Y \cdot A}$ that we will be able to detect with S subjects at the desired α level will be

$$\min(R^2_{Y \cdot A}) = \frac{W_{\text{critical}}}{1 + W_{\text{critical}}} .$$

For example, suppose that you have planned an experiment with $A = 4$ levels. You know that you have only 20 subjects available. What is the minimum $R^2_{Y \cdot A}$ that you will be able to detect? With a total of 20 subjects, you have $S = 20/4 = 5$ subjects per group. Suppose that you are happy with reporting your results with an α level of .05. The first step is to locate in the F_{critical} Table the value corresponding to $\nu_1 = (A - 1) = 3$ and $\nu_2 = A(S - 1) = 16$ degrees of freedom.

We find that for $\alpha = .05$, $F_{\text{critical}} = 3.24$. If we plug these values into the previous equations, we first find:

$$W_{\text{critical}} = \frac{A-1}{A(S-1)} F_{\text{critical}} = \frac{3 \times 3.24}{16} \approx .6075.$$

With this value of W_{critical} , we find that the smallest $R^2_{Y \cdot A}$ that we can detect with $S = 5$ subjects per group is

$$\min(R^2_{Y \cdot A}) = \frac{W_{\text{critical}}}{1 + W_{\text{critical}}} = \frac{.6075}{1 + .6075} \approx .3779.$$

Hence, using an ANOVA with 20 subjects and 4 experimental groups, we will be able to detect an experimental effect only if it explains at least 38% of the variance of the dependent variable.

8.9 Computational formulas

The formulas used so far to compute the sums of squares are the direct translation of their definition. They have, as a consequence, the advantage of being clear and easy to understand. These formulas are named *comprehension* or *defining* formulas. They have, however, the disadvantage of leading to cumbersome computational procedures. When the problem is to *compute* the sums of squares manually, especially when the number of data is large, some other formulas are to be preferred: they are called *computational* formulas.

For example, some elementary algebraic manipulations give alternative computationally easier formulas for the sums of squares. They are obtained by expanding and simplifying the *comprehension* formulas.

$$\begin{aligned} SS_A &= S \sum (M_{a\cdot} - M_{\cdot\cdot})^2 = \left(\frac{\sum Y_{a\cdot}^2}{S} \right) - \left(\frac{Y_{\cdot\cdot}^2}{AS} \right) \\ SS_{S(A)} &= \sum (Y_{as} - M_{a\cdot})^2 = \left(\sum Y_{as}^2 \right) - \left(\frac{\sum Y_{a\cdot}^2}{S} \right) \\ SS_{\text{total}} &= \sum (Y_{as} - M_{\cdot\cdot})^2 = \left(\sum Y_{as}^2 \right) - \left(\frac{Y_{\cdot\cdot}^2}{AS} \right) \end{aligned}$$

In what follows, we present a 7 step computational routine to compute the different sums of squares. We use this opportunity to introduce a new notation named the ‘numbers in the square’. Their construction is explained later on, but this is how they look:

$$\boxed{A}, \quad \boxed{AS}, \quad \boxed{1}.$$

You can skip this section for the first reading and come back to it if you want to learn more about the ‘numbers in the square’, or if you feel the need to compute by hand an analysis of variance.

8.9.1 Back to Romeo and Juliet

The results of the fictitious replication of Bransford and Johnson’s experiment on the effect of context on memory are recalled below.

	No context	Context before	Context after	Partial context
	3	5	2	5
	3	9	4	4
	2	8	5	3
	4	4	4	5
	3	9	1	4
$Y_{a\cdot}$	15	35	16	21
$M_{a\cdot}$	3	7	3.2	4.2
$\sum_s Y_{as}^2$	47	267	62	91

We already saw how to compute the sums of squares corresponding to these data using the comprehension formulas. We will now see how to compute them using 7 values labeled Q1 to Q7 (for quantity) and three numbers in a square ($\boxed{\mathbf{A}}$, $\boxed{\mathbf{AS}}$, and $\boxed{\mathbf{1}}$). Recall that in the following computational procedure

- $A = 4$ is the number of levels of the independent variable $\mathcal{A}$.
- $S = 5$ is the number of subjects *per* group (or number of levels of the factor $\mathcal{S}$).

$$Q1 = Y_{\cdot\cdot} = 3 + 3 + \dots + 5 + 4 = 87.00$$

$$Q2 = \boxed{\mathbf{AS}} = \sum Y_{as}^2 = 3^2 + 3^2 + \dots + 5^2 + 4^2 = 467.00$$

$$Q3 = \boxed{\mathbf{A}} = \sum \frac{Y_{a\cdot}^2}{S} = \frac{15^2}{5} + \dots + \frac{21^2}{5} = 429.40$$

$$Q4 = \boxed{\mathbf{1}} = \frac{Y_{\cdot\cdot}^2}{AS} = \frac{Q1^2}{AS} = \frac{87^2}{20} = 378.45$$

$$Q5 = SS_{\text{total}} = \boxed{\mathbf{AS}} - \boxed{\mathbf{1}} = 467.00 - 378.45 = 88.55$$

$$Q6 = SS_{\text{between}} = SS_A = \boxed{\mathbf{A}} - \boxed{\mathbf{1}} = 429.40 - 378.45 = 50.95$$

$$Q7 = SS_{\text{within}} = SS_{S(A)} = \boxed{\mathbf{AS}} - \boxed{\mathbf{A}} = 467.00 - 429.40 = 37.60$$

Note that due to rounding differences the results differ slightly from those we found with the comprehension formulas.

8.9.2 The 'numbers in the squares'

8.9.2.1 Principles of construction

We have just introduced into the calculation procedure three new quantities called 'numbers in the squares' (several authors call them 'basic ratios', 'abbreviated symbols', 'sums of squares elements', 'squared quantities', or 'computational symbols' —Keppel, 1982; Kirk, 1982; Lee, 1975; Myers, 1979; and Winer, 1971, respectively). Recall the computational formula

(remember that the $\sum$ sign indicates that the summation operates over all the *active indices* of an expression):

$$\boxed{\text{AS}} = \sum Y_{as}^2$$

$$\boxed{\text{A}} = \sum \frac{Y_{a.}^2}{S}$$

$$\boxed{1} = \frac{Y_{..}^2}{AS}$$

In looking at these three formulas, we can deduce the following rules for constructing the ‘*numbers in the squares*’. For example, we could take the number in square $\boxed{\text{A}}$:

1. Write Y^2 .
2. Put as active indices (lower-case letters) of Y the letters in the square. Since the letter A is ‘in the square’, a is the active index. Replace the missing letters with periods. Here the letter S is missing from the square, so the index s is replaced by a period. If we have the number 1 in the square, *it is not an active index*, and, thus, all the indices are replaced by periods.

For our example, we obtain $Y_{a.}^2$.

3. If there are active indices: place the sign $\sum$ in front of Y . (Here a is active, but s , replaced by a period, is inactive.)

This gives us $\sum Y_{a.}^2$.

4. If there are inactive indices, divide the sums by the upper-case letters corresponding to the missing indices. Here s is an inactive index, so divide by S .

This gives finally:

$$\sum \frac{Y_{a.}^2}{S} = \boxed{\text{A}}.$$

8.9.2.2 ‘Numbers in the squares’ and the universe ...

In this section, we look at the relationships among the ‘numbers in the squares’, the labels for the sources of variability, the degrees of freedom, and the comprehension and calculation formulas. Examine the following table:

Source variability	df	Expanded df	Comprehension formula	Computational formula
$\mathcal{A}$	$A - 1$	$A - 1$	$S \sum (M_{a.} - M_{..})^2$	$\boxed{\text{A}} - \boxed{1}$
$S(\mathcal{A})$	$A(S - 1)$	$AS - A$	$\sum (Y_{as} - M_{a.})^2$	$\boxed{\text{AS}} - \boxed{\text{A}}$

Notice that:

- The label for the source of variability gives the number of *df* if we replace the letters *outside the parentheses* with their value minus 1. For example, $\mathcal{A}$ gives $A - 1$ *df*. And we replace the letters *within* the parentheses with their value: $S(\mathcal{A})$ gives $A(S - 1)$ *df*.
- The developed number of *df* (after we do all the multiplications to eliminate the parentheses) gives the indices of the means (or the scores Y if all the indices are present) for the comprehension formulas.

- The developed number of *df* also gives the calculation formulas. It suffices to replace each term in the developed *df* with the corresponding ‘*number in the square*’.

This set of relations among the labels of sources of variability, the number of degrees of freedom, the comprehension formulas and the calculation formulas works for all the designs studied in this book. Here, these relations can appear only as an amusing diversion. Later, for the study of complex factorial designs, these relations constitute the basis of statistical analysis.

Chapter summary

8.10 Key notions of the chapter

Below are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

Null and alternative hypothesis	ANOVA table
Decision rule	Effect size
Fisher’s sampling distributions	Power analysis
Critical value	Comprehension and computational formulas
Intensity of effect and coefficient of correlation	Numbers in the squares

8.11 New notations

Below are the new notations introduced in this chapter. Test yourself on their meaning.

<i>t</i>	A	AS	1
SS_A, MS_A, df_A	$W_{critical}$		
$SS_{S(A)}, MS_{S(A)}, df_{S(A)}$	$R^2_{Y \cdot A} \eta$		

8.12 Key formulas of the chapter

Below are the main formulas introduced in this chapter: try to go through them and understand what they mean.

$$SS_A = S \sum (M_a - M_{..})^2 = \mathbf{A} - \mathbf{1}$$
$$SS_{S(A)} = \sum (Y_{as} - M_a)^2 = \mathbf{AS} - \mathbf{A}$$
$$\mathbf{AS} = \sum Y_{as}^2$$

$$\boxed{\mathbf{A}} = \frac{\sum Y_{a.}^2}{S}$$

$$\boxed{\mathbf{1}} = \frac{Y_{..}^2}{AS}$$

$$R_{Y \cdot A}^2 = \frac{SS_A}{SS_A + SS_{S(A)}} = \frac{SS_A}{SS_{\text{total}}}.$$

$$W_{\text{critical}} = \frac{A-1}{A(S-1)} F_{\text{critical}}$$

8.13 Key questions of the chapter

Below are some questions about the content of this chapter. All the answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- * Sources of variation A and $S(A)$: what are the relationships with the labels *between-group* and *within-group*?
- * What is the relationship between Student's t -test and Fisher's F ?
- * What are the advantages of the new notation using names of independent variables such as A and $S(A)$?
- * What are the differences between comprehension formulas and computational formulas?
- * How can we derive the 'numbers in the square' from the computational formula?
- * How can we detect a small effect
- * When can we use a small number of subjects?
- * How can we quickly estimate the number of subjects needed in an experiment?
- * How can we quickly estimate the size of a detectable effect when the number of subjects is known?

9

ANOVA, one factor: regression point of view

9.1 Introduction

In Chapters 7 and 8 we have seen how to analyze a one-factor experimental design using analysis of variance. In this chapter, we show that this approach can be interpreted in the general framework of regression and correlation. We have already seen, in Chapter 8 (Section 8.8.2.1, pp. 163ff.), that the coefficient $R^2_{Y.A}$ which measures the intensity of the effect of the independent variable on the dependent variable is a coefficient of correlation. In this chapter, we will explore further the relationships between regression analysis and analysis of variance.

In order to use regression for analyzing the data from an analysis of variance design we use a trick that has a lot of interesting consequences. The main idea is to find a way of replacing the nominal independent variable (i.e. the experimental factor) by a numerical independent variable (remember that the independent variable should be numerical to run a regression). One can look at analysis of variance as a technique predicting subjects' behavior from the experimental group. The trick, then, is to find a way of coding those groups. Several choices are possible; an easy one is to *represent a given experimental group by its mean for the dependent variable*.¹ In this framework, the general idea is to try to predict the subjects' scores from the mean of the group to which they belong. The rationale is that, if there is an experimental effect, then the mean of a subject's group should predict the subject's score better than the grand mean. In other words, the larger the experimental effect the better the predictive quality of the group mean. Using the group mean to predict the subjects' performance has an interesting consequence that makes regression and analysis of variance identical: *when we*

¹ The rationale behind regression analysis (presented in Chapter 4 about regression) implies that the independent variable is under the *control* of the experimenter. Using the group mean seems to go against this requirement, because we need to wait until *after* the experiment to know the values of the independent variable. This is why we call our procedure a trick. It works because it is equivalent to more elaborate coding schemes using multiple regression analysis. It has the advantage of being simpler conceptually and computationally.

Control	Experimental
Subject 1: 1	Subject 1: 8
Subject 2: 2	Subject 2: 8
Subject 3: 5	Subject 3: 9
Subject 4: 6	Subject 4: 11
Subject 5: 6	Subject 5: 14
$M_{1.} = M_{\text{control}} = 4$	$M_{2.} = M_{\text{experimental}} = 10$
Grand mean = $M_Y = M_{..} = 7$	

Table 9.1 Results of the ‘memory and imagery’ experiment.

predict the performance of subjects from the mean of their group, the predicted value turns out to be the group mean too!

We will illustrate the regression approach using two examples. In the first, we show how a simple experiment involving two experimental groups can be similarly analyzed using either regression analysis or analysis of variance. In the second example, we extend the comparison between regression and analysis of variance to an experiment involving more than two experimental groups.

9.2 Example 1: memory and imagery

As a first illustration of the relationship between ANOVA and regression we use the ‘memory and imagery’ experiment detailed in Chapter 7. Remember that in this experiment two groups of subjects were asked to learn pairs of words (e.g. ‘beauty–carrot’). Subjects in the first group (control group) were simply asked to learn the pairs of words as best they could. Subjects in the second group (experimental group) were asked to picture each word in a pair and to make an image of the interaction between the two objects. After some delay, subjects in both groups were asked to give the second word (e.g. ‘carrot’) as the answer when prompted with the first word in the pair (e.g. ‘beauty’). For each subject, the number of words correctly recalled was recorded. The purpose of this experiment was to demonstrate an effect of the independent variable (i.e. learning with imagery vs learning without imagery) on the dependent variable (i.e. number of words correctly recalled). The results of the experiment are presented in Table 9.1 and Figure 9.1.

9.3 Analysis of variance for Example 1

As a refresher, we first compute the analysis of variance as seen in Chapter 8. Using the data from Table 9.1, we can compute the different sums of squares of the analysis of variance.

The total sum of squares is:

$$\begin{aligned}
 SS_{\text{total}} &= \sum_a \sum_s (Y_{as} - M_{..})^2 \\
 &= (1 - 7)^2 + (2 - 7)^2 + (5 - 7)^2 + (6 - 7)^2 + (6 - 7)^2 \\
 &\quad + (8 - 7)^2 + (8 - 7)^2 + (9 - 7)^2 + (11 - 7)^2 + (14 - 7)^2
 \end{aligned}$$

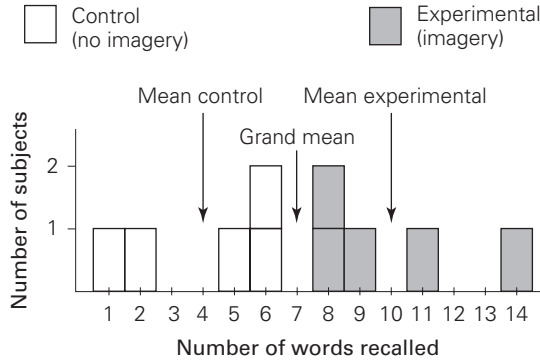


Figure 9.1 Histogram of the results of the experiment on: 'imagery and memory.'

$$\begin{aligned}
 &= (-6)^2 + (-5)^2 + (-2)^2 + (-1)^2 + (-1)^2 \\
 &\quad + 1^2 + 1^2 + 2^2 + 4^2 + 7^2 \\
 &= 36 + 25 + 4 + 1 + 1 + 1 + 1 + 4 + 16 + 49 \\
 &= 138.00 .
 \end{aligned} \tag{9.1}$$

The between-group (or experimental) sum of squares is:

$$\begin{aligned}
 SS_{\text{between}} &= S \sum_a (M_{a\cdot} - M_{\cdot\cdot})^2 = \\
 &= 5 \times [(4 - 7)^2 + (10 - 7)^2] \\
 &= 5 \times [(-3)^2 + 3^2] \\
 &= 5 \times [9 + 9] \\
 &= 90.00 .
 \end{aligned} \tag{9.2}$$

The within-group (or residual or error) sum of squares is:

$$\begin{aligned}
 SS_{\text{within}} &= \sum_a \sum_s (Y_{as} - M_{a\cdot})^2 \\
 &= (1 - 4)^2 + (2 - 4)^2 + (5 - 4)^2 + (6 - 4)^2 + (6 - 4)^2 \\
 &\quad + (8 - 10)^2 + (8 - 10)^2 + (9 - 10)^2 + (11 - 10)^2 + (14 - 10)^2 \\
 &= (-3)^2 + (-2)^2 + 1^2 + 2^2 + 2^2 \\
 &\quad + (-2)^2 + (-2)^2 + (-1)^2 + 1^2 + 4^2 \\
 &= 9 + 4 + 1 + 4 + 4 + 4 + 4 + 1 + 1 + 16 \\
 &= 48.00 .
 \end{aligned} \tag{9.3}$$

As a check we verify that

$$\begin{aligned}
 SS_{\text{total}} &= SS_{\text{within}} + SS_{\text{between}} \\
 &= 48.00 + 90.00 \\
 &= 138.00 .
 \end{aligned} \tag{9.4}$$

We need also to find the number of degrees of freedom of the different sums of squares. For the total sum of squares, we find the following value for the total number of degrees of freedom.

$$\begin{aligned} df_{\text{total}} &= (A \times S) - 1 \\ &= (2 \times 5) - 1 = 10 - 1 \\ &= 9 . \end{aligned} \tag{9.5}$$

The between-group number of degrees of freedom is:

$$\begin{aligned} df_{\text{between}} &= A - 1 \\ &= 2 - 1 \\ &= 1 . \end{aligned} \tag{9.6}$$

The within-group number of degrees of freedom is:

$$\begin{aligned} df_{\text{within}} &= A(S - 1) \\ &= 2(5 - 1) \\ &= 2 \times 4 \\ &= 8 . \end{aligned} \tag{9.7}$$

As a check, we verify that the degrees of freedom add up:

$$\begin{aligned} df_{\text{total}} &= df_{\text{within}} + df_{\text{between}} \\ &= 8 + 1 \\ &= 9 . \end{aligned} \tag{9.8}$$

From the sums of squares and their number of degrees of freedom we can compute the mean squares. The between-group mean square is:

$$\begin{aligned} MS_{\text{between}} &= \frac{SS_{\text{between}}}{df_{\text{between}}} \\ &= \frac{90.00}{1} \\ &= 90.00 . \end{aligned} \tag{9.9}$$

The within-group mean square is:

$$\begin{aligned} MS_{\text{within}} &= \frac{SS_{\text{within}}}{df_{\text{within}}} \\ &= \frac{48.00}{8} \\ &= 6.00 . \end{aligned} \tag{9.10}$$

We can now proceed to the computation of the index F (we will use the formula with the mean squares first):

$$\begin{aligned} F &= \frac{MS_{\text{between}}}{MS_{\text{within}}} \\ &= \frac{90.00}{6.00} \\ &= 15.00 . \end{aligned} \tag{9.11}$$

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Between	1	90.00	90.00	15.00**
Within	8	48.00	6.00	
Total	9	138.00		

Table 9.2 The ANOVA table for the results of the 'Imagery and Memory' experiment. ** $p < .01$

Under the null hypothesis this F ratio will follow a Fisher distribution with $\nu_1 = 1$ and $\nu_2 = 8$ degrees of freedom. Using the critical value procedure, we find that for $\alpha = .01$ the critical value is 11.26. Because the computed value of F is larger than the critical value, we can reject the null hypothesis and, therefore, we conclude that imagery does facilitate memory.

We can also compute F using the coefficient of intensity of effect, $R_{Y \cdot A}^2$. Its value is:

$$R_{Y \cdot A}^2 = \frac{SS_{\text{between}}}{SS_{\text{total}}} = \frac{90.00}{138.00} \approx .6522. \quad (9.12)$$

From the value of $R_{Y \cdot A}^2$, we compute the following F ratio:

$$F = \frac{R_{Y \cdot A}^2}{1 - R_{Y \cdot A}^2} \times \frac{df_{\text{within}}}{df_{\text{between}}} = \frac{.6522}{1 - .6522} \times 8 \approx 15.00. \quad (9.13)$$

which, indeed, agrees with the value computed previously!

The results of the analysis are summarized in the traditional analysis of variance fashion in Table 9.2.

9.4 Regression approach for Example 1: mental imagery

Recall that the general idea behind the regression approach is to find a way to transform or to code the nominal independent variable as a numerical independent variable. Using symbols, we can write that we want to re-code the nominal independent variable A as a quantitative independent variable X for use in a regression analysis. As we have mentioned before, we have chosen a trick to transform A into X : *The value of X for all subjects in a group will be the group mean.*

Why does that work? First, because we want to predict the subjects' performance in the experimental condition, we need to treat all observations in a given condition alike. Second, we want the values of the quantitative variable to describe the performance of the groups as closely as possible,² and the best way³ of describing a group's performance is its mean.

In summary, the general idea when using the regression approach for an analysis of variance problem is to predict subject scores from the mean of the group to which they belong. The rationale for doing so is to consider the group mean as representing the experimental effect,

² We could also have said that we want the values of the quantitative variable to give the best possible prediction. These two formulations are in fact equivalent, as can be seen as a consequence of the digression in Section 9.8.

³ In a least squares sense, that is.

and hence as a predictor of the subjects' behavior. If the independent variable has an effect, the group mean should be a better predictor of the subject's behavior than the grand mean. Formally, we want to predict the score Y_{as} of subject s in condition a from a quantitative variable X that will be equal to the mean of the group a in which the s observation was collected. With an equation, we want to predict each observation as

$$\hat{Y} = a_{\text{regression}} + bX, \quad (9.14)$$

with X being equal to M_a . Please note that in this chapter we use the notation $a_{\text{regression}}$ for the intercept of the regression line instead of a in order to avoid the confusion with a meaning the 'experimental group of the subject'. The particular choice of X has several interesting consequences, but the most important one is that the mean of the predictor M_X is also the mean of the dependent variable M_Y . These two means are also equal to the grand mean of the analysis of variance. With an equation,

$$M_{..} = M_X = M_Y. \quad (9.15)$$

For our example, the mean of the control group, denoted M_{control} , is equal to 4, and the mean of the experimental group, denoted $M_{\text{experimental}}$, is equal to 10. Using the group mean to predict subjects' performance leads to the regression problem given in Table 9.3.

To analyze this regression problem we use the steps detailed in Chapter 4. The first step is to compute the predicted values $\hat{Y}$, by finding the values of the intercept and the slope in the equation $\hat{Y} = a_{\text{regression}} + bX$. The quantities needed to compute the values of $a_{\text{regression}}$ and b are given in Table 9.4. The *slope* is found to be

$$b = \frac{SCP_{XY}}{SS_X} = \frac{90}{90} = 1. \quad (9.16)$$

The intercept is

$$a_{\text{regression}} = M_Y - bM_X = 7 - 1 \times 7 = 0. \quad (9.17)$$

The equation to predict Y from X (keeping in mind that X is the subject group mean M_a .) is therefore

$$\hat{Y} = a_{\text{regression}} + bX = 0 + 1 \times X = M_a. \quad (9.18)$$

This is a remarkable result! We have just found that the *best prediction of a subject's score is the group mean of this score*. As a consequence, when using the regression approach for an analysis of variance problem *the predicted value $\hat{Y}$ and the predictor X have the same value*. This is not a fluke, and we will show later in a digression that this is always the case. Having $\hat{Y}$ equal to M_a is what makes analysis of variance and regression analysis the same technique.

X = M_a . predictor	4	4	4	4	4	10	10	10	10	10
Y (value to be predicted)	1	2	5	6	6	8	8	9	11	14

Table 9.3 The data from Table 9.1 presented as a regression problem. The predictor X is the value of the mean of the subject's group.

	$X = M_{a\cdot}$	x	x^2	Y_{as}	y	y^2	$x \times y$
	4	−3	9	1	−6	36	18
	4	−3	9	2	−5	25	15
	4	−3	9	5	−2	4	6
	4	−3	9	6	−1	1	3
	4	−3	9	6	−1	1	3
	10	3	9	8	1	1	3
	10	3	9	8	1	1	3
	10	3	9	9	2	4	6
	10	3	9	11	4	16	12
	10	3	9	14	7	49	21
Σ	70	0	90	70	0	138	90
	SS_X			SS_Y		SCP_{YX}	

Table 9.4 Quantities needed to compute the values of $a_{\text{regression}}$ and b for the data presented as a regression problem in Table 9.3. The predictor X is the value of the mean of the subject’s group. The mean of X is equal to $M_X = \frac{70}{10} = 7$. The mean of Y is equal to $M_Y = \frac{70}{10} = 7$. (Both are equal to the grand mean of the ANOVA, $M_{\cdot\cdot} = 7$.) The following abbreviations are used: $x = (X - M_X)$, $y = (Y - M_Y)$.

$X = M_{a\cdot}$ predictor	4	4	4	4	4	10	10	10	10	10
Y (value to be predicted)	1	2	5	6	6	8	8	9	11	14

Table 9.5 The data from Table 9.1 presented as a regression problem. The predicted value ($\hat{Y}$) is the value of the mean of the subject’s group ($M_{a\cdot}$).

The predicted values are collected in Table 9.5. From these values we can proceed to compute the sums of squares total, regression and residual, and then the squared coefficient of correlation between the predicted values and the actual values.

The sum of squares total is obtained as the sum of the squared distances from each score to the grand mean. The grand mean being equal to $M_Y = 7$, we obtain:

$$\begin{aligned} SS_{\text{total}} &= \sum_a \sum_s (Y - M_Y)^2 \\ &= (1 - 7)^2 + (2 - 7)^2 + (5 - 7)^2 + (6 - 7)^2 + (6 - 7)^2 \\ &\quad + (8 - 7)^2 + (8 - 7)^2 + (9 - 7)^2 + (11 - 7)^2 + (14 - 7)^2 \\ &= (-6)^2 + (-5)^2 + (-2)^2 + (-1)^2 + (-1)^2 \\ &\quad + 1^2 + 1^2 + 2^2 + 4^2 + 7^2 \\ &= 36 + 25 + 4 + 1 + 1 + 1 + 1 + 4 + 16 + 49 \\ &= 138.00 . \end{aligned} \tag{9.19}$$

The sum of squares regression is obtained by computing the sum of the squared distance from each predicted score (here the mean of the subject’s group) to the grand mean M_Y .

$$SS_{\text{regression}} = \sum (\hat{Y} - M_Y)^2$$

$$\begin{aligned}
&= \underbrace{(4-7)^2 + (4-7)^2 + (4-7)^2 + (4-7)^2 + (4-7)^2}_{5 \text{ times}} \\
&\quad + \underbrace{(10-7)^2 + (10-7)^2 + (10-7)^2 + (10-7)^2 + (10-7)^2}_{5 \text{ times}} \\
&= 5 \times (4-7)^2 + 5 \times (10-7)^2 \\
&= 5 \times 9 + 5 \times 9 \\
&= 90.00 .
\end{aligned} \tag{9.20}$$

The sum of squares residual is obtained by computing the sum of the squared distance between each subject score and each predicted score (the group means).

$$\begin{aligned}
SS_{\text{residual}} &= \sum (Y - \hat{Y})^2 \quad (\text{recall that } \hat{Y} \text{ is the group mean } M_a. \text{ here}) \\
&= (1-4)^2 + (2-4)^2 + (5-4)^2 + (6-4)^2 + (6-4)^2 \\
&\quad + (8-10)^2 + (8-10)^2 + (9-10)^2 + (11-10)^2 + (14-10)^2 \\
&= (-3)^2 + (-2)^2 + 1^2 + 2^2 + 2^2 \\
&\quad + (-2)^2 + (-2)^2 + (-1)^2 + 1^2 + 4^2 \\
&= 9 + 4 + 1 + 4 + 4 + 4 + 4 + 1 + 1 + 16 \\
&= 48.00 .
\end{aligned} \tag{9.21}$$

We can check, incidently, that the sum of squares total is equal to the sum of squares regression and the sum of squares residual.

$$\begin{aligned}
SS_{\text{total}} &= SS_{\text{regression}} + SS_{\text{residual}} \\
&= 90.00 + 48.00 \\
&= 138.00 .
\end{aligned} \tag{9.22}$$

The second step in analyzing our data as a regression problem is to compute the squared coefficient of correlation between the predicted scores and the actual subjects' scores (or, which is equivalent, the correlation between the dependent variable and the independent variable). This is done simply by dividing the sum of squares regression by the sum of squares total (cf. Chapter 4):

$$r_{Y \cdot Y}^2 = r_{Y \cdot X}^2 = R_{Y \cdot X}^2 = \frac{SS_{\text{regression}}}{SS_{\text{total}}} = \frac{90.00}{138.00} = .6522 . \tag{9.23}$$

We are now ready to compute the index F .

$$\begin{aligned}
F &= \frac{r_{Y \cdot X}^2}{1 - r_{Y \cdot X}^2} \times \frac{df_{\text{residual}}}{df_{\text{regression}}} \\
&= \frac{r_{Y \cdot X}^2}{1 - r_{Y \cdot X}^2} \times \frac{N - 2}{1} \\
&= \frac{.6522}{1 - .6522} \times 8 = 1.8752 \times 8 \approx 15.00 .
\end{aligned} \tag{9.24}$$

Regression notation	ANOVA notation
Y	Y_{as}
$\hat{Y}$	$M_{a\cdot}$
X	$M_{a\cdot}$
M_Y	$M_{..}$
M_X	$M_{..}$
$SS_{\text{regression}}$	SS_{between} or SS_A
SS_{residual} or SS_{error}	SS_{within} or $SS_{S(A)}$
$R^2_{Y\cdot X}$ or $r^2_{Y\cdot X}$ or $r^2_{\hat{Y}\cdot Y}$	$R^2_{Y\cdot A}$
N	$A \times S$

Table 9.6 Equivalence of notations between regression and ANOVA. The main difference between these two techniques is in the notations!

9.5 Equivalence between regression and analysis of variance

By now you may have realized that the main difference between regression and analysis of variance boils down to different notations and emphasis on prediction or hypothesis testing. To make clear the similarities between these approaches, we have gathered the formulas for the sums of squares, degrees of freedom, mean squares, and F ratio for the imagery example. Table 9.6 gives the notational equivalences. Study them carefully and make sure you master the ‘double notation’.

Total sum of squares:

$$\begin{aligned} SS_{\text{total}} &= \sum (Y - M_Y)^2 = \sum_a \sum_s (Y_{as} - M_{..})^2 \\ &= (1 - 7)^2 + (2 - 7)^2 + (5 - 7)^2 + (6 - 7)^2 + (6 - 7)^2 \\ &\quad + (8 - 7)^2 + (8 - 7)^2 + (9 - 7)^2 + (11 - 7)^2 + (14 - 7)^2 \\ &= (-6)^2 + (-5)^2 + (-2)^2 + (-1)^2 + (-1)^2 \\ &\quad + 1^2 + 1^2 + 2^2 + 4^2 + 7^2 \\ &= 36 + 25 + 4 + 1 + 1 + 1 + 1 + 4 + 16 + 49 \\ &= 138.00 . \end{aligned} \tag{9.25}$$

Regression *alias* between sum of squares:

$$\begin{aligned} SS_{\text{regression}} &= \sum (\hat{Y} - M_Y)^2 = S \sum_a (M_{a\cdot} - M_{..})^2 = SS_{\text{between}} \\ &= 5 \times [(4 - 7)^2 + (10 - 7)^2] \\ &= 5 \times [9 + 9] \\ &= 90.00 . \end{aligned} \tag{9.26}$$

Residual *alias* within sum of squares:

$$\begin{aligned}
 SS_{\text{residual}} &= \sum (Y - \hat{Y})^2 = \sum_a \sum_s (Y_{as} - M_{a\cdot})^2 = SS_{\text{within}} \\
 &= (1 - 4)^2 + (2 - 4)^2 + (5 - 4)^2 + (6 - 4)^2 + (6 - 4)^2 \\
 &\quad + (8 - 10)^2 + (8 - 10)^2 + (9 - 10)^2 + (11 - 10)^2 + (14 - 10)^2 \\
 &= (-3)^2 + (-2)^2 + 1^2 + 2^2 + 2^2 + (-2)^2 + (-2)^2 + (-1)^2 + 1^2 + 4^2 \\
 &= 9 + 4 + 1 + 4 + 4 + 4 + 4 + 1 + 1 + 16 \\
 &= 48.00 .
 \end{aligned} \tag{9.27}$$

The sums of squares add up:

$$\begin{aligned}
 SS_{\text{total}} &= SS_{\text{residual}} + SS_{\text{regression}} = SS_{\text{within}} + SS_{\text{between}} \\
 &= 48.00 + 90.00 \\
 &= 138.00 .
 \end{aligned} \tag{9.28}$$

The total number of degrees of freedom is

$$\begin{aligned}
 df_{\text{total}} &= N - 1 = (A \times S) - 1 \\
 &= 10 - 1 = (2 \times 5) - 1 \\
 &= 9 .
 \end{aligned} \tag{9.29}$$

Regression *alias* between-group number of degrees of freedom:

$$\begin{aligned}
 df_{\text{regression}} &= df_{\text{between}} = A - 1 \\
 &= 2 - 1 \\
 &= 1 .
 \end{aligned} \tag{9.30}$$

Residual *alias* within-group number of degrees of freedom:

$$\begin{aligned}
 df_{\text{residual}} &= df_{\text{within}} = A(S - 1) \\
 &= 2(5 - 1) \\
 &= 2 \times 4 \\
 &= 8 .
 \end{aligned} \tag{9.31}$$

The degrees of freedom add up:

$$\begin{aligned}
 df_{\text{total}} &= df_{\text{residual}} + df_{\text{regression}} = df_{\text{within}} + df_{\text{between}} \\
 &= 8 + 1 \\
 &= 9 .
 \end{aligned} \tag{9.32}$$

Regression *alias* between-group mean square:

$$\begin{aligned}
 MS_{\text{regression}} &= \hat{\sigma}_{\text{regression}}^2 = \frac{SS_{\text{regression}}}{df_{\text{regression}}} = MS_{\text{between}} = \frac{SS_{\text{between}}}{df_{\text{between}}} \\
 &= \frac{90.00}{1} \\
 &= 90.00 .
 \end{aligned} \tag{9.33}$$

Residual *alias* within-group mean square:

$$\begin{aligned}
 MS_{\text{residual}} &= \hat{\sigma}_{\text{residual}}^2 = \frac{SS_{\text{residual}}}{df_{\text{residual}}} = MS_{\text{within}} = \frac{SS_{\text{within}}}{df_{\text{within}}} \\
 &= \frac{48.00}{8} \\
 &= 6.00 .
 \end{aligned} \tag{9.34}$$

The F ratio is:

$$\begin{aligned}
 F &= \frac{MS_{\text{regression}}}{MS_{\text{residual}}} = \frac{\hat{\sigma}_{\text{regression}}^2}{\hat{\sigma}_{\text{residual}}^2} = \frac{MS_{\text{between}}}{MS_{\text{within}}} \\
 &= \frac{90.00}{6.00} \\
 &= 15.00.
 \end{aligned} \tag{9.35}$$

The F ratio can also be computed from the coefficient of correlation as:

$$F = \frac{r_{Y \cdot X}^2}{1 - r_{Y \cdot X}^2} \times \frac{df_{\text{residual}}}{df_{\text{regression}}} = \frac{R_{Y \cdot A}^2}{1 - R_{Y \cdot A}^2} \times \frac{df_{\text{within}}}{df_{\text{between}}} . \tag{9.36}$$

9.6 Example 2: Romeo and Juliet

This second example is ‘Romeo and Juliet’, the replication of Bransford *et al.*’s (1972) experiment. The rationale and details were given in Chapter 8 (Section 8.7, page 157). We want to illustrate with this example that the correlation/regression approach to the analysis of variance works with more than two groups. The only thing to remember is to correct the sums of squares by dividing them by their degrees of freedom in order to obtain the correct mean squares. The degrees of freedom can also be used directly with the formula for F using the squared coefficients of correlation $R_{Y \cdot A}^2$ or $r_{Y \cdot X}^2$.

As in the two-group case, we consider that predicting an effect of the experimental manipulation is equivalent to predicting the subjects’ score from the mean of their experimental group. We then compute the regression of Y on X ; because X is a coding of the independent variable, this amounts to computing the regression of Y on $\mathcal{A}$ (remember that $\mathcal{A}$ is the name of the independent variable). The necessary steps are detailed in Table 9.7. As previously, we find that when predicting Y from $X = M_a$, the slope is equal to one, and the intercept is equal to zero. Therefore, the predicted value $\hat{Y}$ is equal to M_a .

From Table 9.7, we can now compute the squared coefficient of correlation as:

$$R_{Y \cdot X}^2 = \frac{SS_{\text{regression}}}{SS_{\text{total}}} = \frac{50.95}{88.55} = .5754 .$$

Group	Y	X $M_{q\cdot}$	Reg $(M_{q\cdot} - M_{q\cdot\cdot})$	Reg ² $(M_{q\cdot} - M_{q\cdot\cdot})$	Res $(Y_{as} - M_{q\cdot})$	Res ² $(Y_{as} - M_{q\cdot\cdot})$	Total $(Y_{as} - M_{q\cdot\cdot})$	Total ² $(Y_{as} - M_{q\cdot\cdot})$
1	3	3	-1.35	1.8225	0	0	-1.35	1.8225
1	3	3	-1.35	1.8225	0	0	-1.35	1.8225
1	2	3	-1.35	1.8225	-1	1	-2.35	5.5225
1	4	3	-1.35	1.8225	1	1	-0.35	0.1225
1	3	3	-1.35	1.8225	0	0	-1.35	1.8225
2	5	7	2.65	7.0225	-2	4	0.65	0.4225
2	9	7	2.65	7.0225	2	4	4.65	21.6225
2	8	7	2.65	7.0225	1	1	3.65	13.3235
2	4	7	2.65	7.0225	-3	9	-0.35	0.1225
2	9	7	2.65	7.0225	2	4	4.65	21.6225
3	2	3.2	-1.15	1.3225	-1.2	1.44	-2.35	5.5225
3	4	3.2	-1.15	1.3225	0.8	0.64	-0.35	0.1225
3	5	3.2	-1.15	1.3225	1.6	3.24	0.65	0.4225
3	4	3.2	-1.15	1.3225	0.8	0.64	-0.35	0.1225
3	1	3.2	-1.15	1.3225	-2.2	4.84	-3.35	11.2225
4	5	4.2	-0.15	0.2250	0.8	0.64	0.65	0.4225
4	4	4.2	-0.15	0.2250	-0.2	0.04	-0.35	0.1225
4	3	4.2	-0.15	0.2250	-1.2	1.44	-1.35	1.8225
4	5	4.2	-0.15	0.2250	0.8	0.64	0.65	0.4225
4	4	4.2	-0.15	0.2250	-0.2	0.04	-0.35	0.1225
Σ			0.00	50.95	0.0	37.60	0.00	88.55
			$SS_{\text{regression}}$			SS_{residual}		
			SS_{between}			SS_{within}		
			SS_A			$SS_{S(A)}$		
			SS_{total}			SS_{total}		
			SS_{total}			SS_{total}		

Reg, regression; Res, residual.

Table 9.7 Values needed to compute the regression and residual sums of squares corresponding to ‘Romeo and Juliet’ (Bransford *et al.*’s experiment).

The first formula for the F ratio is

$$\begin{aligned}
 F &= \frac{R_{Y \cdot X}^2}{1 - R_{Y \cdot X}^2} \times \frac{df_{\text{residual}}}{df_{\text{regression}}} \\
 &= \frac{.5754}{1 - .5754} \times \frac{4(5 - 1)}{4 - 1} \\
 &= \frac{.5754}{.4246} \times \frac{16}{3} \\
 &= 7.2270 .
 \end{aligned} \tag{9.37}$$

The second formula for F uses the mean square ‘between’ (*alias regression*) and the mean square ‘within’ (*alias residual*). Each of these mean squares is obtained by dividing the corresponding sum of squares by its number of degrees of freedom.

We obtain the following values for the mean squares:

$$\begin{aligned}
 MS_{\text{regression}} &= MS_{\text{between}} = MS_A \\
 &= \frac{SS_{\text{regression}}}{df_{\text{regression}}} = \frac{SS_{\text{between}}}{df_{\text{between}}} = \frac{SS_A}{df_A} = \frac{SS_A}{A - 1} \\
 &= \frac{50.95}{4 - 1} \\
 &= 16.98
 \end{aligned} \tag{9.38}$$

and

$$\begin{aligned}
 MS_{\text{residual}} &= MS_{\text{within}} = MS_{S(A)} \\
 &= \frac{SS_{\text{residual}}}{df_{\text{residual}}} = \frac{SS_{\text{within}}}{df_{\text{within}}} = \frac{SS_{S(A)}}{df_{S(A)}} = \frac{SS_{S(A)}}{A(S - 1)} \\
 &= \frac{37.60}{4(5 - 1)} \\
 &= 2.35 .
 \end{aligned} \tag{9.39}$$

With these values, we can now compute the F ratio as

$$\begin{aligned}
 F &= \frac{MS_{\text{regression}}}{MS_{\text{residual}}} = \frac{MS_{\text{between}}}{MS_{\text{within}}} = \frac{MS_A}{MS_{S(A)}} \\
 &= \frac{16.88}{2.35} \\
 &= 7.227 .
 \end{aligned} \tag{9.40}$$

which is, as it should be, the same value found previously.

In order to decide if there is an experimental effect from the value of the test, we need to evaluate the probability associated with the computed value of the F ratio. Supposing we use a program to find this value, we shall find that $\Pr(F > 7.2270) = .0028$. This probability is smaller than the α level of .01, hence we can reject the null hypothesis and assume that there is indeed an effect of the experimental manipulation.⁴ In APA style we would say that the type of context during learning influences the amount of material recalled later on, $F(3, 16) = 7.23$, $MS_e = 2.35$, $p < .01$.

⁴ We would reach the same conclusion using the critical value approach.

9.7 If regression and analysis of variance are one thing, why keep two different techniques?

We have seen in this chapter that regression and analysis of variance are, essentially, the same technique with two different points of view. The obvious question is: if regression and analysis of variance are one thing, why keep two different techniques? The first reason is historical, the second more theoretical.

The historical reason is due to the fact that these two approaches were developed in different branches of psychology (analysis of variance in experimental psychology, regression in social and clinical psychology) and that, depending upon their domain, scientists use one or the other terminology. If we want to be able to read journal articles, we have to learn both terminologies.

The theoretical reason is that some problems are more easily solved within one framework. For example, unbalanced designs are more easily analyzed within the regression framework, but designs with several factors are easier to analyze within the analysis of variance framework.

9.8 Digression: when predicting Y from $M_{a.}$, $b = 1$, and $a_{\text{regression}} = 0$

In this section, we examine, in a more general manner, the consequences of using the mean of a group to predict the performance of a subject from that group. Specifically, we show that when the predictor is the group mean, then the predicted value is also *always* the group mean. In other words, we show that the best possible prediction (in a least squares sense) of subjects' performance is their group mean. Therefore, if we want to transform the nominal independent variable $\mathcal{A}$ into a quantitative independent variable X , the group mean is the best possible choice.

The idea of predicting subjects' performance from their mean is sometimes referred to as the *mean model* by statisticians. This approach and this terminology become important when dealing with complex or unbalanced designs. Formally, we want to show that when we predict Y_{as} from $X = M_{a.}$, the predicted value $\hat{Y}_{as}$ is equal to $M_{a.}$. This is equivalent to saying that in the equation,

$$\hat{Y} = a_{\text{regression}} + bX,$$

the slope (b) is equal to one, and the intercept ($a_{\text{regression}}$) is equal to zero.

9.8.1 Remember ...

The main idea behind this proof is essentially to rewrite and substitute in the regression equation the quantities coming from the analysis of variance approach. Let us start by rewriting the regression equation taking into account that the X (the predictor) is the group mean:

$$\hat{Y} = a_{\text{regression}} + bX \quad \text{becomes} \quad \hat{Y}_{as} = a_{\text{regression}} + bM_{a.} \quad (9.41)$$

We need to have present in mind the following facts:

- X is the group mean:

$$X = M_{a.} \quad (9.42)$$

- The mean of X and Y are the same:

$$M_Y = M_X = M_{..} . \quad (9.43)$$

- Equation for the intercept

$$a_{\text{regression}} = M_Y - bM_X . \quad (9.44)$$

- Equation for the slope

$$b = \frac{SCP_{YX}}{SS_X} . \quad (9.45)$$

9.8.2 Rewriting SS_X

The first step is to rewrite SS_X , taking into account that $X = M_{a.}$.

$$\begin{aligned} SS_X &= \sum_{a,s} (X - M_X)^2 \\ &= \sum_{a,s} (M_{a.} - M_{..})^2 \\ &= S \sum_a (M_{a.} - M_{..})^2 . \end{aligned} \quad (9.46)$$

Developing the square in Equation 9.46 gives

$$SS_X = S \sum (M_{a.}^2 + M_{..}^2 - 2M_{..}M_{a.}) . \quad (9.47)$$

Distributing the $\sum$ sign gives

$$SS_X = S \left(\sum M_{a.}^2 + \sum M_{..}^2 - 2 \sum M_{..}M_{a.} \right) . \quad (9.48)$$

We note that

$$\sum_a M_{..}^2 = AM_{..}^2 \quad (9.49)$$

and that

$$-2 \sum M_{..}M_{a.} = -2M_{..} \sum M_{a.} \quad (9.50)$$

$$\begin{aligned} &= -2M_{..}(AM_{a.}) \\ &= -2AM_{..}^2 . \end{aligned} \quad (9.51)$$

Therefore, Equation 9.46 reduces to

$$SS_X = S \left(\sum M_{a.}^2 - AM_{..}^2 \right) . \quad (9.52)$$

9.8.3 Rewriting SCP_{YX}

Recall that the sum of the cross-products between Y and X is equal to

$$SCP_{YX} = \sum (Y_{as} - M_Y)(X_{as} - M_X) . \quad (9.53)$$

This is equivalent to

$$SCP_{YX} = \sum (Y_{as} - M_{..})(M_{a\cdot} - M_{..}) . \quad (9.54)$$

Developing the product gives

$$SCP_{YX} = \sum (Y_{as}M_{a\cdot} - M_{..}M_{a\cdot} - Y_{as}M_{..} + M_{..}^2) . \quad (9.55)$$

Distributing the $\sum$ gives

$$SCP_{YX} = \sum Y_{as}M_{a\cdot} - M_{..} \sum M_{a\cdot} - M_{..} \sum Y_{as} + ASM_{..}^2 . \quad (9.56)$$

Noting that

$$M_{..} \sum_{a,s} Y_{as} = M_{..} S \sum_a M_{a\cdot} = ASM_{..}^2 . \quad (9.57)$$

and that (because $\sum_s Y_{as} = SM_{a\cdot}$)

$$\sum Y_{as}M_{a\cdot} = \sum_a M_{a\cdot} \sum_s Y_{as} = S \sum_a M_{a\cdot}^2 , \quad (9.58)$$

Equation 9.54 becomes

$$SCP_{YX} = S \left(\sum_a M_{a\cdot}^2 - AM_{..}^2 \right) . \quad (9.59)$$

Comparing this last result with Equation 9.52, we find that

$$SCP_{YX} = SS_X , \quad (9.60)$$

therefore

$$b = \frac{SCP_{YX}}{SS_X} = \frac{SS_X}{SS_X} = 1 . \quad (9.61)$$

9.8.4 $a_{\text{regression}} = 0$

To find the value of $a_{\text{regression}}$ we just plug the value of b in the intercept equation:

$$\begin{aligned} a_{\text{regression}} &= M_Y - bM_X \\ &= M_{..} - 1 \times M_{..} \\ &= 0 . \end{aligned} \quad (9.62)$$

Et voilà!

9.8.5 Recap

Finally, we find that, when we use $M_{a\cdot}$ as a predictor for Y_{as} , the regression equation becomes

$$\begin{aligned} \hat{Y} &= a_{\text{regression}} + bX \\ &= 0 + 1 \times M_{a\cdot} \\ &= M_{a\cdot} . \end{aligned} \quad (9.63)$$

9.9 Multiple regression and analysis of variance

We have chosen to link regression and analysis of variance by using the ‘mean model’ (i.e. the mean of their group predicts subjects’ performance). The main advantage of this presentation is to make the use of *simple* regression possible, no matter the number of levels of the independent variable. We could have used different coding schemes.

When dealing with two groups, any coding of the independent variable will do, as long as we give different values to X for each group. For example, we can give the value of $+1$ to the subjects of the first group and the value of -1 to the subjects of the second group (this is called *contrast coding*). The meaning of the expression ‘contrast coding’ should become clear after Chapter 12 dedicated to the notion of contrast. Another coding scheme is to give the value of 1 to the subjects of the first group and the value of 0 to the subjects of the second group. This is called a *group coding* (also sometimes ‘dummy coding’). The sums of squares, degrees of freedom, mean squares, and the F ratio will all be the same as with the mean model coding. The property of having $\hat{Y} = M_a$ will also remain. However, the values of the parameters $a_{\text{regression}}$ and b will depend upon the coding. The main point is that with two groups we have *one* independent variable with *one* degree of freedom. Hence, using one predictor means that we use as many predictors as the independent variable has degrees of freedom to predict the dependent variable. As an exercise, try to compute the regression with these two different codings for the imagery example. What are the values of $a_{\text{regression}}$ and b ?

When dealing with an independent variable with more than two groups (e.g. ‘Romeo and Juliet’), we can also use the *multiple* regression approach. The reasoning behind the multiple regression approach is to use as many predictors (i.e. as many X , T , etc.) as there are degrees of freedom for the independent variable. Because the independent variable named A has $A - 1$ degrees of freedom, we will need $A - 1$ predictors. We can use a contrast approach, or a group approach. The sums of squares, degrees of freedom, mean squares, and F will be the same as with the ‘mean model’. However, they will be more complex to compute. Having a good grasp of this technique requires some understanding of the notion of contrasts, and consequently we postpone its full discussion until after Chapter 12. For now, let us say that if we use the ‘group coding’ approach, we will use $A - 1$ contrasts, the subjects of group 1 will all have a value of 1 for the first predictor, and a value of 0 for all the other predictors. Subjects of group 2 will have a value of 1 for the second predictor and 0 for all the other predictors, and so on until the $(A - 1)$ th group, whose subjects will have a value of 1 on the last predictor [i.e. the $(A - 1)$ th predictor]. The subjects of the last group will have a value of 0 on all predictors. For the contrast approach, let us define a contrast as a set of numbers (not all equal to zero) with a mean of zero. Almost any set of contrasts will do, as long as all subjects in a group are assigned the same number for the contrast. Some sets of contrasts (i.e. pairwise orthogonal) are of special interest as we shall see later.⁵

⁵ If you have not yet read the chapter on contrasts, and if you have the impression of being in the same condition as the subjects in the condition ‘information without context’ of Bransford’s experiment, don’t panic—this really is the case! The last paragraph is included for a fuller understanding on a second (or further) reading.

Chapter summary

9.10 Key notions of the chapter

Below is the main notion introduced in this chapter. If you have problems understanding it, you may want to re-read the part of the chapter in which it is defined and used. One of the best ways is to write down a definition of this notion by yourself with the book closed.

Mean model. When we predict the performance of subjects from the mean of their group, the predicted value turns out to be the group mean too.

9.11 Key formulas of the chapter

Below are the main formulas introduced in this chapter: try to go through them and understand what they mean.

$$\hat{Y} = a_{\text{regression}} + bX = 0 + 1 \times X = M.$$

Regression notation	ANOVA notation
Y	Y_{as}
$\hat{Y}$	$M_{a.}$
X	$M_{a.}$
M_Y	$M_{..}$
M_X	$M_{..}$
$SS_{\text{regression}}$	SS_{between} or SS_A
SS_{residual} or SS_{error}	SS_{within} or $SS_{S(A)}$
$R^2_{Y \cdot X}$ or $r^2_{Y \cdot X}$ or $r^2_{\hat{Y} \cdot Y}$	$R^2_{Y \cdot A}$
N	$A \times S$

Table 9.8 More key formulas ...

$$F = \frac{MS_{\text{regression}}}{MS_{\text{residual}}} = \frac{\hat{\sigma}^2_{\text{regression}}}{\hat{\sigma}^2_{\text{residual}}} = \frac{MS_{\text{between}}}{MS_{\text{within}}}$$

9.12 Key questions of the chapter

Below are some questions about the content of this chapter. All the answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- * If regression and analysis of variance are one thing, why keep two different techniques?
- * What is the purpose of the mean model?
- * When using the mean model, what is the value of the intercept? What is the value of the slope?

10

ANOVA, one factor: score model

10.1 Introduction

10.1.1 Motivation: why do we need the score model?

Running an analysis of variance is equivalent to separating or partitioning the variability of the scores of an experiment into several sources. The deviation from each score Y_{as} to the grand mean $M_{..}$ is now decomposed into two basic components: the first one is the deviation from the group mean to the grand mean (i.e. the between-group deviation), the second one is the deviation from each score to the group mean (i.e. the within-group deviation). When this partition is formalized it is called the *score model* of the analysis of variance.

In addition to being clear and precise, the score model has several uses:

- It is necessary to express correctly the so-called ‘validity assumptions’ that are needed (in addition to the null hypothesis) to derive the sampling distribution of F .
- It is the key to understanding complex factorial designs (i.e. designs involving more than one independent variable). In particular, one needs the score model to find out what mean squares should be used in order to compute the F ratio.
- It is used also to evaluate the size of the experimental effect (i.e. how much power does the experimental manipulation have to affect subjects’ behavior).
- The statistical rationale behind the analysis of variance is directly derived from the score model.

10.1.2 Decomposition of a basic score

Most of what has been presented up to now is based on the relationship:

$$\begin{array}{llll} \text{Deviation of the score} & = & \text{between-group} & + & \text{within-group} \\ \text{from the grand mean} & & \text{deviation} & & \text{deviation} \end{array} \quad (10.1)$$

Or, with a formula,

$$Y_{as} - M_{..} = (M_{a.} - M_{..}) + (Y_{as} - M_{a.}). \quad (10.2)$$

This relation can be rewritten as

$$Y_{as} = M_{..} + (M_{a.} - M_{..}) + (Y_{as} - M_{a.}) \quad (10.3)$$

which is equivalent to partitioning Y_{as} into three terms.

10.1.3 Fixed effect model

At this point we need to add an assumption about the independent variable, namely that all the levels of the independent variable are included in the experiment. When this is the case, we say that we have a *fixed* factor. For a fixed factor any replication of our experiment will use the same levels as the original experiment, and the experimenter is interested only in the levels of the factor used in the experiment. Fixed factors are contrasted with *factors* which include only *some* of the levels of the independent variable of interest (these last factors are called *random*, see Section 10.2, page 197).

When $\mathcal{A}$ is a fixed factor, the components of the decomposition of the basic score Y_{as} can be interpreted as follows:

- $M_{..}$.
This is the grand mean. It corresponds to the expected value of the dependent variable when the effect of the independent variable is not taken into account.
- $M_{a.} - M_{..}$.
This is the deviation, or the distance, from the a th experimental group to the grand mean. It expresses the effect of the a th experimental condition on the subjects' behavior.
- $Y_{as} - M_{a.}$.
This is the deviation from the score of subject s in the a th experimental condition to the mean of all the subjects in the a th experimental condition. It expresses the differences of behavior or responses of subjects placed in the *same experimental condition*. Intuitively, there are two main reasons why subjects in the same experimental condition would give different responses. The first is the *error* of measurement (e.g. a value is misread by the experimenter, the measurement device is faulty). The second can be the result of a differential effect of the experimental condition on the subject (i.e. all the subjects do not react in the same way to the experimental manipulation). If this is the case we say that there is an *interaction* between the subject factor and the experimental factor (the notion of interaction will be detailed more in the chapters dealing with multiple factor experiments, see, e.g., Chapter 16). Because $\mathcal{S}$ (the subject factor) is nested in $\mathcal{A}$ (the experimental factor), it is not possible to assess separately the effects of each of these possible causes.

In summary, a score Y_{as} can be decomposed as

$$Y_{as} = \begin{array}{lll} \text{mean value of the} & + & \text{effect of the} & + & \text{error linked to} \\ \text{DV when} & & \text{ath level of} & & \text{subject } s \text{ nested in} \\ \text{ignoring the IV} & & \text{the IV} & & \text{the ath level of the} \\ & & & & \text{IV} \end{array} \quad (10.4)$$

(DV, dependent variable; IV, independent variable.)

This decomposition is made using the values Y_{as} , $M_{a.}$, and $M_{..}$ which are computed from the sample of subjects used in the experiment. However (cf. Chapter 8, on the notion of statistical test), the sample of subjects is used to make an *inference* on the whole *population*.

So, in order to distinguish the statistics computed on a sample from the equivalent *parameters* of the population, a specific notation is needed.

The grand mean of the population is denoted $\mu_{..}$ (read ‘mu dot dot’). The mean of the group a for the population is denoted $\mu_{a\cdot}$ (read ‘mu a dot’).

So,

- $M_{..}$ is a statistic computed for a sample, estimating the $\mu_{..}$ parameter of the population (whose exact value is unknown).
- $M_{a\cdot}$ is a statistic computed for a sample, estimating the $\mu_{a\cdot}$ parameter of the population (whose exact value is unknown).

Using these notations we can now define the ‘score model’ in terms of the parameters of the population:

$$Y_{as} = \mu_{..} + (\mu_{a\cdot} - \mu_{..}) + (Y_{as} - \mu_{a\cdot}) \quad (10.5)$$

with:

- $\mu_{..}$ mean of the dependent variable for the population.
- $(\mu_{a\cdot} - \mu_{..})$ effect of the a th level of the independent variable for the population.
- $(Y_{as} - \mu_{a\cdot})$ error associated with the score Y_{as} .

In the same way as (for a balanced design)

$$M_{..} = \frac{\sum M_{a\cdot}}{A} \quad (10.6)$$

the condition

$$\mu_{..} = \frac{\sum \mu_{a\cdot}}{A}, \text{ which is equivalent to } A\mu_{..} = \sum \mu_{a\cdot} \quad (10.7)$$

is imposed on the parameters of the population. As a consequence,

$$\sum (\mu_{a\cdot} - \mu_{..}) = \sum \mu_{a\cdot} - (A\mu_{..}) = A\mu_{..} - A\mu_{..} = 0. \quad (10.8)$$

In general, the score model is expressed using a slightly more condensed notation. Specifically, the score Y_{as} is decomposed as

$$Y_{as} = \mu_{..} + \alpha_a + e_{s(a)}. \quad (10.9)$$

This notation is expanded as

$$\alpha_a = \mu_{a\cdot} - \mu_{..} \quad (10.10)$$

where α_a is the effect of the a th level of the independent variable. Note also that:

$$\sum_a \alpha_a = 0 \quad (10.11)$$

(cf. Equation 10.8). The parameter α_a (note that this is the Greek letter alpha) is estimated by $(M_{a\cdot} - M_{..})$, and $e_{s(a)}$ by $(Y_{as} - M_{a\cdot})$.

Warning. The previous notation is quite standard, and that is the reason why it is used here. Be careful, however, to take note that α_a in the previous equation has nothing to do with the notation α used to denote the Type I error (e.g. as in the expression $\alpha = .01$).

10.1.4 Some comments on the notation

The term α_a denotes a parameter of the population. Because the factor $\mathcal{A}$ is assumed to be fixed, if the experiment is replicated, the value of α_a will be the same for all replications (and the same levels of $\mathcal{A}$ will be used as well). In general, in order to denote the effect of a fixed factor we will use the lower case Greek letter equivalent to the letter used to denote the name of the source. For example, for the factor $\mathcal{A}$, the lower case Greek letter will be α and the subscript will be a , and the effect of the a th level of $\mathcal{A}$ will be denoted α_a .

The notation $e_{s(a)}$ denotes the effect of the s th level of the factor $\mathcal{S}$ nested in (the a th level of) the factor $\mathcal{A}$, as the subscript makes clear. This effect is considered as being equivalent to the experimental error, which is the reason why the letter e (like *error*) is used to denote it. If the experiment is replicated, the subjects of the new experiment will be different from the subjects of the previous one, hence the error components of the new experiment will be different from the ones from the old experiment (compare this with the previous paragraph about α_a). Because we suppose that the samples of subjects are *random* samples, the error components will also be random [i.e. the factor $\mathcal{S}(\mathcal{A})$ is random]. In general, the effect of a random factor will be denoted with a lower case roman letter (specifically here e).

In all cases (i.e. random or fixed) the subscripts will correspond to the name of the source. For example, the subscript $s(a)$ in $e_{s(a)}$ corresponds to the factor $\mathcal{S}(\mathcal{A})$.

10.1.5 Numerical example

In the experiment ‘imagery and memory’ (see Table 7.2, page 135) used to introduce the basic notions of the analysis of variance, Subject 2 of the experimental group (i.e. level 2 of factor $\mathcal{A}$) obtained a score of 8. The mean of the second group is 10, the grand mean is 7. Hence the score of subject 2 in group 2 is decomposed as

$$Y_{22} = \mu_{..} + \alpha_2 + e_{2(2)} . \quad (10.12)$$

However, the values of $\mu_{..}$, α_2 , as well as $e_{2(2)}$ are unknown. We can estimate them from the equivalent statistics computed from the sample. Specifically,

- $\mu_{..}$ is estimated by $M_{..} = 7$.
- α_2 is estimated by $\text{est}\{\alpha_2\} = \hat{\alpha}_2 = M_{2.} - M_{..} = 10 - 7 = 3$ (read ‘estimation of alpha two’ or ‘alpha-two hat’).
- $e_{2(2)}$ is estimated by $\text{est}\{e_{2(2)}\} = (Y_{22} - M_{2.}) = 8 - 10 = -2$.

With these estimations, the score Y_{22} is then expressed as:

$$Y_{22} = 7 + 3 - 2 = 8 . \quad (10.13)$$

Along the same lines, Y_{15} , the score of subject 5 in group 1, is 6 and is decomposed as

$$Y_{15} = \mu_{..} + \alpha_1 + e_{5(1)} = 6 \quad (10.14)$$

with

- α_1 estimated by $\text{est}\{\alpha_1\} = \hat{\alpha}_1 = M_{1.} - M_{..} = 4 - 7 = -3$
[note in passing that $\sum \text{est}\{\alpha_a\} = \hat{\alpha}_1 + \hat{\alpha}_2 = 3 - 3 = 0$].
- $e_{5(1)}$ estimated by $\text{est}\{e_{5(1)}\} = (Y_{15} - M_{1.}) = 6 - 4 = 2$.

As an exercise try to express all the scores of this example, and check that

$$\sum_s^S \text{est} \{e_{s(a)}\} = 0 \quad \text{for all } a. \quad (10.15)$$

10.1.6 Score model and sum of squares

The different components of the sum of squares can also be expressed with the score model. The *grand mean* can be expressed by (recall that $\sum \alpha = 0$):

$$M_{..} = \frac{Y_{..}}{AS} = \frac{1}{AS} \sum_{as} (\mu_{..} + \alpha_a + e_{s(a)}) = \mu_{..} + \bar{e}_{.(.)}, \quad (10.16)$$

where

$$\bar{e}_{.(.)} = \frac{e_{.(.)}}{AS} \quad (10.17)$$

The bar in $\bar{e}$ indicates that this is the average error, read ‘*e*-bar dot dot’. An additional caveat is that the mean of the error is supposed to be zero for the whole population of interest, *not* for the sample (because of the sampling error) even though the mean of the estimate is zero.

The *group means* are given by

$$M_{a.} = \frac{Y_{a.}}{S} = \frac{1}{S} \sum_s (\mu_{..} + \alpha_a + e_{s(a)}) = \mu_{..} + \alpha_a + \bar{e}_{.(a)}, \quad (10.18)$$

where

$$\bar{e}_{.(a)} = \frac{1}{S} \sum_s e_{s(a)} \quad (10.19)$$

is the average error for group a .

With these notations, the sum of squares between groups becomes

$$SS_A = S \sum (M_{a.} - M_{..})^2 = S \sum (\alpha_a + \bar{e}_{.(a)} - \bar{e}_{.(.)})^2 = S \sum \text{est} \{\alpha_a\}^2, \quad (10.20)$$

and the sum of squares within-groups becomes

$$SS_{S(A)} = \sum (Y_{as} - M_{a.})^2 = \sum (e_{s(a)} - \bar{e}_{.(a)})^2 = \sum \text{est} \{e_{s(a)}\}^2. \quad (10.21)$$

Note that the within-group sum of squares is composed only of error terms. By contrast, the between-group sum of squares is composed of an error term *plus* a term reflecting the effect of the independent variable. We have seen previously that the sums of squares are used to compute the mean squares when divided by their degrees of freedom. If the experiment is replicated, the values computed for the statistics will differ from one replication to the other (because of the error terms of the sums of squares). What would be their most probable values? Equivalently, if we repeat the experiment an infinite number of times, what would be the average values of these statistics? Intuitively, the average value of a statistic when computed on an infinite number of samples should be the value of the parameter it estimates. When this is the case, the parameter is said to be *unbiased* (cf. Mood *et al.* 1974). The sums of squares, as well as the mean squares, have the nice property of being unbiased estimators. In other words, the *expected value* of the statistic is the parameter (when the statistic is an unbiased estimator, the notion of expected value is explained in Appendices D and F). The expected value can be computed quite straightforwardly following a set of simple rules (described in detail in Appendix F).

Using these rules, the expected values of the mean squares are found to be

$$\begin{aligned}\mathbf{E}\{MS_A\} &= \sigma_e^2 + S\vartheta_a^2 \\ \mathbf{E}\{MS_{S(A)}\} &= \sigma_e^2,\end{aligned}\tag{10.22}$$

with

$$\vartheta_a^2 = \frac{\sum \alpha_a^2}{(A-1)} \quad (\text{read 'theta squared a'}), \tag{10.23}$$

and

$$\sigma_e^2 : \text{error variance} \quad (\text{read 'sigma squared e'}). \tag{10.24}$$

When the null hypothesis is true, all the α_a values are null (there is no effect of the independent variable) and therefore all the α_a^2 values are equal to 0, and hence $\vartheta_a^2 = 0$. Consequently,

$$\mathbf{E}\{MS_A\} = \mathbf{E}\{MS_{S(A)}\}. \tag{10.25}$$

So, when the null hypothesis is true, the mean square between group and the mean square within group estimate the same quantity, namely the *error variance*. Also, because these estimations are based on two different sums of squares, they are independent (at least when the validity assumptions detailed later hold true). It can be shown that the sampling distribution of the ratio of two independent estimations of the same variance is a *Fisher* distribution (e.g. Edwards, 1964; Reny, 1966; Hoel, 1971; Mood *et al.*, 1974; Hays, 1981). The proof of this statement is a bit technical and can require some mathematical sophistication, but, as far as we are concerned, we can accept it as given. The point here is that the score model is the necessary tool for the derivation.

10.1.7 Digression: why ϑ_a^2 rather than σ_a^2 ?

When $\mathcal{A}$ is a fixed factor, the A levels of $\mathcal{A}$ represent *all* the possible levels of $\mathcal{A}$. Because of this, they constitute the *entire population* of possible levels of $\mathcal{A}$. Strictly speaking the term 'variance' is reserved for denoting the variance of a population. When a population consists of a finite number of elements, the variance is obtained by dividing the sum of squares by the *number of elements* in that population. We divide the sum of squares by $A - 1$ only when we wish to *estimate* the variance of a (theoretically infinite) population on the basis of a sample drawn from that population. Precisely, the variance of factor $\mathcal{A}$ is given by

$$\sigma_a^2 = \frac{\sum \alpha_a^2}{A}. \tag{10.26}$$

Recall that since $\sum \alpha_a = 0$ the mean of all the effects of $\mathcal{A}$ is zero. Therefore, $\sum \alpha_a^2$ is the sum of squares of deviations from the mean. We obtain the variance by dividing the sum of squares by A . The expected value of the sum of squares of $\mathcal{A}$ is equal to

$$\mathbf{E}\{SS_A\} = (A-1)\sigma_e^2 + S\sum \alpha_a^2$$

To obtain the expected value of the mean square of $\mathcal{A}$, we divide the sum of squares by $A - 1$, giving

$$\mathbf{E}\{MS_A\} = \sigma_e^2 + \frac{S\sum \alpha_a^2}{A-1}.$$

It is very important to distinguish between the two quantities,

$$\frac{\sum \alpha_a^2}{A} \quad \text{and} \quad \frac{\sum \alpha_a^2}{A-1}.$$

Therefore, we introduce a specific notation, distinguishing between

$$\sigma_a^2 = \sum \frac{\alpha_a^2}{A},$$

which we call the *variance* of $\mathcal{A}$; and

$$\vartheta_a^2 = \sum \frac{\alpha_a^2}{A-1},$$

which represents the effect of the variable $\mathcal{A}$ and is called the component of the *mean square* of $\mathcal{A}$.

Note that the difference between σ_a^2 and ϑ_a^2 is not trivial. For example, when $A = 2$, ϑ_a^2 is twice as large as σ_a^2 .

In general, note that ϑ_a^2 and σ_a^2 are related as follows:

$$\sigma_a^2 = \frac{A-1}{A} \vartheta_a^2 \quad \text{or} \quad \vartheta_a^2 = \frac{A}{A-1} \sigma_a^2.$$

The distinction between ϑ_a^2 and σ_a^2 becomes particularly important in constructing measures of the intensity of the effect of $\mathcal{A}$.

10.2 ANOVA with one random factor (Model II)

10.2.1 Fixed and random factors

In the preceding paragraphs we developed the ANOVA for one fixed factor; that is, for the case where the different levels of the independent variable presented in the experiment are the only ones of interest to the experimenter. Model I designs involving fixed factors make up most of the one-factor designs in psychological research. Nevertheless, it can happen that the independent variable is a random factor, meaning that the levels of the independent variable are drawn from a population, and that a probability of being drawn is associated with each level of the independent variable (see Appendix C for a refresher on probability). We owe the distinction between fixed and random factors—that is, between Model I and Model II—to Eisenhart *et al.* (1947). With Model I (fixed effects) designs we are generally interested in the effect of each level of the independent variable. In contrast, with Model II (random effects) designs we are concerned with the overall effect of the independent variable, *because we cannot know the levels of a random factor a priori*. Some authors introduce the distinction between fixed and random factors by presenting the former as measured ‘without error’, and the latter as measured ‘with an error term’. In fact, these different presentations are equivalent, if we identify the error term with fluctuations due to sampling. Nevertheless, the initial presentation, which is clearer and more precise, is to be preferred.

For an experimental design involving one factor—that is, an $\mathcal{S}(\mathcal{A})$ design—the distinction between Models I and II is essentially a methodological point. The calculations and the ANOVA table are identical in the two cases. In particular, the value of F is identical, and we compare it to the same critical value F_{critical} . The essential difference lies in the researcher’s focus of interest. With a fixed factor, the interest is in the effect of each level of the independent variable

	Experimental groups				
	Group 1	Group 2	Group 3	Group 4	Group 5
	40	53	46	52	52
	44	46	45	50	49
	45	50	48	53	49
	46	45	48	49	45
	39	55	51	47	52
	46	52	45	53	45
	42	50	44	55	52
	42	49	49	49	48
$M_{a.}$	43	50	47	51	49

Table 10.1 Results of (fictitious) experiment on face perception (see text for explanation).

and therefore in the intensity of the α_a values. In contrast, with a random factor, the interest is with the overall influence of the levels of the independent variable on the dependent variable. In that case the particular levels of the independent variable hold no intrinsic interest, *because they change with each replication of the experiment*.

10.2.2 Example: $\mathcal{S}(\mathcal{A})$ design with $\mathcal{A}$ random

In a series of experiments on face perception we examined whether the degree of attention devoted to each face varies from face to face. In order to verify this hypothesis, we assign 40 Nundergraduate students to five experimental conditions. For each condition we have a man’s face drawn at random from a collection of several thousand faces. We use the subjects’ pupil dilation when viewing the face as an index of the attentional interest evoked by the face. The results are presented in Table 10.1 (with pupil dilation expressed in arbitrary units).

The previous procedures for calculation apply here as well. By way of review make your calculations on a separate piece of paper. You should obtain the following ANOVA table:

Source	df	SS	MS	F	$Pr(F)$
$\mathcal{A}$	4	320.00	80.00	10.00	.000020
$\mathcal{S}(\mathcal{A})$	35	280.00	8.00		
Total	39	600.00			

From this table it is clear that the research hypothesis is supported by the experimental results: all faces do not attract the same amount of attention.

10.3 The score model: Model II

The score model for a one-factor ANOVA, Model II, needs to take into account that the independent variable is a random variable. The score model will be essentially similar to that for Model I, except for this difference. To call attention to this difference we shall modify our

notation as follows:

$$Y_{as} = \mu_{..} + a_a + e_{s(a)}$$

with

- Y_{as} : score for the s th individual in group a ;
- $\mu_{..}$: population mean;
- a_a : effect of level a of the independent variable
(note the use of the Roman letter a , as opposed to the Greek α);
- $e_{s(a)}$ error associated with observation s in group a .

As in the case with a fixed factor, we impose the following conditions on the error term:

$$\mathbf{E}\{e_{s(a)}\} = 0 \quad \text{and} \quad \mathbf{E}\{e_{s(a)}^2\} = \sigma_e^2.$$

An important difference between the two models is that the sum of the a_a values is not zero for the random model. However, the expected value of a_a is zero:

$$\mathbf{E}\{a_a\} = 0.$$

In addition, we require that the a_a be distributed normally with variance σ_a^2 :

$$\mathbf{E}\{a_a^2\} = \sigma_a^2$$

(note that we write σ_a^2 here, and not ϑ_a^2).

Following this model, the expected values of the mean squares will be:

$$\mathbf{E}\{MS_A\} = \sigma_e^2 + S\sigma_a^2$$

$$\mathbf{E}\{MS_{S(A)}\} = \sigma_e^2.$$

Notice that the principal difference between the fixed factor design and the random factor design is that the MS_A includes a term corresponding to the variance for the effect of the independent variable (i.e. the term σ_a^2) instead of a variance-like corresponding to the effect of a fixed independent variable (i.e. the term ϑ_a^2).

The null hypothesis assumes the value of all the a_a in the population to be zero. In particular it supposes a value of zero for the a_a corresponding to the modalities randomly selected to represent the independent variable. Consequently, under the null hypothesis $\sigma_a^2 = 0$ and

$$\mathbf{E}\{MS_A\} = \sigma_e^2.$$

Thus MS_A and $MS_{S(A)}$ estimate *the same error variance* under the null hypothesis. When the 'validity assumptions (see Chapter 11, Section 11.2, page 211) are satisfied, and when the null hypothesis is true, the F ratio of these two mean squares will follow a Fisher distribution (with the appropriate degrees of freedom as parameters). We can thus evaluate the probability associated with each possible value of the criterion, and decide to reject the null hypothesis when the associated probability is less than the alpha level.

10.4 $F < 1$ or the strawberry basket

When $MS_A < MS_{S(A)}$, then $F < 1$. This could be an indication that the ANOVA model is not correct. One question of interest in this case is to determine whether F is really smaller than one. In other words, is $MS_{S(A)} > MS_A$? This corresponds to a statistical test with H_0 stating

the equality of the two variances estimated by the mean squares and H_1 stating that $MS_{S(A)}$ is larger than MS_A . Because the direction of the difference is now specified, the test becomes a one-tailed test. To find the probability associated with a value of $F < 1$ (which comes down to considering the area under the F distribution to the *left* of the observed value, rather than the area to the *right* that we use for $F > 1$). In such cases an interesting property of the Fisher distribution is that the probability of obtaining a value of $F < 1$ with ν_1 and ν_2 degrees of freedom is identical to the probability associated with $1 / F$ with ν_2 and ν_1 degrees of freedom. (Note the reversal of the df parameters.) For example, to find the critical value of $F < 1$ with $\nu_1 = 5$, $\nu_2 = 20$ and $\alpha = .05$, look up the value for $\nu_1 = 20$ and $\nu_2 = 5$. We find $F_{\text{critical}} = 4.56$. F_{critical} with $F < 1$ for $\nu_1 = 5$ and $\nu_2 = 20$ is therefore equal to $1 / 4.56 = .22$.

Note that when $F < 1$, we reject H_0 when $F < F_{\text{critical}}$. To be precise, a value of F *significantly* smaller than 1 indicates that the group means are more similar to each other than we would expect on the basis of fluctuations due to sampling. There are two principal sources from which this could arise:

- first, there may be a factor ‘homogenizing’ the group means (as in the case of the ‘strawberry basket’ described below);
- second, there may be a factor affecting the dependent variable that is ignored in the experimental design.

10.4.1 The strawberry basket

To illustrate the effect of homogenization of groups, consider the parable of the ‘Strawberry farmer and his baskets’. Once upon a time a farmer who was about to take his strawberries to the market became infatuated with random sampling (he had just read *The Joy of Randomization*). So he randomly assigned the strawberries to the baskets. He had 100 baskets, and for each strawberry he looked up a two-digit number in a random-number table to pick the basket to put it in. He ended up with some beautiful baskets brimming over with great big strawberries, some medium baskets that were more or less full of medium-sized strawberries, and some skimpy baskets which had a preponderance of small strawberries. Naturally the people who came to the market preferred the beautiful baskets and ignored the skimpy baskets. As a result, the farmer sold all his beautiful baskets, but none of the skimpy baskets.

To avoid this sad ending, the next time the farmer took his strawberries to the market he *homogenized* the baskets instead of using random assignment. That is, he arranged the baskets so that it *looked* as though chance could not have been responsible for the distribution of strawberries into baskets—the baskets all looked very much alike (in fact *too much* alike!). And because of his love for statistics, he satisfied himself by means of an ANOVA using strawberries as subjects, the weights of the strawberries as the dependent variable, and baskets as the random factor A . He found $F < 1$, and after selling all the baskets congratulated himself on his success in homogenizing the baskets.

This parable is somewhat classic for statisticians. Therefore, when we suspect that some factor is making the group means *too* similar, we will call this a ‘strawberry basket effect’.

10.4.2 A hidden factor augmenting error

Consider another example: You wish to design an experiment with two conditions corresponding to two methods for the teaching of lists of words. Since you have a grant,

you hire two experimenters to carry out the experiment. Having taken to heart Chapters 1 (which you should have read) and 15 (which you may not have read yet ...) of this book, you wish to avoid confounded variables, and so you ask each experimenter to serve in each of the experimental conditions. So far, so good. Now suppose that one of the experimenters is fairly charismatic and inspires the subjects, while the other experimenter is far from inspiring. *If you do not take this variable into account, you will increase the error in the two conditions.* In effect, half the subjects in each condition will perform better and half will perform worse, all due to the difference between the experimenters. In other words, the effect of the hidden independent variable ‘experimenter’ is *confounded* with the experimental error, since we have neglected to take it into account. Because of this, the eventual effect of the independent variable might be masked by the artificial increase of error. If this increase is substantial, you could obtain a value of F that is less than 1. Note in passing that a slight increase in error can result in a failure to detect a real effect of the independent variable. This is an additional good reason not to accept H_0 . *Failure to detect an effect of the independent variable can simply indicate that you have forgotten to take into account a factor that is confounded with the experimental error.* One cure is to include the variable as an independent variable in the analysis. Another cure is to bring the variable under better control (as, for example, by hiring more comparable experimenters).

10.5 Size effect coefficients derived from the score model: ω^2 and ρ^2

From the score model we can derive two coefficients which estimate the size of the experimental effect as the ratio of the experimental variance to the sum of all sources of variance. The computation of these indices is different for the fixed effect and the random effect and therefore these indices have different names: ω^2 is used for a fixed effect model and ρ^2 is used for a random effect model. The estimations of the indices are denoted with a caret (i.e. $\hat{\omega}^2$ and $\hat{\rho}^2$)

In an $S(\mathcal{A})$, there are two sources of variance, $\mathcal{A}$ and $S(\mathcal{A})$. Therefore when $\mathcal{A}$ is fixed, the effect size parameter for the population, denoted $\omega^2_{\mathcal{A} \cdot Y}$, is defined as

$$\omega^2_{\mathcal{A} \cdot Y} = \frac{\sigma_a^2}{\sigma_e^2 + \sigma_a^2},$$

where

$$\sigma_a^2 = \frac{\sum \alpha_a^2}{A} = \frac{A-1}{A} s_a^2 = \text{variance of } \mathcal{A}$$

$$\sigma_e^2 = \text{error variance} = \mathbf{E} \left\{ e_{s(e)}^2 \right\}.$$

When $\mathcal{A}$ is random, the effect size parameter for the population, denoted $\rho^2_{\mathcal{A} \cdot Y}$, is defined as

$$\rho^2_{\mathcal{A} \cdot Y} = \frac{\sigma_a^2}{\sigma_e^2 + \sigma_a^2},$$

where:

$$\sigma_a^2 = \text{variance of } \mathcal{A} = \mathbf{E} \{ a_a^2 \}$$

$$\sigma_e^2 = \text{error variance} = \mathbf{E} \left\{ e_{s(e)}^2 \right\}.$$

These effect size coefficients can also be interpreted as the proportion of the dependent variable that can be explained by the independent variable, or the information provided by the independent variable.

In order to estimate the value of these coefficients we replace the population variances by their estimation. This procedure is only an approximation because (cf. Hays, 1981; Smith, 1982) the ratio of two estimations is not equal to the estimation of this ratio.

10.5.1 Estimation of $\omega_{\mathcal{A},Y}^2$

From the score model we have

$$\begin{aligned}\mathbf{E}\{MS_A\} &= \sigma_e^2 + S\vartheta_a^2 \\ \mathbf{E}\{MS_{S(A)}\} &= \sigma_e^2\end{aligned}$$

which can be rewritten as

$$MS_A = \text{est}\{\sigma_e^2 + S\vartheta_a^2\} = \text{est}\{\sigma_e^2\} + \text{est}\{S\vartheta_a^2\}$$

(with $\text{est}\{\}$ meaning ‘estimation of’). We also have:

$$MS_{S(A)} = \text{est}\{\sigma_e^2\}.$$

If we remember that:

$$\sigma_a^2 = \frac{\sum \alpha_a^2}{A} = \frac{A-1}{A} \vartheta_a^2,$$

We can estimate σ_a^2 from MS_A and $MS_{S(A)}$ with the following steps:

1. $MS_A - MS_{S(A)} = \text{est}\{\sigma_e^2\} + \text{est}\{S\vartheta_a^2\} - \text{est}\{\sigma_e^2\} = \text{est}\{S\vartheta_a^2\} = S \times \text{est}\{\vartheta_a^2\}$
2. $\frac{MS_A - MS_{S(A)}}{S} = \text{est}\{\vartheta_a^2\}$
3. $\frac{(A-1)(MS_A - MS_{S(A)})}{AS} = \frac{A-1}{A} \text{est}\{\vartheta_a^2\} = \text{est}\{\sigma_a^2\}.$

This gives:

$$\begin{aligned}\widehat{\omega}_{\mathcal{A},Y}^2 &= \frac{\text{est}\{\sigma_a^2\}}{\text{est}\{\sigma_e^2\} + \text{est}\{\sigma_a^2\}} \\ &= \frac{[(A-1)(MS_A - MS_{S(A)})] / (AS)}{MS_{S(A)} + \{[(A-1)(MS_A - MS_{S(A)})] / (AS)\}} \\ &= \frac{SS_A - (A-1)MS_{S(A)}}{SS_{\text{total}} + MS_{S(A)}}. \quad (10.27)\end{aligned}$$

We can illustrate the estimation of $\widehat{\omega}_{\mathcal{A},Y}^2$ with the example of ‘Romeo and Juliet’ (see page 157) where $A = 4$, $S = 5$, $SS_A = 50.95$ and $SS_{S(A)} = 37.60$. Plugging in these values into Equation 10.27 gives:

$$\widehat{\omega}_{\mathcal{A},Y}^2 = \frac{50.95 - 3 \times 2.35}{88.55 + 2.35} = \frac{43.9}{90.9} = .48.$$

10.5.2 Estimating $\rho_{\mathcal{A}.Y}^2$

When $\mathcal{A}$ is a random factor, a procedure similar to the fixed effect case is used to estimate $\rho_{\mathcal{A}.Y}^2$ from the expected value of the mean squares:

$$\mathbf{E}\{MS_A\} = \sigma_e^2 + S\sigma_a^2$$

$$\mathbf{E}\{MS_{S(A)}\} = \sigma_e^2$$

First we estimate σ_e^2 and σ_a^2 :

$$\text{est}\{\sigma_e^2\} = MS_{S(A)}$$

$$\text{est}\{\sigma_a^2\} = \frac{MS_A - MS_{S(A)}}{S}.$$

Then $\rho_{\mathcal{A}.Y}^2$ is estimated as:

$$\begin{aligned}\hat{\rho}_{\mathcal{A}.Y}^2 &= \frac{\text{est}\{\sigma_a^2\}}{\text{est}\{\sigma_e^2\} + \text{est}\{\sigma_a^2\}} \\ &= \frac{(MS_A - MS_{S(A)})/S}{MS_{S(A)} + [(MS_A - MS_{S(A)})/S]} \\ &= \frac{MS_A - MS_{S(A)}}{MS_{S(A)} + (S-1)MS_{S(A)}}.\end{aligned}\tag{10.28}$$

As an illustration, suppose that we have the following results from a random factor design: $A = 5$, $S = 9$, $MS_A = 87.00$, $MS_{S(A)} = 9.00$. Plugging in these values into Equation 10.28 gives:

$$\hat{\rho}_{\mathcal{A}.Y}^2 = \frac{87.00 - 9.00}{87.00 + 8 \times 9.00} = .49.$$

10.5.3 Negative values for ω and ρ

When MS_A is smaller than $MS_{S(A)}$, we obtain negative values for estimated values of both $\hat{\omega}_{\mathcal{A}.Y}^2$ and $\hat{\rho}_{\mathcal{A}.Y}^2$. Because these parameters cannot be negative, we report an estimated value of zero in these cases. Note that when the estimated values of $\hat{\omega}_{\mathcal{A}.Y}^2$ or $\hat{\rho}_{\mathcal{A}.Y}^2$ are negative, the F ratio is smaller than 1 (see Section 10.4, page 199).

10.5.4 Test for the effect size

The null hypothesis for the effect size is equivalent to the null hypothesis of the analysis of variance and therefore the standard F test applies to both null hypotheses.

10.5.5 Effect size: which one to choose?

The effect size coefficients $\hat{\omega}_{\mathcal{A}.Y}^2$ and $\hat{\rho}_{\mathcal{A}.Y}^2$ are derived from the score model. As we have seen in Section 8.8, we can also evaluate the size of the experimental effect using the $R_{Y.A}^2$ coefficient. These different effect size coefficients correspond to different interests. The $\hat{\omega}_{\mathcal{A}.Y}^2$ and $\hat{\rho}_{\mathcal{A}.Y}^2$ coefficients estimate the population effect size (they are equivalent to the shrunken

$\hat{\tau}^2$ described in Section 3.3.1, page 56). The coefficient $R_{Y.A}^2$ is a *descriptive* statistic which evaluates the importance of the effect for a sample. So the choice seems simple: we just need to choose the statistic that matches our interest. But it is often not that simple. For example, $R_{Y.A}^2$ has the advantage of being easy to understand and to compute and this coefficient makes a natural connection with the interpretation of the analysis of variance as regression. Therefore this coefficient is preferred when this connection is important (cf. Kerlinger, 1973; Cohen and Cohen, 1975; Pedhazur, 1982; Draper and Smith, 1982). In addition, contrary to $\hat{\omega}_{A.Y}^2$ and $\hat{\rho}_{A.Y}^2$, coefficients such as $R_{Y.A}^2$ can be computed for all types of designs. This probably explains the frequent use of this coefficient. Incidentally, $R_{Y.A}^2$ is always larger than $\hat{\omega}_{A.Y}^2$ and $\hat{\rho}_{A.Y}^2$ (and this may add to its popularity). In any case, when the number of degrees of freedom is larger than 80 all these coefficients will give roughly the same value (see Fowler, 1985) and then the choice is only a matter of taste.

10.5.6 Interpreting the size of an effect

Strictly speaking, the size of an effect is meaningful only for a random effect model (cf. Glass and Hakstian, 1969; Dooling and Danks, 1975; Keren and Lewis, 1979; Maxwell *et al.* 1981) because in a fixed effect model it is always possible to manipulate the effect size via a control of the error factors. So in a fixed effect model the size of the effect is often an indication of the *quality* of the experiment rather than of the size of the effect *per se*.

10.6 Three exercises

The following three experiments provide some key points concerning the analysis of an $\mathcal{S}(\mathcal{A})$ experimental design. The first experiment has one fixed factor, and can be analyzed using a straightforward ANOVA. The second experiment shows how an ANOVA can be calculated when the data supplied include only means and standard deviations. The final experiment presents the experimental dilemma of deciding whether a factor should be considered as fixed or random.

10.6.1 Images ...

In a new experiment on mental imagery, we have three groups of 5 students each (psychology majors for a change!) learn a list of 40 concrete nouns and recall them 1 hour later. The first group learns each word with its definition, and draws the object denoted by the word (the *built image condition*). The second group was treated just like the first, but had simply to *copy* a drawing of the object instead of making it up themselves (the *given image condition*). The third group simply read the words and their definitions (the *control condition*). Table 10.2 shows the number of words recalled 1 hour later by each subject.

Do these results allow us to conclude that there is an effect of the different instructions on memorization? We need to run an ANOVA. The experimental design is clearly $\mathcal{S}(\mathcal{A})$, with $S = 5$, $A = 3$, and $\mathcal{A}$ is a fixed factor (that is, the only levels of interest are present in this experiment). From this we obtain the ANOVA results displayed in Table 10.2 (the detail of the computations is left to the reader as an exercise).

Thus, we can conclude that instructions had an effect on memorization.

	Experimental condition		
	Built image	Given image	Control
	22	13	8
	17	9	7
	24	14	10
	23	18	14
	24	21	16
$\sum$	110	75	55
$M_{a\cdot}$	22	15	11

Table 10.2 Results of the mental imagery experiment.

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Pr(F)</i>
$\mathcal{A}$	2	310.00	155.00	10.33**	.0026
$S(\mathcal{A})$	12	180.00	15.00		
Total	14	490.00			

Table 10.3 ANOVA table for the mental imagery experiment

10.6.2 The fat man and not so very nice numbers ...

Stein and Bransford (1979) wanted to show the effect of elaboration during encoding on memory. To do this, they designed the following experiment. Forty subjects were assigned (randomly) to four groups. Each subject in each group read twelve short sentences like:

- The fat man looked at the sign.
- The tall man took the jar.

A *control* group read only these short sentences. The three other groups read the sentences augmented with phrases.

The *precise elaboration* group read the sentences completed with a phrase that provided an elaboration agreeing with the adjective modifying ‘man’, for example:

- The fat man looked at the sign indicating that the chair was fragile.
- The tall man took the jar from a high shelf.

The *imprecise elaboration* group read the sentences completed with a phrase that *didn’t* provide an elaboration agreeing with the adjective modifying ‘man’, for example:

- The fat man looked at the sign which was 6 feet high.
- The tall man took the jar of jam.

Finally, subjects in the *personal elaboration* group were asked to complete the sentences with phrases of their own creation.

After a ten minute delay the subjects in each group were asked to recall all the adjectives modifying ‘man’ that they had read. The means and standard deviations obtained in each group are presented in Table 10.4.

Because only means and standard deviations for each of the experimental groups are provided, the comprehension formulas must be used to compute the analysis of variance.

	Experimental condition			
	Control	Precise	Imprecise	Personal
$M_{a\cdot}$	4.2	7.4	2.2	5.8
$\hat{\sigma}_a$	1.08167	1.20830	1.15758	1.10905

Table 10.4 Results for Stein and Bransford's (1979) experiment

Recall the formula for standard deviation:

$$\hat{\sigma}_a = \sqrt{\sum (Y_{as} - M_{a\cdot})^2 / (S - 1)}$$

Calculation of the sum of squares of $\mathcal{A}$:

$$\begin{aligned} SS_A &= S \sum (M_{a\cdot} - M_{\cdot\cdot})^2 \\ &= 10[(4.20 - 4.90)^2 + (7.40 - 4.90)^2 + (2.20 - 4.90)^2 + (5.80 - 4.90)^2] \\ &= 148.40 \end{aligned}$$

Calculation of the mean square of $\mathcal{A}$:

$$MS_A = SS_A / (A - 1) = 148.40 / 3 = 49.47$$

Calculation of the mean square of $\mathcal{S}(\mathcal{A})$

$$\begin{aligned} MS_{S(A)} &= \sum (Y_{as} - M_{a\cdot})^2 \cdot / A(S - 1) = \sum \hat{\sigma}_a^2 / A \\ &= (1.08167^2 + 1.20830^2 + 1.15758^2 + 1.10905^2) / 4 = 1.30 \end{aligned}$$

Calculation of F :

$$F_{\text{cal}} = MS_A / MS_{S(A)} = 49.47 / 1.30 = 38.05.$$

The critical value of F for $\alpha = .01$ with $\nu_1 = 3$ and $\nu_2 = 36$ is 4.38, and clearly leads to rejection of the null hypothesis at the .01 level.

10.6.3 How to choose between fixed and random—taking off with Elizabeth Loftus ...

Elizabeth Loftus (Loftus and Palmer, 1974), in a series of experiments on the theme of eyewitness testimony, wanted to demonstrate the influence of the wording of a question on the later responses of witnesses. To do this, she showed subjects a film of a car accident. Following the film she asked them a series of questions. Among the questions was one of five versions of a critical question concerning the speed at which the vehicles had been traveling in miles per hour. Here are the versions of this question, listed with the experimental condition in which they occurred:

1. HIT: About how fast were the cars going when they *hit* each other?
2. SMASH: About how fast were the cars going when they *smashed into* each other?
3. COLLIDE: About how fast were the cars going when they *collided with* each other?
4. BUMP: About how fast were the cars going when they *bumped* each other?
5. CONTACT: About how fast were the cars going when they *contacted* each other?

	Experimental groups				
	Contact	Hit	Bump	Collide	Smash
	21	23	35	44	39
	20	30	35	40	44
	26	34	52	33	51
	46	51	29	45	47
	35	20	54	45	50
	13	38	32	30	45
	41	34	30	46	39
	30	44	42	34	51
	42	41	50	49	39
	26	35	21	44	55
$M_{a.}$	30	35	38	41	46

Table 10.5 Results for a fictitious replication of Loftus's experiment.

The results are presented in Table 10.5. The dependent variable is the speed in miles per hour; the independent variable is the verb used in the critical question. Loftus wanted to show that the connotations of the verb affected the judgment of the subjects concerning speed. The important point is that Loftus wanted to generalize these results to all verbs signifying 'coming in contact with'. Since she could not, strictly speaking, sample from those verbs randomly, she selected a representative sample (see Brunswick, 1956). The problem is to decide whether the factor is fixed or random. If we decide that the verbs that Loftus selected are a *representative sample*, we can decide in favor of a random factor (see the discussion started by Clark, 1973). If we decide that the levels were in fact chosen *arbitrarily*, then we can call the factor fixed and the conclusions we draw can only be extended to the levels of the factor actually present in the design. Whatever we decide, the decision is open to criticism.

Here the distinction between fixed and random factors can appear to be without importance since the decision to reject the null hypothesis or not will be identical in the two cases. *This will not be the case for more complex designs*. Clark's (1973) central argument was that some psycholinguistic research led to erroneous conclusions as a result of confusion between fixed and random factors (cf. Wike and Church, 1976), specifically by taking random factors as fixed. Chastaing (1986) responded to this discussion by pointing out that another body of psycholinguistic research was invalid because researchers had taken fixed factors as random!

In the present case, the choice between the two models has no effect on the conclusions drawn from the statistical analysis (though it might affect the *psychological* interpretation). Analysis of variance lets us conclude that in any case there was an effect of verb on estimated speed. The ANOVA table follows:

Source	df	SS	MS	F	Pr(F)
$\mathcal{A}$	4	1,460.00	365.00	4.56	.0036
$S(\mathcal{A})$	45	3,600.00	80.00		
Total	49	5,060.00			



Figure 10.1 Results from replication of Loftus and Palmer’s (1974)s experiment.

Thus the verb used in the critical question influenced the estimation of speed that the subjects made, $F(4, 45) = 4.56$, $MS_e = 80.00$, $p < .01$. See Figure 10.1 in which we see that the verb ‘smash’ led to higher estimates.

Chapter summary

10.7 Key notions of the chapter

Below are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

Score model.	Difference in score models when $\mathcal{A}$ is fixed vs random.
Decomposition of a score (a Y value).	
Expected values of mean squares for $\mathcal{A}$ and $\mathcal{S}(\mathcal{A})$.	Estimation of the intensity of effect of the independent variable in completing the ANOVA
Expressing sums of squares in terms of the score model.	Importance of the <i>selection</i> of the random levels of a random factor.

10.8 New notations

Below are the new notations introduced in this chapter. Test yourself on their meaning.

$$Y_{as} = \mu_{..} + \alpha_a + e_{s(a)}$$

$$\alpha_a = \mu_{a.} - \mu_{..}$$

$$\bar{e}_{(\cdot)} = \frac{e_{(\cdot)}}{AS} = \frac{\sum e_{s(a)}}{AS}$$

$$\bar{e}_{(a)} = \frac{1}{S} \sum_s e_{s(a)}$$

$$\mathbf{E}\{MS_A\} = \sigma_e^2 + S\vartheta_a^2$$

$$\mathbf{E}\{MS_{S(A)}\} = \sigma_e^2$$

with

$$\vartheta_a^2 = \frac{\sum \alpha_a^2}{(A-1)}, \quad (\text{read 'theta squared a'})$$

and

σ_e^2 : error variance,

(read 'sigma squared e')

$$\omega_{\mathcal{A},Y}^2 (\text{effect size for a fixed effect model}) \quad (10.29)$$

and

$$\rho_{\mathcal{A},Y}^2 (\text{effect size for a random effect model}) \quad (10.30)$$

10.9 Key formulas of the chapter

Below are the main formulas introduced in this chapter: try to go through them and understand what they mean.

$$Y_{as} = M_{..} + (M_{a\cdot} - M_{..}) + (Y_{as} - M_{a\cdot})$$

Y_{as} = mean value of the DV when ignoring the IV + effect of the ath level of the IV + error linked to subject s nested in the ath level of the IV

$$Y_{as} = \mu_{..} + (\mu_{a\cdot} - \mu_{..}) + (Y_{as} - \mu_{a\cdot})$$

Fixed effect model:

$$Y_{as} = \mu_{..} + \alpha_a + e_{s(a)} \quad \text{expanded as} \quad \alpha_a = \mu_{a\cdot} - \mu_{..} \quad \text{with} \quad \sum_a \alpha_a = 0.$$

$$\mathbf{E}\{MS_A\} = \sigma_e^2 + S\vartheta_a^2$$

$$\mathbf{E}\{MS_{S(A)}\} = \sigma_e^2$$

with

$$\vartheta_a^2 = \frac{\sum \alpha_a^2}{(A-1)} \quad \text{and} \quad \sigma_e^2 : \text{error variance}$$

Random effect model:

$$Y_{as} = \mu_{..} + a_a + e_{s(a)}$$

$$\mathbf{E}\{MS_A\} = \sigma_e^2 + S\sigma_a^2$$

$$\mathbf{E}\{MS_{S(A)}\} = \sigma_e^2$$

$$\omega_{\mathcal{A},Y}^2 = \frac{\sigma_a^2}{\sigma_e^2 + \sigma_a^2}.$$

$$\rho_{\mathcal{A},Y}^2 = \frac{\sigma_a^2}{\sigma_e^2 + \sigma_a^2}.$$

$$\widehat{\omega}_{\mathcal{A},Y}^2 = \frac{SS_A - (A-1)MS_{S(A)}}{SS_{\text{total}} + MS_{S(A)}}.$$

$$\widehat{\rho}_{\mathcal{A},Y}^2 = \frac{MS_A - MS_{S(A)}}{MS_{S(A)} + (S-1)MS_{S(A)}}.$$

10.10 Key questions of the chapter

Below are some questions about the content of this chapter. All the answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- * What are the four main reasons for studying the score model?
- * How do you estimate the a_a and $e_{s(a)}$ components of the score model?
- * In the score model, when do you use a Greek letter to indicate the source of an effect, and when a Roman letter?
- * Under what conditions do MS_A and $MS_{S(A)}$ have the same expected values?
- * Why do we use ϑ_a^2 instead of σ_a^2 in writing the expected value of $\mathbf{E}\{MS_A\}$ when $\mathcal{A}$ is fixed?
- * When do we observe $F < 1$? How do we test the 'significance' of $F < 1$?
- * When is the choice between $R_{Y,\mathcal{A}}^2$, $\widehat{\omega}_{\mathcal{A},Y}^2$, and $\widehat{\rho}_{\mathcal{A},Y}^2$ a matter of taste?
- * Why is the estimation of the effect size meaningful only for a random effect model?

11

Assumptions of analysis of variance

11.1 Introduction

In order to evaluate the probability associated with the F criterion, we use its sampling distribution (i.e. it is the family of Fisher's F distributions). In order to compute this sampling distribution we assume that the null hypothesis is true, but we also need to make additional technical assumptions, namely that the score model holds, that the error has zero mean, is independent of the measurements, and is distributed like a normal distribution. In this chapter we examine these assumptions in more detail and we look at possible ways of evaluating the validity of these assumptions.

In the first part of the chapter, we look at the assumptions behind the analysis of variance, then we review two procedures which test the veracity of these assumptions: first we show how to test if the variance of error is the same for all groups. This is called testing the homogeneity of variance. Second we show how to test if the error follows a normal distribution. Finally, we present some transformations of scores that can be used to analyze data when the assumptions are not met.

11.2 Validity assumptions

First of all, a truism: analysis of variance, in spite of its name, is a technique for comparing means. As a result, the operations of addition, division, etc., on the dependent variable must be possible and 'relevant' in terms of the theory of measurement we are using (cf. Krantz *et al.*, 1971; Roberts, 1979). More specifically, those operations need to have some meaning in the context of the measurements that were made. Technically speaking, the dependent variable must be measured at least at the level of an *interval scale*. If we have doubts that the dependent variable was measured on an interval scale we can use other methods of data analysis, such as the 'non-parametric' or 'ordinal scale' methods, e.g. the Friedman and Kruskal–Wallis tests on rank order (Siegel, 1956; Conover, 1971; Leach, 1979). Nevertheless, we can be reassured that ANOVA is remarkably *robust*, and leads to valid conclusions even with ordinal or binary data (for example, rank orders or data consisting of zeros and ones, such as 'yes' and 'no' answers).

In order to derive the sampling distribution of the criterion F , obtained as the ratio $MS_A/MS_{S(A)}$, statisticians need to assume that H_0 is true, along with several ‘validity assumptions’:

- Condition 1. The dependent variable comes from a population that is normally distributed in each group, which in conjunction with the score model implies that the error terms (the $e_{s(a)}$) are also normally distributed.
- Condition 2. The errors have a mean of zero. (That is, the errors cancel out in the long run.) Moreover, the intensity of the error term must be independent of the levels of the independent variable. In other words, the independent variable does not influence the experimental error—the error variance is *homoscedastic*, and this condition is called *homoscedasticity* (quite a tongue twister...).
- Condition 3. The various observations are independent of each other. This should hold not only within each experimental group, but also between groups (which rules out, in particular, the repeated use of the same subjects in each group¹).

We can summarize these conditions by stating that the error is $\mathcal{N}(0, \sigma_e^2)$. This is read: ‘The error follows a normal distribution with a mean of zero and a variance of sigma squared e ’.

Using statistical tests we can verify that conditions 1 and 2 are satisfied. For condition 1 we use a ‘test of normality’, of which the most effective is that of Lilliefors (1967). There are many tests for checking homoscedasticity, here we present the O’Brien test for homogeneity of variance (for a review of the various tests of homogeneity, see Martin and Games, 1977; Games *et al.*, 1979; O’Brien, 1979).

However, thanks to a variety of studies using computer simulations involving ‘Monte-Carlo’ methods (Linguist, 1953; Box, 1954; Games and Lucas, 1966; Scheffé, 1959; Bevan, *et al.*, 1974, among others), we can show that conditions 1 and 2 may be violated without invalidating the results of the analysis. This is true especially when the design of the experiment is balanced (that is, when there are the same number of subjects in each group). In spite of these results, one is sometimes advised to *transform* the scores on the dependent variable to make them conform to the desired scale of measurement. Condition 3 is usually satisfied by using random assignment of subjects to groups. We should keep in mind that each subject is going to contribute *only one score* to the set of data; that is, measurements on a single subject are not repeated. Even so, we should still be aware of different conditions which may militate against independence. (For example, the subjects might copy one another’s answers, or communicate with each other, or be swayed by other subtle effects due to the group or to the experimenter.)

In addition, random factor designs seem to be more susceptible to violations of the validity assumptions than designs with fixed factors. We need to be more cautious, therefore, in interpreting them when we suspect that those conditions are not satisfied. Nevertheless, the really crucial point about random factors lies in the way the levels of the independent variable are selected. *It is essential that the selection be random.*

¹ As we will see in Chapter 18, we can use analysis of variance with repeated measurements, but the procedure and the assumptions are somewhat different.

11.3 Testing the homogeneity of variance assumption

The homogeneity of variance assumption is one of the critical assumptions underlying most parametric statistical procedures including analysis of variance and it is important to be able to test this assumption. In addition, showing that several samples do not come from populations with the same variance is sometimes of importance *per se*. Among the many procedures used to test this assumption, one of the most sensitive is that by O'Brien (1979, 1981). The null hypothesis for this test is that the samples under consideration come from populations with the same variance. The alternative hypothesis is that the populations have different variances.

Compared to other tests of homogeneity of variance, the advantage of the O'Brien test resides in its versatility and its compatibility with standard analysis of variance designs (cf. Martin and Games, 1977; Games *et al.*, 1979; O'Brien, 1979). It is also optimal because it minimizes both Type I and II errors (see Chapter 3 for these notions). The essential idea behind the O'Brien test is to replace, for each sample, the original scores by transformed scores such that the transformed scores reflect the variance of the sample. Then, a standard analysis of variance based on the transformed scores will test the homogeneity of variance assumption.

11.3.1 Motivation and method

There are several tests available to detect if several samples come from populations having the same variances. In the case of two samples, the ratio of the population estimates (computed from the samples) is distributed as a Fisher distribution under the usual assumptions. Unfortunately there is no straightforward extension to this approach for designs involving more than two samples. By contrast, the O'Brien test is designed to test the homogeneity of variance assumption for several samples at once and with the versatility for analysis of variance designs including contrast analysis and sub-design analysis.

The main idea behind the O'Brien test is to transform the original scores so that the transformed scores reflect the variation of the original scores. An analysis of variance on the transformed scores will then reveal differences in the variability (i.e. variance) of the original scores and therefore this analysis will test the homogeneity of variance assumption. A straightforward application of this idea will be to replace the original scores by the absolute value of their deviation to the mean of their experimental group (cf. Levene, 1960; Glass and Stanley, 1970). So, if we denote by Y_{as} the score of subject s in experimental condition a whose mean is denoted by $M_{a\cdot}$, this first idea amounts to transforming Y_{as} into v_{as} as:

$$v_{as} = |Y_{as} - M_{a\cdot}|.$$

This transformation has the advantage of being simple and easy to understand, but, unfortunately, it creates some statistical problems (i.e. the F distribution does not model the probability distribution under the null hypothesis) and in particular, it leads to an excess of Type I errors (i.e. we reject the null hypothesis more often than the α level indicates; for more details see, e.g., Miller, 1968; Games *et al.* 1979; O'Brien, 1979).

A better approach is to replace each score by its absolute distance to the *median* of its group. Specifically, each score is replaced by

$$w_{as} = |Y_{as} - Md_{a\cdot}|$$

with $Md_{a\cdot}$: median of group a . This transform gives very satisfactory results for an omnibus testing of the homogeneity of variance assumption. However, in order to implement

more sophisticated statistical procedures (e.g. contrast analyses, multiple comparisons, see Chapter 12), a better transformation has been proposed by O'Brien (1979, 1981). Here, the scores are transformed as:

$$u_{as} = \frac{S_a(S_a - 1.5)(Y_{as} - M_{a\cdot})^2 - .5SS_a}{(S_a - 1)(S_a - 2)},$$

with:

S_a number of observations of group a

$M_{a\cdot}$ mean of group a

SS_a sum of the squares of group a : $SS_a = \sum_s (Y_{as} - M_{a\cdot})^2$.

When all the experimental groups have the same size, this formula can be simplified as

$$u_{as} = \frac{S(S - 1.5)(Y_{as} - M_{a\cdot})^2 - .5SS_a}{(S - 1)(S - 2)}$$

where S is the number of observations per group.

11.4 Example

In this section we detail the computation of the median and the O'Brien transforms. We use data from a memory experiment reported by Hunter (1964, see also Abdi, 1987).

11.4.1 One is a bun ...

In this experiment, Hunter wanted to demonstrate that it is easier to remember an arbitrary list of words when we use a mnemonic device such as the peg-word technique. In this experiment, 64 participants were assigned to either the control or the experimental group. The task for all participants was to learn an arbitrary list of pairs of words such as 'one-sugar', 'two-tiger', ... 'ten-butterfly'. Ten minutes after they had learned their list, the participants were asked to recall as many pairs as they could. Participants in the control group were told to try to remember the words as best as they could. Participants from the experimental group were given the following instructions:

A good way to remember a list is to first learn a 'nursery-rhyme' such as: '*one is a bun, two is a shoe, three is a tree, four is door, five is a hive, six is a stick, seven is heaven, eight is a gate, nine is a mine, and ten is a hen*'. When you need to learn a pair of words, start by making a mental image of the number and then make a mental image of the second word and try to link these two images. For example, in order to learn 'one-cigarette' imagine a cartoon-like bun smoking a cigarette.

The results are given in Table 11.1 and are illustrated in Figure 11.1. The results suggest that the participants from the experimental group did better than the participants from the control group. To confirm this interpretation, an analysis of variance was performed (see Table 11.2) and the F test indicates that, indeed, the average number of words recalled is significantly larger in the experimental group than in the control group.

Figure 11.1 also shows that a large proportion of the participants of the experimental group obtained a perfect score of 10 out of 10 words (cf. the peak at 10 for this group). This is called a *ceiling* effect: some of the participants of the experimental group could have

Number of words recalled	Control group	Experimental group
5	5	0
6	11	1
7	9	2
8	3	4
9	2	9
10	2	16
$Y_{a\cdot}$	216	293
$M_{a\cdot}$	6.750	9.156
$Md_{a\cdot}$	6.500	9.500
SS_a	58.000	36.219

Table 11.1 Data from Hunter (1964). Frequency of subjects recalling a given number of words. For example, 11 subjects in the control group recalled 6 words from the list they had learned.

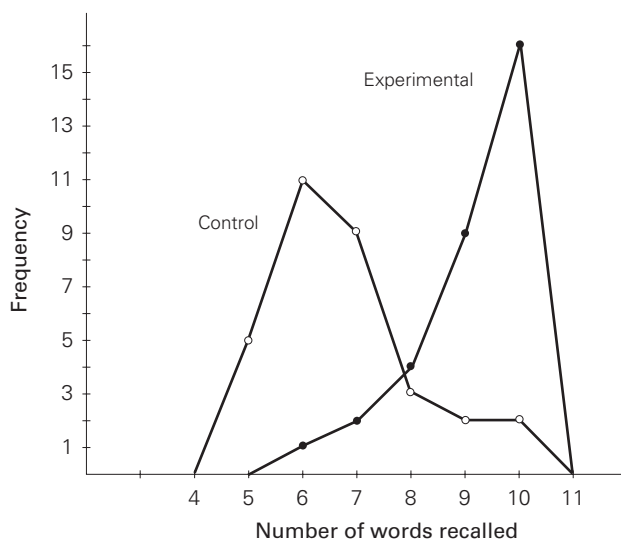


Figure 11.1 Results of the 'peg-word' experiment (one is a bun) from Hunter (1964).

Source	df	SS	MS	F	$Pr(F)$
Experimental	1	92.64	92.64	60.96**	.000 000 001
Error	62	94.22	1.52		
Total	63	186.86			

** : p smaller than $\alpha = .01$. $R^2_{A \cdot Y} = .496$.

Table 11.2 Analysis of variance for the experiment of Hunter (1964). Raw data.

performed even better if they had had more words to learn. As a consequence of this ceiling effect, the variance of the experimental group is likely to be *smaller* than it should be because the ceiling effect eliminates the differences between the participants with a perfect score. In order to decide if this ceiling effect does reduce the size of the variance of the experimental group, we need to compare the variance of the two groups as shown below.

The first step to test the homogeneity of variance is to transform the original scores. For example, the transformation of Y_{as} into w_{as} for a score of 5 from the control group gives

$$w_{as} = |Y_{as} - Md_a| = 5 - 6.5 = 1.5 .$$

The transformation of Y_{as} into u_{as} for a score of 5 from the control group gives

$$\begin{aligned} u_{as} &= \frac{S(S - 1.5)(Y_{as} - M_a)^2 - .5SS_a}{(S - 1)(S - 2)} \\ &= \frac{32(32 - 1.5)(5 - 6.75)^2 - .5 \times 58}{31 \times 30} \\ &= 3.1828 . \end{aligned}$$

The recoded scores are given in Tables 11.3 and 11.4. The analysis of variance table obtained from the analysis of transformation w_{as} is given in Table 11.5. The analysis of variance table obtained from the analysis of transformation u_{as} is given in Table 11.6.

Tables 11.5 and 11.6 indicate that the ceiling effect observed in Figure 11.1 cannot be shown to significantly reduce the variance of the experimental group compared to the control group.

Number of words recalled	Control group		
	Frequency	w_{as}	u_{as}
5	5	1.5	3.1828
6	11	0.5	0.5591
7	9	0.5	0.0344
8	3	1.5	1.6086
9	2	2.5	5.2817
10	2	3.5	11.0538

Table 11.3 Recoded scores: control group.

Number of words recalled	Experimental group		
	Frequency	w_{as}	u_{as}
5	0	4.5	—
6	1	3.5	10.4352
7	2	2.5	4.8599
8	4	1.5	1.3836
9	9	0.5	0.0061
10	16	0.5	0.7277

Table 11.4 Recoded scores: experimental group.

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Pr(F)</i>
Experimental	1	0.77	0.77	1.16 (<i>ns</i>)	.2857
Error	62	41.09	0.66		
Total	63	41.86			

ns: no significant difference.

Table 11.5 Homogeneity of variance test. Recoded scores: w_{as} . (Median.)

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Pr(F)</i>
Experimental	1	7.90	7.90	1.29 (<i>ns</i>)	.2595
Error	62	378.59	6.11		
Total	63	386.49			

ns: no significant difference.

Table 11.6 Homogeneity of variance test. Recoded scores: u_{as} . (O'Brien test.)

11.5 Testing normality: Lilliefors

In order to show that the F criterion follows a Fisher distribution, we need to assume that the error has a mean of zero and that its distribution is normal. This is called the *normality* assumption. This is a major assumption of most standard statistical procedures, and it is important to be able to assess it. In addition, showing that a sample does not come from a normally distributed population is sometimes of importance *per se*. Among the many procedures used to test this assumption, one of the most well known is a modification of the Kolomogorov–Smirnov test of goodness of fit, generally referred to as the *Lilliefors test for normality*, which was developed independently by Lilliefors (1967) and by Van Soest (1967). The null hypothesis is that the error is normally distributed (i.e. there is no difference between the observed distribution of the error and a normal distribution). The alternative hypothesis is that the error is not normally distributed.

Like most statistical tests, this test of normality defines a criterion and gives its sampling distribution. When the probability associated with the criterion is smaller than a given α -level, the alternative hypothesis is accepted (i.e. we conclude that the sample does *not* come from a normal distribution). An interesting peculiarity of the Lilliefors test is the technique used to derive the sampling distribution of the criterion. In general, mathematical statisticians derive the sampling distribution of the criterion using analytical techniques. However, in this case, this approach failed and consequently Lilliefors decided to calculate an approximation of the sampling distribution by using the Monte-Carlo technique. Essentially, the procedure consists of extracting a large number of samples from a normal population and computing the value of the criterion for each of these samples. The empirical distribution of the values of the criterion gives an approximation of the sampling distribution of the criterion under the null hypothesis.

Specifically, both Lilliefors and Van Soest used, for each sample size chosen, 1,000 random samples derived from a standardized normal distribution to approximate the sampling

distribution of a Kolomogorov–Smirnov criterion of goodness of fit. The critical values given by Lilliefors and Van Soest are quite similar, the relative error being of the order of 10^{-2} .

According to Lilliefors (1967) this test of normality is more powerful than other procedures for a wide range of non-normal conditions. Dagnelie (1968) indicated that the critical values reported by Lilliefors can be approximated by an analytical formula. Such a formula facilitates writing computer routines because it eliminates the risk of creating errors when keying in the values of the table. Recently, Molin and Abdi (1998) refined the approximation given by Dagnelie and computed new tables using a larger number of runs (i.e. $K = 100,000$) in their simulations.

11.6 Notation

The sample for the test is made of N scores, each of them denoted X_n . The sample mean is denoted M_X and computed as

$$M = \frac{1}{N} \sum_n^N X_n, \quad (11.1)$$

the sample variance is denoted

$$\hat{\sigma}^2 = \frac{\sum_n^N (X_n - M_X)^2}{N - 1}, \quad (11.2)$$

and the standard deviation of the sample denoted $\hat{\sigma}$ is equal to the square root of the sample variance.

The first step of the test is to transform each of the X_n scores into Z -scores as follows:

$$Z_n = \frac{X_n - M_X}{\hat{\sigma}}. \quad (11.3)$$

For each Z_n -score we compute the proportion of score smaller or equal to its value: This is called the *frequency associated* with this score and it is denoted $\mathcal{S}(Z_n)$. For each Z_n -score we also compute the probability associated with this score if it comes from a ‘standard’ normal distribution with a mean of zero and a standard deviation of one. We denote this probability by $\mathcal{N}(Z_n)$:

$$\mathcal{N}(Z_n) = \int_{-\infty}^{Z_n} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} Z_n^2 \right\}. \quad (11.4)$$

This value can be obtained from some scientific calculators, spreadsheets, or from Table 1 on page 497 in the Appendix.

The criterion for the Lilliefors test is denoted L . It is calculated from the Z -scores:

$$L = \max_n \{ |\mathcal{S}(Z_n) - \mathcal{N}(Z_n)|, |\mathcal{S}(Z_n) - \mathcal{N}(Z_{n-1})| \}. \quad (11.5)$$

So L is the absolute value of the biggest split between the probability associated with Z_n when Z_n is normally distributed, and the frequencies actually observed. The term $|\mathcal{S}(Z_n) - \mathcal{N}(Z_{n-1})|$ is needed to take into account that, because the empirical distribution is discrete, the maximum absolute difference can occur at either endpoint of the empirical distribution.

The critical values are listed in Table 4 in the Appendix. The critical value is denoted L_{critical} . The null hypothesis is rejected when the L criterion is greater than or equal to the critical value L_{critical} .

	G. 1	G. 2	G. 3	G. 4
	3	5	2	5
	3	9	4	4
	2	8	5	3
	4	4	4	5
	3	9	1	4
$Y_{a\cdot}$	15	35	16	21
$M_{a\cdot}$	3	7	3.2	4.2

Table 11.7 The data from 'Romeo and Juliet'. Twenty subjects were assigned to four groups. The dependent variable is the number of ideas correctly recalled.

11.7 Numerical example

We now look at an analysis of variance example for which we want to test the 'normality assumption'. We will use the data from our 'Romeo and Juliet' example (see Section 8.7, page 157). Recall that the data correspond to memory scores obtained by 20 subjects who were assigned to one of four experimental groups (hence five subjects per group). The within-group mean square $MS_{S(A)}$ (equal to 2.35) corresponds to the best estimation of the population error variance. For convenience, the data are reprinted in Table 11.7

The normality assumption states that the error is normally distributed. In the analysis of variance framework, the error corresponds to the *residuals* which are equal to the deviations of the scores from the mean of their group. So in order to test the normality assumption for the analysis of variance, the first step is to compute the residuals from the scores. We denote X_n the residual corresponding to the i th observation (with i ranging from 1 to 20). The residuals are given in the following table:

Y_{as}	3	3	2	4	3	5	9	8	4	9
X_n	0	0	-1	1	0	-2	2	1	-3	2
Y_{as}	2	4	5	4	1	5	4	3	5	4
X_n	-1.2	.8	1.8	.8	-2.2	.8	-.2	-1.2	.8	-.2

Next we transform the X_n values into Z_n values using the following formula:

$$Z_n = \frac{X_n}{\sqrt{MS_{S(A)}}} \quad (11.6)$$

because $MS_{S(A)}$ is the best estimate of the population variance, and the mean of X_n is zero. Then, for each Z_n value, the frequency associated with $\mathcal{S}(Z_n)$ and the probability associated with Z_n under the normality condition $\mathcal{N}(Z_n)$ are computed [we have used Table 1 of the normal distribution in the Appendix on page 497 to obtain $\mathcal{N}(Z_n)$]. The results are presented in Table 11.8.

The value of the criterion is (see Table 11.8)

$$L = \max_n \{ |\mathcal{S}(Z_n) - \mathcal{N}(Z_n)|, |\mathcal{S}(Z_n) - \mathcal{N}(Z_{n-1})| \} = .250. \quad (11.7)$$

X_n	N_i	F_i	Z_n	SZ_n	$\mathcal{N}(Z_n)$	D_0	D_{-1}	max
-3.0	1	1	-1.96	.05	.025	.025	.050	.050
-2.2	1	2	-1.44	.10	.075	.025	.075	.075
-2.0	1	3	-1.30	.15	.097	.053	.074	.074
-1.2	2	5	-.78	.25	.218	.032	.154	.154
-1.0	1	6	-.65	.30	.258	.052	.083	.083
-.2	2	8	-.13	.40	.449	.049	.143	.143
.0	3	11	.00	.55	.500	.050	.102	.102
.8	4	15	.52	.75	.699	.051	.250	.250
1.0	2	17	.65	.85	.742	.108	.151	.151
1.8	1	18	1.17	.90	.879	.021	.157	.157
2.0	2	20	1.30	1.00	.903	.097	.120	.120

Table 11.8 How to compute the criterion for the Lilliefors test of normality. N_n stands for the absolute frequency of a given value of X_n , F_n stands for the absolute frequency associated with a given value of X_n (i.e. the number of scores smaller or equal to X_n), Z_n is the Z-score corresponding to X_n , $S(Z_n)$ is the proportion of scores smaller than Z_n , $\mathcal{N}(Z_n)$ is the probability associated with Z_n for the standard normal distribution, $D_0 = |S(Z_n) - \mathcal{N}(Z_n)|$, $D_{-1} = |S(Z_n) - \mathcal{N}(Z_{n-1})|$, and max is the maximum of $\{D_0, D_{-1}\}$. The value of the criterion is $L = .250$.

Taking an α level of $\alpha = .05$, with $N = 20$, we find from Table 4 in the Appendix on page 505 that $L_{\text{critical}} = .192$. Because L is larger than L_{critical} , the null hypothesis is rejected and we conclude that the residuals in our experiment are not distributed normally.

11.8 Numerical approximation

The available tables for the Lilliefors' test of normality typically report the critical values for a small set of α values. For example, Table 4 in the Appendix reports the critical values for $\alpha = [.20, .15, .10, .05, .01]$. These values correspond to the 'habitual' α levels. To find other α values, a first approach is to generate the sampling distribution 'on the fly' for each specific problem (i.e. to run a new simulation every time we make a test) and to derive the specific critical values which are needed. Another approach to finding critical values for unusual values of α , is to using a numerical approximation for the sampling distributions. Molin and Abdi (1998, see also Abdi and Molin, 2007) proposed such an approximation and showed that it was accurate for at least the first two significant digits.

11.9 Transforming scores

When the assumptions behind the analysis of variance are not met, we can transform the scores in order to make them satisfy the assumptions and then run an analysis of variance on these transformed scores. Specifically, we can transform the scores to take care of the following problems:

- homogeneity of variance
- normality of the error distribution
- additivity of the effects.

We have already mentioned that the analysis of variance was robust relative to the first two conditions, and that therefore these are not as critical as the last one (cf. Budescu and Appelbaum, 1981; but see also Wilcoxon, 1987; Box and Cox, 1964). It is worth noting that often a transformation tailored for one condition is likely to work for the other ones.

One word of caution about transformation of scores: the results of the analysis are valid only for the *transformed scores*, not for the original scores.

11.9.1 Ranks

When there are outliers in the data, or if we doubt that the measurements are very precise, we can replace the scores by their ranks in the distribution. For example, for two groups having, each, the following three observations: group 1: 1, 100, 102; group 2: 9, 101, 1000; we obtain the following ranks: group 1: 1, 3, 5; group 2: 2, 4, 6. The rank transformation is equivalent to making a *non-parametric* analysis of variance—it is, in practice, a very robust transform (cf. Conover, 1971; but see also Thompson and Ammann, 1990).

11.9.2 The log transform

The logarithm transform (or ‘log’ for short, often denoted by $\ln$) is used when the group variances are proportional to their means, or when we think that the experimental effects are multiplicative (instead of additive as assumed by the score model). The log transform replaces Y_{as} by Y'_{as} as:

$$Y'_{as} = \ln(Y_{as}) . \quad (11.8)$$

If the scores are all positive with some of them close to zero, a better transform is obtained as:

$$Y'_{as} = \ln(Y_{as} + 1).$$

11.9.3 Arcsine transform

The arcsine transform is used when the data are binary (i.e. 0 or 1). In this case, the sampling distribution is derived from the binomial distribution (see Appendix D, pages 453 ff.) for which the variance is proportional to the mean (i.e. the homogeneity of variance assumption is not met). Traditionally, the arcsine transform is used for binary data, it is obtained as:

$$Y'_{as} = \arcsin\left(\sqrt{Y_{as}}\right) = \sin^{-1}\left(\sqrt{Y_{as}}\right) . \quad (11.9)$$

The arcsine transform is given by most scientific calculators or by spreadsheets. It is worth noting that even though this transform is often recommended, it may be superfluous as suggested by numerical simulations performed by Hsu and Feld (1969), and Lunney (1970). These authors showed that the sampling distribution of binary data are very close to the F distribution as long as the design is balanced and that the number of degrees of freedom of the mean square of error is ≥ 20 , and that the group means are between .2 and .8. The advantage of not transforming the data is that results are easier to interpret (it is not easy to make a concrete image of the arcsine of a percentage). Another, more modern, approach when dealing with binary data is to use logistic regression (cf. Pampel, 2000; Menard, 2001).

Chapter summary

11.10 Key notions of the chapter

Below are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

Estimation of the intensity of effect of the independent variable in completing the ANOVA

Conditions for the validity of the ANOVA (normality and homoscedasticity).

Robustness of ANOVA with respect to violations of the validity assumptions.

Testing the assumption of equal variance of error

Testing the normality assumption

Transforming scores (ranks, logarithms, arcsine transform)

11.11 New notations

Below are the new notations introduced in this chapter. Test yourself on their meaning.

O'Brien transformation: w_{as} , u_{as}

Lilliefors' test: $S(Z_n)$, $\mathcal{N}(Z_n)$, L , L_{critical}

11.12 Key formulas of the chapter

Below are the main formulas introduced in this chapter: try to go through them and understand what they mean.

$$w_{as} = |Y_{as} - Md_{a\cdot}|.$$

The transformation of Y_{as} into u_{as} for a score of 5 from the control group gives

$$u_{as} = \frac{S(S-1.5)(Y_{as} - M_{a\cdot})^2 - .5SS_a}{(S-1)(S-2)}.$$

$$L = \max_i \{ |S(Z_n) - \mathcal{N}(Z_n)|, |S(Z_n) - \mathcal{N}(Z_{n-1})| \}.$$

11.13 Key questions of the chapter

Below are some questions about the content of this chapter. The answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- * What are the 'validity assumptions' of the sampling distribution for F ?
- * Why and how do we test the homogeneity of variance assumption?
- * Why and how do we test the normality assumption?
- * Why and how do we transform the original scores?

12

Analysis of variance, one factor: planned orthogonal comparisons

12.1 Introduction

The statistical test using the F -test in the analysis of variance gives an indication of the existence of a *global effect* of the independent variable on the dependent variable. Because of that, the F ratio is often called an *omnibus* test (from the latin *omnibus* meaning 'for all'). The problem is that, in general, experimenters want to draw a more precise conclusion than a global statement such as: 'the experimental manipulation has an effect on the subjects' behavior'. For example, if we have one control group and several experimental groups, one purpose of the experiment is to compare each of the experimental groups with the control. We are probably willing, also, to compare some experimental groups to some others, etc. In order to make a fine-grained analysis of the effect of the independent variable, the strategy is to use specific or focused *comparisons*. An almost synonymous term is '*contrast analysis*'. As we shall see later, a contrast is always a comparison but not all comparisons are contrasts (but for the time being, this distinction is of no importance). Contrasts and comparisons have other functions beside decomposing the effect of the independent variable. First, they give a comprehensive view of the different experimental designs. Second, they are a great intellectual tool for understanding the similarity between two major techniques of data analysis, namely analysis of variance and regression. However, some problems are created by the use of multiple comparisons. The most important problem is that the greater the number of comparisons, the greater the risk of rejecting the null hypothesis when it is actually true (i.e. making a Type I error). The general strategy adopted to take this problem into account depends upon the number of comparisons to be performed. As a consequence, some distinctions need to be made. The most important one separates what are called *planned* and *post hoc* comparisons.

- The *planned* comparisons (also called *a priori* comparisons) are selected before running the experiment. In general, they correspond to the research

hypotheses that the researcher wants to test. If the experiment has been designed to confront two or more alternative theories (e.g. with the use of *rival* hypotheses), the comparisons are derived from those theories. When the experiment is actually run, it is possible to see if the results support or eliminate one of the theories. Because these comparisons are planned they are usually *few* in number.

- The *post hoc* comparisons (also called *a posteriori*, or ‘after the fact’ comparisons, or ‘data snooping’) are decided upon after the experiment has been run. For example, if you want to compare the best group with the worst group in an experiment, you need to collect the data *before*¹ you can know which group is what. The aim of *post hoc* comparisons is to make sure that some patterns that can be seen in the results, but were not expected, are reliable.

In this chapter, we will present the planned (*a priori*) comparisons. Among these it is customary to distinguish *orthogonal* (named also *independent*) contrasts from *non-orthogonal* contrasts (two contrasts are independent or orthogonal if their correlation coefficient is zero). Only the *orthogonal* case will be presented in this chapter. A set of comparisons is composed of orthogonal comparisons if the hypotheses corresponding to each comparison are independent of each other. If you wonder, now, about the number of possible orthogonal comparisons one can perform, the answer is one less than the number of levels of the independent variable (i.e. $A - 1$). We shall discuss later the reason for that. For the time being, the interesting thing to note is that $A - 1$ is also the number of degrees of freedom of the sum of squares for A . Despite their apparent diversity, all these different types of comparisons can be performed following the same procedure:

1. Formalization of the comparison, and expression of the comparison as a set of weights for the means.
2. Computation of the F_{comp} ratio (this is the usual F ratio adapted for the case of testing a comparison).
3. Evaluation of the probability associated with F_{comp} . This final step is the only step that changes with the different types of comparisons.

12.2 What is a contrast?

Suppose we have an experiment planned to test the effects of two different presentations of a barbiturate on video game performance. The experimental design consists of four groups:

- A *control* group in which subjects simply play with the video game.
- A *placebo* group in which subjects are given a placebo.

¹ Which means that the comparison to be made can be determined only *after* the data have been collected. *Post hoc* means ‘after the fact’, and *a posteriori* means ‘after’.

- A *red-pill* group in which subjects are given the barbiturate within a bright red pill.
- A *blue-pill* group in which subjects are given the barbiturate within a pale blue pill.

Suppose, also, that the experimenters want to show first that the barbiturate has an effect (whatever the color of the pill). This amounts simply to comparing the two barbiturate groups with the placebo group. If the barbiturate does not affect video game performance, then the two barbiturate groups should be equivalent to the placebo group. Hence, under the null hypothesis the average of the barbiturate group should be equal to the placebo group. Precisely, the null hypothesis (recall that statistical hypotheses always deal with the parameters of the population, not with the statistics actually computed) is:

$$H_0 : \mu_2 = \frac{(\mu_3 + \mu_4)}{2}.$$

A more workable way of writing the equation is to eliminate the fractional notation:

$$H_0 : 2\mu_2 = (\mu_3 + \mu_4).$$

Or, equivalently,

$$H_0 : 2\mu_2 - \mu_3 - \mu_4 = 0.$$

Basically, we have contrasted two sets of means (the positive ones vs the negative ones); hence, the name of *contrast*. It is traditional to denote a contrast by the Greek letter ψ (i.e. the letter ‘psy’ like in ‘psychology,’ it is read ‘sigh’). With this notation, the previous equation can be written:

$$\psi_1 = 2\mu_2 - \mu_3 - \mu_4.$$

(ψ_1 is pronounced ‘sigh-one’; the subscript one means that this is the first contrast of a set of contrasts). The null hypothesis, then, can be written as

$$H_0 : \psi_1 = 0$$

and the alternative hypothesis as

$$H_1 : \psi_1 \neq 0.$$

Contrasts can also be written in a more general manner. For example, ψ_1 can be written as:

$$\psi_1 = (0 \times \mu_1) + (2 \times \mu_2) + (-1 \times \mu_3) + (-1 \times \mu_4).$$

Note that μ_1 , the mean of the control group, is multiplied by 0 because it is not involved in this contrast. A contrast appears here as a *weighted sum* of the population means corresponding to each of the experimental group means. The weights are usually denoted $C_{a,c}$, with a standing for the group’s index, and c for the contrast’s index (i.e. the same index used for ψ). Some authors (i.e. Rosenthal and Rosnow, 1985; Estes, 1991) prefer using λ_a instead of C_a ; in this text, however, we have decided to stick to the more general notation C_a . Here, the C values for ψ_1 are

$$C_{1,1} = 0, \quad C_{2,1} = 2, \quad C_{3,1} = -1, \quad C_{4,1} = -1.$$

As an exercise, suppose that the experimenter wants to show a *placebo* effect. This effect will be revealed by contrasting the *placebo* group and the control group. That is, by the following $C_{a,2}$ weights:

$$C_{1,2} = 1, \quad C_{2,2} = -1, \quad C_{3,2} = 0, \quad C_{4,2} = 0.$$

This second contrast is written as

$$\psi_2 = (1 \times \mu_1) + (-1 \times \mu_2) + (0 \times \mu_3) + (0 \times \mu_4).$$

The null hypothesis is

$$H_0: \psi_2 = 0.$$

The *real* value of ψ is not known (because the values of the parameters μ_a are not known). However, ψ can be estimated from the means actually computed from the sample (i.e. the M_a values). The estimated value of ψ is denoted $\hat{\psi}$ (pronounced ‘sigh-hat’). The ‘hat’ indicates that this is an *estimation* of a parameter of a population from a statistic computed on a sample. For example,

$$\begin{aligned}\hat{\psi}_1 &= 2M_{2.} - M_{3.} - M_{4.} \\ \hat{\psi}_2 &= M_{1.} - M_{2.}\end{aligned}\tag{12.1}$$

To oppose the two different forms of the barbiturate, the contrast ψ_3 is estimated by $\hat{\psi}_3$:

$$\hat{\psi}_3 = M_{3.} - M_{4.}.$$

The null hypothesis is

$$H_0: \psi_3 = 0.$$

Note: ψ without a ‘hat’ (because statistical hypotheses are stated in terms of parameters of some population, not in terms of estimates).

12.2.1 How to express a research hypothesis as a contrast

When a research hypothesis is precise, it is possible to express it as a contrast. A research hypothesis, in general, can be expressed as an *a priori* order on the experimental means. This order can be expressed as a set of ranks, each rank corresponding to one mean. Actually, in some cases, it may even be possible to predict quantitative values (e.g. the mean reaction time for the first experimental group should be 455 ms, the mean of the second group should be 512 ms, etc.). This case rarely occurs in psychology. Anyway, if it happens it is fairly easy to deal with by replacing the ranks by the precise values. To convert the set of ranks into a contrast, it suffices to subtract the mean rank from each rank, and then to convert these values into integers. The transformed ranks, then, give the C_a weights expressing the research hypothesis as a contrast.² Some examples follow to help mastering the technique.

² Another way could be to transform the rank into Z-scores, because Z-scores, having a mean of zero, are always contrasts, but they are not, in general, ‘nice’ numbers.

12.2.1.1 Example: rank order

If we have an experiment with four groups (such as ‘Romeo and Juliet’ used in Chapter 8, Section 8.7, page 157) and if we predict that the second group will be superior to the other three groups, then the following rank order should emerge:

C_1	C_2	C_3	C_4	Mean	
1	2	1	1	5/4.	(12.2)

Subtracting the mean from each rank, the weights become:

C_1	C_2	C_3	C_4	Mean	
$-1/4$	$3/4$	$-1/4$	$-1/4$	0.	(12.3)

Finally, transforming the fractions into integers gives

C_1	C_2	C_3	C_4	Mean	
-1	3	-1	-1	0 ,	(12.4)

which is the contrast we are looking for.

12.2.1.2 A bit harder

Assume that for a 4-group design, the theory predicts that the first and second groups should be equivalent. The third group should perform better than these two groups. The fourth group should do better than the third, and the advantage of the fourth group over the third should be twice the gain of the third over the first and the second. When translated into a set of ranks this prediction gives:

C_1	C_2	C_3	C_4	Mean	
1	1	2	4	2	(12.5)

which translates into the contrast:

C_1	C_2	C_3	C_4	Mean	
-1	-1	0	2	0 .	(12.6)

In case of doubt, a good heuristic is to draw the predicted configuration of results, and then to represent the position of the means by ranks.

12.3 The different meanings of alpha

Experimenters are generally interested in making several contrasts. A set of contrasts is called a *family of contrasts* or *family of comparisons*. The problem with testing multiple contrasts³ is that the meaning of α (the significance level) becomes ambiguous. The main reason may be illustrated as follows. Suppose that the general null hypothesis is true, that is, all the population’s means (μ_a) have the same value. If several comparisons computed from

³ Or, more generally, when using the same data set to test several statistical hypotheses.

the group means (i.e. the M_a values) are tested, it is possible that some of these will reach significance due to random fluctuations.

To make the point even clearer, imagine the following ‘pseudo-experiment’:

I toss 20 coins on the table, and I imagine that I am able to compel the coins to fall on the heads side (I believe that I have some parapsychological powers like telekinesis). As I am a statistician, I know that by using the ‘binomial test’ (which is described in Appendix E) I can say that I will be able to reject the null hypothesis with the $\alpha = .05$ level if the number of heads is greater than or equal to 15. That is, the probability of getting 15 or more heads due to chance alone is less than $\alpha = .05$. As I think that my parapsychological powers vary from time to time, I decide to repeat the experiment 10 times.

Suppose that this pseudo-experiment is performed. On the 10 trials (i.e. the 10 *comparisons*), one gives the results 16 heads vs 4 tails. Does that mean that I influenced the coins on that occasion? Of course not! It is clear that the more I repeat the experiment the greater the probability of detecting a low-probability event (like 16 vs 4). Actually, if I wait long enough, I will be able to observe any given event as long as its probability is not zero. Specifically, if I toss the coins 10 times, we will see later on that I have a .40 probability to find at least one result allowing the rejection of the null hypothesis (with a .05 α level) *when the null hypothesis is, in fact, true*.

12.3.1 Probability in the family

In this section we see how to compute the probability of rejecting the null hypothesis at least once in a family of comparisons when the null hypothesis is true. In other words, we will show how to compute the probability of making at least one Type I error in the family of comparisons.

For convenience suppose that we set the significance level at $\alpha = .05$. For one comparison (i.e. one trial in the example of the coins) the probability of making a Type I error is equal to $\alpha = .05$. The events ‘making a Type I error’ and ‘not making a Type I error’ cannot occur simultaneously: they are *complementary events*. That is, the probability of one event of the pair of complementary events is equal to

$$1 - \{\text{probability of the other event}\}.$$

Thus, the probability of *not making a Type I error* on one trial is equal to

$$1 - \alpha = 1 - .05 = .95.$$

It happens that when two events are independent of each other, the probability of observing these two events simultaneously is given by the product of the probability of each event (see Appendix C). Thus, if the comparisons are independent of each other (as one can assume for the coins example), the probability of not making a Type I error on the first comparison *and* not making a Type I error on the second comparison is

$$.95 \times .95 = (1 - .05)^2 = (1 - \alpha)^2.$$

With three comparisons, the same reasoning gives for the probability of not making a Type I error on all three comparisons:

$$.95 \times .95 \times .95 = (1 - .05)^3 = (1 - \alpha)^3.$$

Suppose now that there are C comparisons in the family. The probability of *not* making a Type I error for the *whole family* (i.e. not on the first, nor the second, nor the third, etc.) is

$$(1 - \alpha)^C.$$

For the coins example, the probability of not making a Type I error on the family when each of the 10 comparisons (i.e. trials) uses a .05 level is

$$(1 - \alpha)^C = (1 - .05)^{10} = .599.$$

Now, what we are looking for is the probability of making one or more Type I errors on the family of comparisons. This event is the complement of the event *not making a Type I error on the family of comparisons*. Its probability is thus

$$1 - (1 - \alpha)^C.$$

For the example we find

$$1 - (1 - .05)^{10} = .401.$$

In other words: if a significance level of .05 is selected for *each* comparison, then the probability of wrongly rejecting the null hypothesis on 10 comparisons is .401.

This example makes clear the need to distinguish between two meanings of α :

- The probability of making a Type I error when dealing only with a specific comparison. This probability is denoted $\alpha[PC]$ (pronounced ‘alpha per comparison’).
- The probability of making at least one Type I error for the whole family of comparisons. This probability is denoted $\alpha[PF]$ (pronounced ‘alpha per family of comparisons’).

12.3.2 A Monte-Carlo illustration

The use of the so-called ‘Monte-Carlo’ technique may help you grasp the difference between $\alpha[PC]$ and $\alpha[PF]$. The Monte-Carlo technique consists of running a simulated experiment many times using random data, with the aim of obtaining a pattern of results showing what would happen just on the basis of chance. We have already used this technique to show how the sampling distribution of the F -ratio can be approximated for correlation (cf. Chapter 3, Section 3.2.5.1, pages 45ff.). We are going to use it again to show what can happen when the null hypothesis is actually true and when we perform several comparisons per experiment.

Suppose that 6 groups with 100 observations per group are created with data randomly sampled from a normal population. By construction, the null hypothesis is true (i.e. all population means are equal). Call that procedure an *experiment*. Now, construct 5 *independent comparisons* from these 6 groups (we will see later on how to do that). For each comparison, compute an F -test. If the probability associated with the statistical index is smaller than $\alpha = .05$, the comparison is said to reach significance (i.e. $\alpha[PC]$ is used). Then redo the experiment say 10,000 times or better: ask a computer to do it.

In sum, there are 10,000 experiments, 10,000 families of comparisons and $5 \times 10,000 = 50,000$ comparisons. The results of one simulation are given in Table 12.1.

Number of families with X Type I errors	X: Number of Type I errors per family	Number of Type I errors
7,868	0	0
1,907	1	1,907
192	2	384
20	3	60
13	4	52
0	5	0
10,000		2,403

Table 12.1 Results of a Monte-Carlo simulation. Numbers of Type I errors when performing $C = 5$ comparisons for 10,000 analyses of variance performed on a 6-group design when the null hypothesis is true. How to read the table? For example, 192 families over 10,000 have 2 Type I errors, this gives $2 \times 192 = 384$ Type I errors.

Table 12.1 shows that the null hypothesis is rejected for 2,403 comparisons over the 50,000 comparisons actually performed ($5 \text{ comparisons} \times 10,000 \text{ experiments}$). From these data, an estimation of $\alpha[PC]$ is computed as:

$$\begin{aligned}\alpha[PC] &= \frac{\text{number of comparisons having reached significance}}{\text{total number of comparisons}} \\ &= \frac{2,403}{50,000} = .0479.\end{aligned}\quad (12.7)$$

This value falls close to the theoretical value of $\alpha = .05$ (recall that the null hypothesis is rejected whenever an observed F comes from the 5% ‘rare events’ when the null hypothesis is true).

It can be seen also that for 7,868 families (i.e. experiments) no comparison reaches significance. Equivalently for 2,132 families ($10,000 - 7,868$) at least one Type I error is made. From these data, $\alpha[PF]$ can be estimated as:

$$\begin{aligned}\alpha[PF] &= \frac{\text{number of families with at least one Type I error}}{\text{total number of families}} \\ &= \frac{2,132}{10,000} = .2132.\end{aligned}\quad (12.8)$$

This value falls close to the theoretical value given by the formula detailed previously:

$$\alpha[PF] = 1 - (1 - \alpha[PC])^C = 1 - (1 - .05)^5 = .226.$$

12.3.3 The problem with replications of a meaningless experiment: ‘alpha and the captain’s age’

The problem linked with the two different meanings of the α level occurs also when an experimenter replicates an experiment several times ‘trying’ various (meaningless) independent variables, and decides to publish the results of an experiment as soon as one reaches significance. It suffices to recall the example of ‘parapsychology and the coins’ (actually, the same problem would appear when using different dependent variables), to show that, if we try long enough, we are almost sure of finding a spurious significant result.

As an illustration, suppose you want to test the following theoretically utmost important hypothesis:

The age of the captain in the room next door has an influence on learning.

To do so, you decide to perform a two-group experiment. Subjects in the first group will learn and recall a series of pairs of syllables with a *young captain* in the room next door. Subjects in the second group will learn the same series of syllables but with an *old captain* in the room next door. In order to guarantee a reliable experiment, a ‘double-blind’ procedure is followed: neither the subjects nor the experimenters will know anything about the existence of any captain in the room next door. Your first attempt at proving your hypothesis fails, but you think that the cause of the failure comes from the color of the experimental room that interferes with the subjects’ concentration. So you decide to repaint the room, and try again. After a new failure, you realize that the color of the room of the captain was inadequate, and you decide to have it repainted. Then, you realize that the experiment should be run in the morning rather than in the afternoon (you have read a paper suggesting that sailors tend to be less effective in the afternoon). If the experiment is repeated 20 times with such unimportant variations, there is a .64 probability of finding at least one experiment significant. Repeat the experiment 50 times and the probability is .92.

Moral: *Chance seems to favor persistent experimenters!*

The problem, indeed, is that other experimenters will have trouble replicating such important (although controversial) results. The same problem occurs when several experimenters try independently to run the same (or almost the same) experiment (e.g. because the idea is ‘in the air’, and the topic is fashionable, etc.). This is complicated by the fact that people tend to report *only* experiments that reach significance.

To summarize, when the null hypothesis is true, if an important number of attempts is made to show the existence of a ‘non-existent’ effect, then it is highly probable that one of these attempts will give the spurious impression of having detected this effect. As this apparently successful attempt has a greater chance of being published than the non-successful ones (editors are notoriously reluctant to publish ‘non-significant’ results, or results that try to ‘prove the null hypothesis’), researchers in the field may be led to believe in its existence. If they replicate the original experiment, they will probably fail to replicate the (spurious) finding. However, if a large number of experimenters try to replicate, some at least may also obtain the spurious positive result. If the editor’s bias for ‘significant’ results is taken into account, then it is possible to obtain a series of papers where failures to replicate are followed by rebuttals followed by other failures to replicate *when in fact no effect exists*. As a consequence, some ‘experimental results’ may be very hard to duplicate simply because they do not really exist and are only artifacts!

12.3.4 How to correct for multiple comparisons: Šidàk and Bonferroni, Boole, Dunn

In the previous discussion about multiple comparisons, it was shown that the probability of making *as least one* Type I error for a family of C comparisons was

$$\alpha[PF] = 1 - (1 - \alpha[PC])^C.$$

with $\alpha[PF]$ being the Type I error for the family of comparisons; and $\alpha[PC]$ being the Type I error per comparison. This equation can be rewritten as

$$\alpha[PC] = 1 - (1 - \alpha[PF])^{1/C}.$$

This formula is derived assuming *independence* or *orthogonality* of the comparisons. It is sometimes called the Šidák equation.

Using it means that in order to obtain a given $\alpha[PF]$ level, we need to adapt the $\alpha[PC]$ values used for each comparison.

Because the Šidák equation involves a fractional power, it was quite difficult to compute by hand before the age of computers. Hence, an approximation was derived. Because it was derived independently by several authors, it is known under different names: *Bonferroni* inequality (the most popular name), *Boole* inequality, or even *Dunn* inequality. It is an approximation of the Šidák formula. It has the advantage of not involving a power function. This inequality relates $\alpha[PC]$ to $\alpha[PF]$ by

$$\alpha[PC] \approx \frac{\alpha[PF]}{C}.$$

The Šidák and the Bonferroni, Boole, Dunn correction formulas are linked to each other by the inequality

$$\alpha[PC] = 1 - (1 - \alpha[PF])^{1/C} \geq \frac{\alpha[PF]}{C}.$$

The Šidák and the Bonferroni formula give, in general, values very close to each other. As can be seen, the Bonferroni, Boole, Dunn inequality is a pessimistic estimation (it always does worse than Šidák equation). Consequently Šidák should be preferred. However, the Bonferroni, Boole, Dunn inequality is more well known (probably because it was easier to compute some time ago), and hence is used and cited more often.

The Šidák equation or Bonferroni, Boole, Dunn inequality are used to find a correction on $\alpha[PC]$ in order to keep $\alpha[PF]$ fixed. The general idea of the procedure is to correct $\alpha[PC]$ in order to obtain the overall $\alpha[PF]$ for the experiment. By deciding that the family is the unit for evaluating the Type I error, the Šidák equation gives the value for each $\alpha[PC]$, whereas the Bonferroni, Boole, Dunn inequality gives an approximation for each $\alpha[PC]$.

For example, suppose you want to perform four *independent* comparisons, and that you want to limit the risk of making at least one Type I error to an overall value of $\alpha[PF] = .05$, you will consider that any comparison of the family reaches significance if the probability associated with it is smaller than

$$\alpha[PC] = 1 - (1 - \alpha[PF])^{1/C} = 1 - (1 - .05)^{1/4} = .0127.$$

This is a change from the usual .05 and .01 values!

If you decide to use the Bonferroni, Boole, Dunn approximation, you will decide to reject the null hypothesis for each comparison if the probability associated with it is smaller than

$$\alpha[PC] = \frac{\alpha[PF]}{C} = \frac{.05}{4} = .0125,$$

which is very close to the exact value of .0127.

The fact that we may be in need of values different than the usual .01 and .05 values shows the interest of computer programs that give the actual probability associated with F .

If you do not have access to a computer running a program computing the probability associated with F , you can use the tables in the Appendix (see Table 5, page 506). These tables give the values of $F_{\text{critical Sidak}}$ to be used for testing the null hypothesis for a family of contrasts they can be used only to test comparisons with 1 degree of freedom (i.e. contrasts).

12.4 An example: context and memory

This example is inspired by an experiment by Smith (1979). The main purpose in this experiment was to show that to be in the same context for learning and for test can give a better performance than being in different contexts. More specifically, Smith wanted to explore the effect of putting oneself *mentally* in the same context. The experiment was organized as follows. During the learning phase, subjects learned a list made of 80 words in a room painted with an orange color, decorated with posters, paintings and a decent amount of paraphernalia. A first test of learning was then given, essentially to give subjects the impression that the experiment was over. One day later, subjects were unexpectedly re-tested for their memory. An experimenter would ask them to write down all the words of the list they could remember. The test took place in five different experimental conditions. Fifty subjects (10 per group) were randomly assigned to one of the five experimental groups.

The formula of the experimental design is $S(\mathcal{A})$ or $S_{10}(\mathcal{A}_5)$. The dependent variable measured is the number of words correctly recalled. The five experimental conditions were:

1. *Same context.* Subjects are tested in the same room in which they learned the list.
2. *Different context.* Subjects are tested in a room very different from the one in which they learned the list. The new room is located in a different part of the campus, is painted grey, and looks very austere.
3. *Imaginary context.* Subjects are tested in the same room as subjects from Group 2. In addition, they are told to try to remember the room in which they learned the list. In order to help them, the experimenter asks them several questions about the room and the objects in it.
4. *Photographed context.* Subjects are placed in the same condition as Group 3, and, in addition, they are shown photos of the orange room in which they learned the list.
5. *Placebo context.* Subjects are in the same condition as subjects in Group 2. In addition, before starting to try to recall the words, they are asked first to perform a warm-up task, namely to try to remember their living room.

Several research hypotheses can be tested with those groups. Let us accept that the experiment was designed to test the following research hypotheses:

- *Research hypothesis 1.* Groups for which the context at test matches the context during learning (i.e. is the same or is simulated by imaging or photography) will perform differently (precisely they are expected to do better) than groups with a different context or than groups with a placebo context.
- *Research hypothesis 2.* The group with the same context will differ from the group with imaginary or photographed context.
- *Research hypothesis 3.* The imaginary context group differs from the photographed context group.
- *Research hypothesis 4.* The different context group differs from the placebo group.

12.4.1 Contrasted groups

The four research hypotheses are easily transformed into statistical hypotheses. For example, the first research hypothesis is equivalent to stating the following null hypothesis:

The means of the population for groups 1, 3, and 4 have the same value as the means of the population for groups 2 and 5.

This is equivalent to contrasting groups 1, 3, 4 and groups 2, 5. This first contrast is denoted ψ_1 :

$$\psi_1 = 2\mu_1 - 3\mu_2 + 2\mu_3 + 2\mu_4 - 3\mu_5.$$

The null hypothesis to be tested is

$$H_{0,1} : \psi_1 = 0.$$

Note that a second subscript has been added, in order to distinguish the different null hypotheses from each other. The notation is a bit awkward, because the subscript of the contrast has the same status as the subscript meaning ‘null hypothesis’. However, it looks better than a notation like $H_0^{(1)}$ or even ${}_1H_0$. The first contrast is equivalent to defining the following set of coefficients C_a :

Gr. 1	Gr. 2	Gr. 3	Gr. 4	Gr. 5	$\sum_a C_a$	
+2	-3	+2	+2	-3	0	(12.9)

Note: The sum of the coefficients C_a is zero, as it should be for a contrast. Using the same line of reasoning, the second research hypothesis is translated into the contrast ψ_2 with the following set of coefficients: defining the second contrast (denoted ψ_2) as

Gr. 1	Gr. 2	Gr. 3	Gr. 4	Gr. 5	$\sum_a C_a$
+2	0	-1	-1	0	0

$$\psi_2 = 2\mu_1 + 0\mu_2 - 1\mu_3 - 1\mu_4 + 0\mu_5.$$

The null hypothesis to be tested is

$$H_{0,2} : \psi_2 = 0.$$

As an exercise, try to translate the other two hypotheses into contrasts. The following table gives the set of all four contrasts.

Comparison	Gr. 1	Gr. 2	Gr. 3	Gr. 4	Gr. 5	$\sum_a C_a$
ψ_1	+2	-3	+2	+2	-3	0
ψ_2	+2	0	-1	-1	0	0
ψ_3	0	0	+1	-1	0	0
ψ_4	0	+1	0	0	-1	0

Have you noted that the number of contrasts is the same as the number of degrees of freedom of the sum of squares for $\mathcal{A}$? Now the problem is to decide if the contrasts constitute

an orthogonal family. The first method would be to see if *logically* each comparison is independent of all the other comparisons in the family. For example, comparisons 3 and 4 involve different groups. Hence, they should be independent. Using the C_a coefficients gives an alternative way, more rapid and more secure, to check the independence of two contrasts.

12.5 Checking the independence of two contrasts

The coefficients C_a describe the *shape* of a contrast. One way to test whether two contrasts are independent (or orthogonal) could be to compute a coefficient of correlation (the Pearson r seen in Chapter 2) between the set of coefficients representing each contrast. If $C_{a,1}$ describes the set of coefficients for the first contrast, and $C_{a,2}$ the set of coefficients for the second contrast, then it is shown in the next section that this coefficient of correlation will be zero *if and only if*:

$$(C_{1,1} \times C_{1,2}) + (C_{2,1} \times C_{2,2}) + (C_{3,1} \times C_{3,2}) + (C_{4,1} \times C_{4,2}) + (C_{5,1} \times C_{5,2}) = 0.$$

Under the *normality assumption*, a zero correlation implies independence. Therefore, *two contrasts are independent when their correlation is zero*. With a more formal notation, two contrasts i and j are independent if and only if:

$$\sum_{a=1}^A C_{a,i} C_{a,j} = 0.$$

The term $\sum C_{a,i} C_{a,j}$ is often called the *scalar-product* or sometimes the *cross-product*⁴ of the contrast's coefficients. The term scalar-product comes from linear algebra where the sets of coefficients are seen as *vectors*, when the scalar-product between two vectors is zero, the vectors are orthogonal to each other. For example, to test the independence between ψ_1 and ψ_2 , the coefficients $C_{a,1}$ are +2, -3, +2, +2, -3, the coefficients $C_{a,2}$ are +2, 0, -1, -1, 0. The scalar-product between those two sets of coefficients is

$$(2 \times 2) + (-3 \times 0) + (2 \times -1) + (2 \times -1) + (-3 \times 0) = \sum_a C_{a,1} C_{a,2} = 0.$$

For a set of comparisons to be an orthogonal family, each possible pair of comparisons should be an orthogonal pair. An easy way to check whether the family is orthogonal is to build a table with all the pairs of comparisons, by putting comparisons in rows and in columns and by writing in each cell the cross-product of the coefficients. For the example on context and memory, our check for orthogonality can be presented as in Table 12.2.

This shows that all possible pairs of contrasts are orthogonal and hence, the family of contrast is orthogonal.

12.5.1 For experts: orthogonality of contrasts and correlation

This section shows that it is equivalent for two contrasts to have a zero coefficient of correlation or to have a null cross-product. You can skip it if you are already convinced.

⁴ Even though it has nothing to do with the cross-product of the physicists or engineers.

	ψ_1	ψ_2	ψ_3	ψ_4
ψ_1	*	0	0	0
ψ_2		*	0	0
ψ_3			*	0
ψ_4				*

Table 12.2 Orthogonality check for contrasts.

First, recall the formula of the coefficient of correlation between two sets of values, Y_s and X_s , with M_Y and M_X denoting the mean of respectively Y and X :

$$r_{Y \cdot X} = \frac{\sum (Y_s - M_Y)(X_s - M_X)}{\sqrt{\sum (Y_s - M_Y)^2 \sum (X_s - M_X)^2}}.$$

Obviously, $r_{Y \cdot X}$ is zero only if the numerator $\sum (Y - M_Y)(X - M_X)$ is zero. Now, for the particular case of a contrast, the sum of the C_a values is zero, and hence the mean of the C_a values is zero too. So, the numerator of the coefficient of correlation between the coefficients of two contrasts is

$$\sum_{a=1}^A C_{a,i} C_{a,j},$$

which is the *cross-product* of the two sets of coefficients. Therefore, if the cross-product is zero, the coefficient of correlation is also zero.

12.6 Computing the sum of squares for a contrast

A first possible way of computing the sum of squares for a contrast comes from the fact that a contrast is a particular case of an analysis of variance. A contrast is equivalent to putting the data into two groups: one group corresponding to the C_a values with a *plus* sign, and one group corresponding to the C_a values with a *minus* sign. Hence the sum of squares can be computed using the usual routine (described in Chapter 7).

Because this analysis is performed with two groups, the sum of squares for a contrast has just one degree of freedom. The mean squares used for evaluating the significance of the contrast, however, will come from the complete analysis of variance. This is because the mean square within group uses all the possible information in the experiment, and as a consequence should be the best possible estimator of the experimental error.

Actually, the C_a coefficients provide a more practical way of computing the sum of squares for a contrast. Specifically, the sum of squares for a contrast is denoted (depending upon the context) SS_ψ or SS_{comp} . It is computed as

$$SS_\psi = \frac{S(\sum C_a M_a)^2}{\sum C_a^2} = \frac{S\hat{\psi}^2}{\sum C_a^2},$$

where S is the number of subjects in a group, and $\hat{\psi} = \sum C_a M_a$. This formula will be illustrated shortly with Smith's memory and context example.

Some textbooks prefer using an alternative formula with the Y_a rather than the group means (i.e. the M_a values). With some elementary algebraic manipulations the previous formula can be expressed as

$$SS_{\psi} = \frac{(\sum C_a Y_a)^2}{S \sum C_a^2}.$$

12.7 Another view: contrast analysis as regression

We have seen previously that in order to assess the quality of a prediction of the dependent variable by a (quantitative) independent variable, we can use regression analysis. Remember that, in regression analysis, we try to predict Y , the dependent variable, from X , the independent variable. In the case of a contrast, we can consider that the C_a values constitute the prediction of the scores of the subjects in the a th experimental condition. Hence, we can test the quality of the prediction of a given contrast by computing the squared coefficient of correlation between the C_a and the Y_{as} values. This coefficient is expressed using two equivalent notations, $r_{Y \cdot \psi}^2$ and $R_{Y \cdot \psi}^2$. Because the sum of the C_a is zero, it is possible to find a convenient way of expressing $r_{Y \cdot \psi}^2$ in terms of the C_a . The formula is first given and its derivation is detailed in a digression that you can skip if you are willing to accept it as is. So the squared correlation between the C_a (i.e. a contrast ψ) and the scores obtained by the subjects is

$$r_{Y \cdot \psi}^2 = R_{Y \cdot \psi}^2 = \frac{\widehat{S\psi}^2}{\sum C_a^2} \times \frac{1}{SS_{\text{total}}},$$

which is equivalent to

$$r_{Y \cdot \psi}^2 = R_{Y \cdot \psi}^2 = \frac{SS_{\psi}}{SS_{\text{total}}}.$$

This last equation gives an alternative way of computing SS_{comp} or SS_{ψ} from the coefficient of correlation $R_{Y \cdot \psi}^2$ and the total sum of squares:

$$SS_{\psi} = SS_{\text{comp}} = R_{Y \cdot \psi}^2 \times SS_{\text{total}}.$$

In addition to providing a bridge between the techniques of regression and analysis of variance, the coefficient $r_{Y \cdot \psi}^2$ (or $R_{Y \cdot \psi}^2$) gives an index of the importance of a given contrast relative to the total sum of squares. As such it complements the analysis of variance by estimating the importance of an effect in addition to the significance level reached or not by the effect.

12.7.1 Digression: rewriting the formula of $R_{Y \cdot \psi}^2$

You can skip this section if you are scared by long formulas or if you are willing to accept the results of the previous paragraph without a proof. In this section we show that

$$\begin{aligned} r_{Y \cdot \psi}^2 &= R_{Y \cdot \psi}^2 \\ &= \frac{\widehat{S\psi}^2}{\sum C_a^2} \times \frac{1}{SS_{\text{total}}} \\ &= \frac{SS_{\psi}}{SS_{\text{total}}}. \end{aligned} \tag{12.10}$$

Let us start by recalling the formula of the squared coefficient of correlation between two variables (say Y and X):

$$r_{X,Y}^2 = \frac{\left[\sum (X_s - M_X)(Y_s - M_Y) \right]^2}{\sum (X_s - M_X)^2 \sum (Y_s - M_Y)^2}.$$

The first step is to replace the variable X by the C_a coefficients. Because the sum of the C_a is zero, their mean is also zero, and it can be dropped from the formula. We need also to keep in mind that the summation should be carried for all the values of s and a . Taking these two remarks into account, we can now rewrite the previous formula for $R_{Y,\psi}^2$ as

$$\begin{aligned} R_{Y,\psi}^2 &= r_{Y,\psi}^2 \\ &= \frac{\left[\sum_a^A \sum_s^S C_a (Y_{as} - M_{..}) \right]^2}{\sum_a^A \sum_s^S C_a^2 \sum_a^A \sum_s^S (Y_{as} - M_{..})^2}, \end{aligned}$$

but $\sum_a^A \sum_s^S (Y_{as} - M_{..})^2 = SS_{\text{total}}$, so we obtain

$$r_{Y,\psi}^2 = \frac{\left[\sum_a^A \sum_s^S C_a (Y_{as} - M_{..}) \right]^2}{\sum_a^A \sum_s^S C_a^2 SS_{\text{total}}}.$$

Because $\sum_a^A \sum_s^S C_a^2 = S \sum_a^A C_a^2 = S \sum C_a^2$, we can simplify the denominator:

$$r_{Y,\psi}^2 = \frac{\left[\sum_a^A \sum_s^S C_a (Y_{as} - M_{..}) \right]^2}{\left(S \sum C_a^2 \right) \times SS_{\text{total}}}.$$

Now, we can distribute the term C_a in the numerator to obtain

$$r_{Y,\psi}^2 = \frac{\left[\sum_a^A \sum_s^S (C_a Y_{as} - C_a M_{..}) \right]^2}{\left(S \sum C_a^2 \right) \times SS_{\text{total}}}.$$

Distributing the $\sum$ signs in the numerator gives

$$r_{Y,\psi}^2 = \frac{\left[\sum_a^A \sum_s^S C_a Y_{as} - \sum_a^A \sum_s^S C_a M_{..} \right]^2}{\left(S \sum C_a^2 \right) \times SS_{\text{total}}},$$

but $\sum_a^A C_a = 0$, hence $\sum_a^A \sum_s^S C_a M_{..} = M_{..} \sum_a^A \sum_s^S C_a = 0$,

and the numerator can be simplified as

$$r_{Y \cdot \psi}^2 = \frac{\left[\sum_a^A \sum_s^S C_a Y_{as} \right]^2}{\left(S \sum_a^A C_a^2 \right) \times SS_{\text{total}}}.$$

Recall that $\sum_s^S Y_{as} = Y_{a \cdot} = S M_{a \cdot}$, and simplify the numerator

$$\begin{aligned} r_{Y \cdot \psi}^2 &= \frac{\left[S \sum_a^A C_a M_{a \cdot} \right]^2}{\left(S \sum_a^A C_a^2 \right) \times SS_{\text{total}}} \\ &= \frac{S^2 \left(\sum_a^A C_a M_{a \cdot} \right)^2}{S \left(\sum_a^A C_a^2 \right) \times SS_{\text{total}}} \\ &= \frac{S \left(\sum_a^A C_a M_{a \cdot} \right)^2}{\left(\sum_a^A C_a^2 \right) \times SS_{\text{total}}} \\ &= \frac{S \left(\sum_a^A C_a M_{a \cdot} \right)^2}{\sum_a^A C_a^2} \times \frac{1}{SS_{\text{total}}} \\ &= \frac{S \psi^2}{\sum_a^A C_a^2} \times \frac{1}{SS_{\text{total}}} \\ &= \frac{SS_{\psi}}{SS_{\text{total}}}. \end{aligned} \tag{12.11}$$

Et voilà! (It was quite a long formula, hey?)

12.8 Critical values for the statistical index

Planned orthogonal comparisons are equivalent to independent questions addressed to the data. Because of this independence, there is almost a strong (but not complete—see what follows) consensus now between statisticians and psychologists alike, to consider each comparison as *if it were a single isolated analysis of variance*. This is equivalent to considering that for each contrast the Type I error per family is equal to the Type I error per comparison. You may have noticed that this is equivalent to ignoring the distinction between $\alpha[PF]$ and $\alpha[PC]$. The problem, however, in practice is not always that crucial,

especially because a contrast analysis approach is very often used with rival hypotheses. In this case what really matters is knowing what comparison(s) or theory(ies) can be supported by the data or which one(s) can be eliminated.

In brief, the current procedure is to act as if each contrast were the only contrast tested in the analysis of the data. Hence the null hypothesis for a given contrast is tested by computing

$$F_{\psi} = \frac{MS_{\psi}}{MS_{S(A)}},$$

which follows (when the comparison is considered as a single analysis of variance) a Fisher sampling distribution with $\nu_1 = 1$ and $\nu_2 = A(S - 1)$ degrees of freedom.

12.9 Back to the context ...

The data and results of the replication of Smith's experiment are given in Tables 12.3 and 12.4.

	Experimental context				
	Group 1 same	Group 2 different	Group 3 imagery	Group 4 photo	Group 5 placebo
	25	11	14	25	8
	26	21	15	15	20
	17	9	29	23	10
	15	6	10	21	7
	14	7	12	18	15
	17	14	22	24	7
	14	12	14	14	1
	20	4	20	27	17
	11	7	22	12	11
	21	19	12	11	4
$Y_{a\cdot}$	180	110	170	190	100
$M_{a\cdot}$	18	11	17	19	10
$M_{a\cdot} - M_{\cdot\cdot}$	3	-4	2	4	-5
$\sum(Y_{as} - M_{a\cdot})^2$	218	284	324	300	314

Table 12.3 Results of a replication of an experiment by Smith (1979). The dependent variable is the number of words recalled (see text section 12.4 for details).

Source	df	SS	MS	F	Pr(F)
$\mathcal{A}$	4	700.00	175.00	5.469**	.00119
$S(\mathcal{A})$	45	1,440.00	32.00		
Total	49	2,140.00			

Table 12.4 ANOVA table for a replication of Smith's (1979) experiment.

Recall that the sum of squares for a contrast is noted SS_{comp} or SS_{ψ} . The easiest way to compute it is to use the formula:

$$SS_{\psi} = \frac{S(\sum C_a M_a)^2}{\sum C_a^2} = \frac{S\hat{\psi}^2}{\sum C_a^2}$$

or the equivalent formula:

$$SS_{\psi} = \frac{(\sum C_a Y_a)^2}{S \sum C_a^2}.$$

Because the sum of squares for a contrast has one degree of freedom

$$MS_{\psi} = \frac{SS_{\psi}}{df_{\psi}} = \frac{SS_{\psi}}{1} = SS_{\psi}.$$

Note that strictly speaking, this is true only for contrasts, not for all comparisons. As we will see later, comparisons can have more than one degree of freedom; in this case, they are often called *sub-designs*. A contrast, actually, is a comparison with one degree of freedom.

The statistical index F_{ψ} is now computed as

$$F_{\psi} = \frac{MS_{\psi}}{MS_{S(A)}}.$$

For example, the steps for the computations of $SS_{\text{comp } 1}$ are given in the following table:

Group	M_a	C_a	$C_a M_a$	C_a^2
1	18.00	+2	+36.00	4
2	11.00	−3	−33.00	9
3	17.00	+2	+34.00	4
4	19.00	+2	+38.00	4
5	10.00	−3	−30.00	9
		0	45.00	30

$$SS_{\psi 1} = \frac{S(\sum C_a M_a)^2}{\sum C_a^2} = \frac{10 \times (45.00)^2}{30} = 675.00$$

$$MS_{\text{comp } 1} = 675.00$$

$$F_{\text{comp } 1} = \frac{MS_{\text{comp } 1}}{MS_{S(A)}} = \frac{675.00}{32.00} = 21.094. \quad (12.12)$$

When the null hypothesis is true, when the technical assumptions hold, and when the contrast is considered as being the only contrast of the family, then F_{ψ} will follow a Fisher distribution with 1 and $A(S - 1) = 45$ degrees of freedom. The critical value table gives the value 4.06 for $\alpha = .05$, and 7.23 for $\alpha = .01$. If we report this result in a journal article, we could write something like:

The first contrast is clearly significant, $F(1, 45) = 21.094$, $MS_e = 32.00$, $p < .01$. Hence, it seems that reinstating at test the context experienced during learning improves subjects' performance.

As additional practice, the computation of the other contrasts is detailed below. For Contrast 2, the sum of squares SS_{ψ_2} is:

Group	M_a	C_a	$C_a M_a$	C_a^2
1	18.00	+2	+36.00	4
2	11.00	0	0.00	0
3	17.00	-1	-17.00	1
4	19.00	-1	-19.00	1
5	10.00	0	0.00	0
		0	0.00	6

$$SS_{\psi_2} = \frac{S(\sum C_a M_a.)^2}{\sum C_a^2} = \frac{10 \times (0.0)^2}{6} = 0.$$

Therefore,

$$F_{\psi_2} = 0$$

and is not significant. For Contrast 3, we find:

Group	M_a	C_a	$C_a M_a$	C_a^2
1	18.00	0	0.00	0
2	11.00	0	0.00	0
3	17.00	+1	+17.00	1
4	19.00	-1	-19.00	1
5	10.00	0	0.00	0
		0	-2.00	2

$$SS_{\psi_3} = \frac{S(\sum C_a M_a.)^2}{\sum C_a^2} = \frac{10 \times (-2.00)^2}{2} = 20.00.$$

Therefore,

$$F_{\psi_3} = \frac{20.00}{32.00} = 0.625$$

and is not significant. For Contrast 4, we find:

Group	M_a	C_a	$C_a M_a$	C_a^2
1	18.00	0	0.00	0
2	11.00	+1	+11.00	1
3	17.00	0	0.00	0
4	19.00	0	0.00	0
5	10.00	-1	-10.00	1
		0	1.00	2

$$SS_{\psi_4} = \frac{S(\sum C_a M_a.)^2}{\sum C_a^2} = \frac{10 \times (1.00)^2}{2} = 5.00.$$

Therefore,

$$F_{\psi 4} = \frac{5.00}{32.00} = 0.156$$

and is not significant.

And now, an easy addition:

$$\begin{aligned} SS_{\psi 1} &= 675.00 \\ SS_{\psi 2} &= 0.00 \\ SS_{\psi 3} &= 20.00 \\ SS_{\psi 4} &= 5.00 \\ \hline SS_A &= 700.00 \end{aligned} \quad (12.13)$$

As you can see, the sums of squares for the 4 orthogonal contrasts add up to the sum of squares for factor A . Actually we have simply *sliced* the experimental sum of squares in 4 orthogonal slices: one per degree of freedom. The procedure is, in fact, the same procedure used previously to divide the total sum of squares into experimental sum of squares and within-group sum of squares. When the sums of squares are orthogonal, the degrees of freedom are added the same way as the sums of squares. Hence, the maximum number of orthogonal comparisons is given by the number of degrees of freedom of the experimental sum of squares.

12.10 Significance of the omnibus F vs significance of specific contrasts

When the family of comparisons is orthogonal, the information used (i.e. the sum of squares and the degrees of freedom) is the same as the information used for the omnibus F ratio. Hence, these two procedures are not independent. However, because the sums of squares are transformed into mean squares, the conclusions reached by both approaches can be discordant. For example, the omnibus F can be non-significant but some comparisons can reach significance. This can be the case, for example, when a strong part of the experimental sum of squares is ‘explained’ by a given contrast. The sum of squares for the contrast is also the mean square for the contrast. On the other hand the experimental sum of squares will be divided by the experimental degrees of freedom. Hence the experimental mean square can end up being too small to be significant. A numerical example will make that point clear. Suppose an experiment with 4 groups and 5 subjects per group in which $SS_A = 90$, and $MS_{S(A)} = 10$. The number of degrees of freedom for A is $(A - 1) = 3$, the number of degrees of freedom within groups $[S(A)]$ is equal to $A(S - 1) = 16$. The mean square for A is equal to $90/3 = 30$. The table of critical values for F gives 3.24 and 5.29 for 3 and 16 degrees of freedom and for $\alpha = .05$ and $\alpha = .01$. The criterion for A is

$$F_A = \frac{MS_A}{MS_{S(A)}} = \frac{30}{10} = 3.00.$$

This value is too small to reach significance. Suppose now, that the sum of squares for a contrast is 50. Then the criterion for ψ is

$$F_{\psi} = \frac{MS_{\psi}}{MS_{S(A)}} = \frac{50}{10} = 5.00.$$

The table of critical values for F gives 4.49 and 8.53 for 1 and 16 degrees of freedom and for $\alpha = .05$ and $\alpha = .01$. Hence, F_ψ is large enough for the contrast to be declared significant at the $\alpha = .05$ level.

The inverse pattern of results can also happen: the omnibus F can be significant with none of the comparisons reaching significance. Let us go back to the previous example. Suppose now that the experimental sum of squares is equal to 120. The mean square for $\mathcal{A}$ is now equal to $120/3 = 40$. The criterion for $\mathcal{A}$ is

$$F_A = \frac{MS_A}{MS_{S(A)}} = \frac{40}{10} = 4.00.$$

This is large enough for the effect of $\mathcal{A}$ to reach significance with 3 and 16 degrees of freedom. Suppose now that we have a family made of 3 orthogonal contrasts, each of them explaining the same proportion of the total sum of squares. Then each of them will have a sum of squares equal to $120/3 = 40$. This will give the following value for the F_ψ criterion for ψ :

$$F_\psi = \frac{MS_\psi}{MS_{S(A)}} = \frac{40}{10} = 4.00,$$

which is too small to reach significance with 1 and 16 degrees of freedom.

12.11 How to present the results of orthogonal comparisons

Because the omnibus F ratio and the planned comparisons use the same information, they should not be used simultaneously. Ideally, the choice of one strategy of data analysis should be done *before* running the experiment. Because a set of orthogonal comparisons decomposes the sum of squares of $\mathcal{A}$, it is often the custom to report the result of the data analysis with the following type of analysis of variance table:

Source	df	SS	MS	F
Experimental treatment	(4)	(700.00)	—	—
$\psi_1 : (1 + 3 + 4) \text{ vs } (2 + 5)$	1	675.00	675.00	21.094**
$\psi_2 : (1) \text{ vs } (3 + 4)$	1	0.00	0.00	0.000 <i>ns</i>
$\psi_3 : (3) \text{ vs } (4)$	1	20.00	20.00	0.625 <i>ns</i>
$\psi_4 : (2) \text{ vs } (5)$	1	5.00	5.00	0.156 <i>ns</i>
Error: $S(\mathcal{A})$	45	1,440.00	32.00	
Total	49	2,140.00		

ns: suspend judgment for $\alpha = .05$

** $p < .01$

The table allows us to evaluate the relative importance of each contrast. For example the first contrast just by itself explains 96% of the experimental sum of squares ($675.00/700.00 = 0.96$). Hence we can conclude that most of the experimental effect can be explained by the opposition between having a correct context (imagined or real) and having no context or an inappropriate one. Figure 12.1 confirms that interpretation of the results.

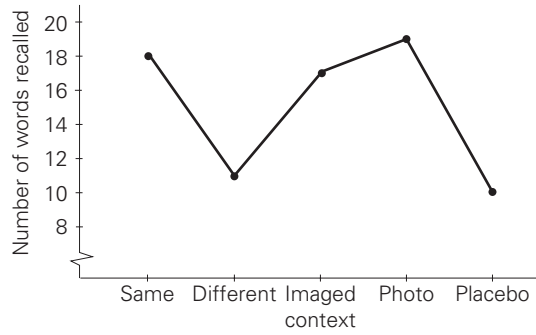


Figure 12.1 Mean number of words recalled one day after learning as a function of type of context at test [design $\mathcal{S}(4)$]. Data from a replication of Smith's experiment (1979).

12.12 The omnibus F is a mean!

When a family of orthogonal contrasts uses all the available experimental degrees of freedom, the mean of the F_ψ values of the contrasts gives the omnibus F .

This is shown with the following derivation:

$$F_{\text{omnibus}} = \frac{SS_A/A - 1}{MS_{S(A)}} = \frac{\sum SS_{\text{comp}}/A - 1}{MS_{S(A)}} = \frac{\sum F_{\text{comp}}}{A - 1} = \frac{1}{A - 1} \sum F_{\text{comp}}.$$

Remember: the sum of $(A - 1)$ orthogonal contrasts is equal to the experimental sum of squares (cf. Equation 12.13).

12.13 Sum of orthogonal contrasts: sub-design analysis

In some cases, the experimenter wants to analyze a sub-design out of the complete design specifying the experiment. As an illustration, suppose we take a modified version of the first example in which performance for a video game was studied. Suppose we are designing a new experiment with four experimental groups: a placebo group, and three groups receiving three different forms of the same drug (say a barbiturate). The experimenter is interested only in two questions:

- *First.* Does taking a barbiturate (whatever its form) affect video game performance?
- *Second.* Do the different forms of the barbiturate differ in the way they affect video game performance?

The first question can easily be translated into a contrast opposing the first group (placebo) to the three barbiturate groups. This contrast will use the following set of C_a coefficients:

$$+3 \quad -1 \quad -1 \quad -1.$$

The second question, however, cannot be expressed by a contrast. It is, actually, equivalent to testing the null hypothesis stating that 'the population means of groups 2, 3, and 4 are equal'. Or with a formula:

$$H_{0,2} : \mu_2 = \mu_3 = \mu_4.$$

This is equivalent to analyzing a part of the whole experiment. This part is called a *sub-design*. It is a 3-level $\mathcal{S}(\mathcal{A})$ sub-design extracted from a 4-level $\mathcal{S}(\mathcal{A})$ design. Because the sub-design involves three groups, it has 2 degrees of freedom. Hence the experimental sum of squares of that sub-design can be decomposed into two orthogonal contrasts. A possible set of orthogonal contrasts could be composed by a first contrast opposing one form of the barbiturate to the other two forms, and a second contrast opposing the last two forms to each other. The following table gives a possible set of orthogonal contrasts. The first contrast corresponds to the first question, the second and third contrasts correspond to the second question.

	Experimental group			
	Placebo	Barbiturate 1	Barbiturate 2	Barbiturate 3
ψ_1	3	-1	-1	-1
ψ_2	0	2	-1	-1
ψ_3	0	0	1	-1

These three contrasts are orthogonal to each other (check it as an exercise). Hence, if we build a 2 degree of freedom comparison by adding contrast 2 to contrast 3 ($\psi_2 + \psi_3$), this comparison will be orthogonal to the first contrast. Note that the notion of comparison is broader than the notion of contrast. Specifically, a contrast is a comparison with only one degree of freedom. A comparison or a sub-design can have several degrees of freedom. The sum of squares and the degrees of freedom for a comparison are easily computed by adding the sums of squares and degrees of freedom of its orthogonal components. For the previous example:

$$\begin{aligned} SS_{\text{comp } 1} &= SS_{\psi_1}, \\ SS_{\text{comp } 2} &= SS_{\psi_2} + SS_{\psi_3}, \end{aligned} \quad (12.14)$$

and

$$df_{\text{comp } 2} = df_{\psi_2} + df_{\psi_3} = 1 + 1 = 2.$$

12.13.1 Sub-design analysis: an example

The subjects' performance for the experiment described in Section 12.2 (page 225) was measured by taking their score on some video game. Ten subjects were assigned to each experimental group. The within-group mean square [$MS_{S(\mathcal{A})}$] is 400.00, and the group means are the following:

Experimental group			
Placebo	Barbiturate 1	Barbiturate 2	Barbiturate 3
105	123	119	133

The three contrasts give the following sums of squares:

$$\begin{aligned} SS_{\psi_1} &= 3,000.00, \\ SS_{\psi_2} &= 60.00, \end{aligned}$$

$$SS_{\psi_3} = 980.00, \quad (12.15)$$

with

$$MS_{S(A)} = 400.00.$$

The first comparison is evaluated by computing

$$F_{\text{comp } 1} = \frac{MS_{\text{comp } 1}}{MS_{S(A)}} = \frac{MS_{\psi_1}}{MS_{S(A)}} = \frac{3,000.00}{400.00} = 7.50.$$

When the null hypothesis is true and when the technical assumptions hold, $F_{\text{comp } 1}$ is distributed as a Fisher F with 1 and 36 degrees of freedom. Using a .05 α level, the critical value is 4.11 and the null hypothesis is rejected. The sum of squares for the second comparison is obtained by adding the sums of squares of contrasts 2 and 3:

$$SS_{\text{comp } 2} = SS_{\psi_2} + SS_{\psi_3} = 60.00 + 980.00 = 1,040.00.$$

This sum of squares has 2 degrees of freedom, hence the mean square for comparison 2 is computed by dividing the sum of squares by its degrees of freedom:

$$MS_{\text{comp } 2} = \frac{SS_{\text{comp } 2}}{df_{\text{comp } 2}} = \frac{1,040.00}{2} = 520.00.$$

The second comparison is evaluated by computing

$$F_{\text{comp } 2} = \frac{MS_{\text{comp } 2}}{MS_{S(A)}} = \frac{520.00}{400.00} = 1.30.$$

When the null hypothesis is true and when the technical assumptions hold, $F_{\text{comp } 2}$ is distributed as a Fisher F with 2 and 36 degrees of freedom. Using a .05 α level, the critical value is 3.26 and the null hypothesis cannot be rejected.

12.14 Trend analysis

A particular case of orthogonal contrast analysis can be applied when the independent variable is quantitative. In this case, we often want to analyze the data in terms of *trends*. Specifically, we want to know if the pattern in the data can be fitted by simple shapes which correspond to basic polynomial functions. The most simple polynomial is a line (this is a degree one polynomial). In this case, we want to know if the group means are positioned on a line. The next polynomial is a *quadratic* polynomial. In this case, the group means are positioned such that the means are a function of the squared values of the independent variable. The next step is a *cubic* polynomial. In this case, the group means are positioned such that the means are a function of the cube of the independent variable. The next step will involve a polynomial of degree four, and so on. With A levels of the independent variable, the largest degrees for a polynomial is $(A - 1)$, which is the number of degrees of freedom associated with this independent variable. Formally, if we denote X_a the a th level of the independent variable, the experimental means are expressed as the following function:

$$M_{a \cdot} = w_0 + w_1 X_1 + w_2 X_2^2 + \cdots + w_{A-1} X_{A-1}^{A-1}. \quad (12.16)$$

In this equation, the w_a term represents the a th coefficient of the polynomial, and the term $w_a X_a$ represents the component of degree a (i.e. when $a = 1$, this is the *linear*, when $a = 2$, this is the *quadratic*, ...).

The goal of trend analysis is to determine the importance of each of these components and also to assess their significance. The basic idea is simple: in order to assess the importance of a component, it suffices to express each component of the polynomial as a contrast and run a contrast analysis. The analysis is particularly easy, when the design is *balanced* and when the levels of the independent variable are equally spaced because we can use sets of contrast coefficients that are designed to be orthogonal and to express each component of the polynomial.⁵ The contrasts corresponding to these coefficients are called *orthogonal polynomial* and they are listed in Table 7 (page 512) in the Appendix. For example, suppose that we have $A = 5$ levels of a quantitative independent variable, with values of 1, 2, 3, 4, and 5. Figure 12.2 plots the values of the orthogonal polynomial obtained from the table in the appendix. We can see that the linear trend, for example, is described by the

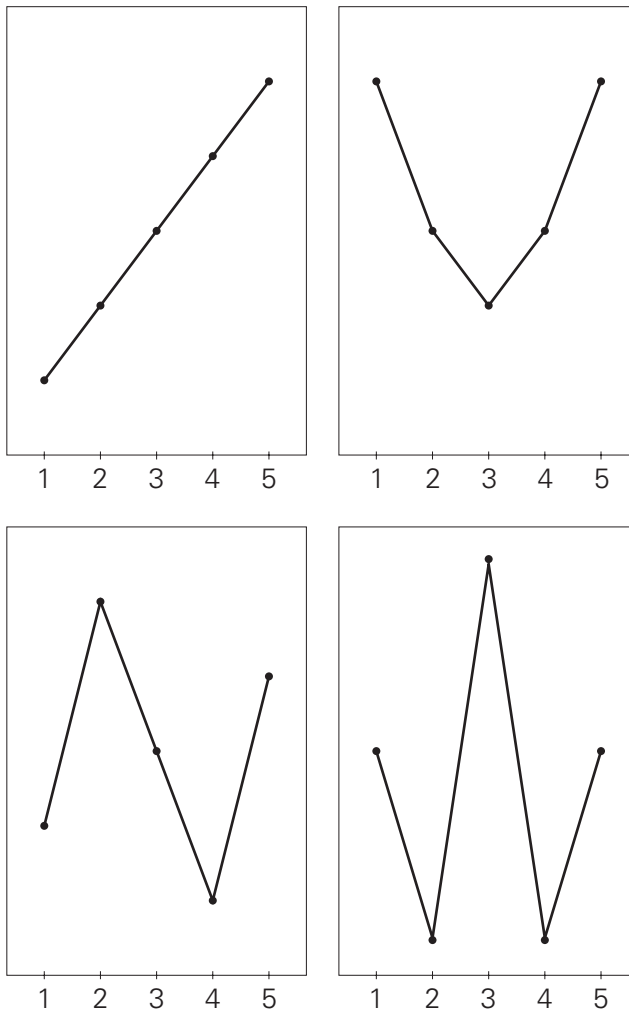


Figure 12.2 Examples of trend: (1) linear, (2) quadratic, (3) cubic and (4) fourth degree.

⁵ When the design is unbalanced or when the levels of the independent variable are not equally spaced, we need to go back to a standard regression analysis approach (see Gaito, 1965).

set of coefficients: $-2, -1, 0, 1$, and $+2$. The quadratic trend is described by the set of coefficients: $2, -1, -2, -1$, and $+2$.

Let us take an example. We have assigned 70 participants to one of 5 groups for wine tasting (i.e. $S = 14$, and $A = 5$). The wines had been manipulated such that the participants from group 1 would drink wines with an alcohol content of 8%, participants from groups 2, 3, 4 and 5 would taste wines with an alcohol content respectively of 10%, 12%, 14%, and 16%. We collected for each participant the average hedonic scores for the wines tasted (as is the tradition for wine tasting, the participants responding by giving a hedonic score ranging from 0 to 100, with 100 being a ‘perfect’ wine). The average scores per group are listed in Table 12.5. We first ran an analysis of variance, and found that $MS_{S(A)} = 500.00$. From the table of orthogonal polynomials, we found the following coefficients.

Linear	−2	−1	0	1	2
Quadratic	2	−1	−2	−1	2
Cubic	−1	2	0	−2	1
Fourth	1	−4	6	−4	1

With these coefficients, we can perform a standard contrast analysis whose results are displayed in Table 12.6. From this table and from Figure 12.3, we find that only the linear and the quadratic trends are of importance for this experiment. The linear component indicates that our tasters feel that the more alcohol the better the wine, the quadratic component indicates that at a certain point (i.e. past 14%) adding alcohol *decreases* the hedonic component.

Note that with trend analysis, the difference between contrast analysis and regression analysis becomes blurred because we are now evaluating the correlation between a quantitative independent variable and the dependent variable and this looks very much like orthogonal multiple regression.

X_a Percent alcohol content	8	10	12	14	16
M_a Average hedonic score	40	70	90	100	80

Table 12.5 Average hedonic score ($S = 14$) for 5 groups of wine tasting.

Source	$R^2_{A \cdot \psi}$	df	SS	MS	F	p
Experimental	—	(4)	(29,680)	—	—	
ψ_1 : Linear	.571	1	16,940	16,940	33.88**	.000,000,2
ψ_2 : Quadratic	.408	1	12,100	12,100	24.19**	.000,996,2
ψ_3 : Cubic	.018	1	560	560	1.12 <i>ns</i>	.292,856,1
ψ_4 : Fourth degree	.003	1	80	80	0.16 <i>ns</i>	.690,468,0
Error: $S(A)$	—	65	32,500	500		
Total	—	69	62,180			

ns: suspend judgment for $\alpha = .05$ ** $p < .01$

Table 12.6 Analysis of variance table for the wine tasting experiment and trend analysis.

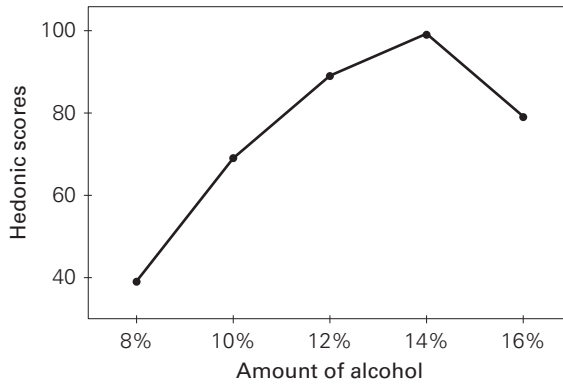


Figure 12.3 Average hedonic score ($S = 14$) for 5 groups of wine tasting.

Chapter summary

12.15 Key notions of the chapter

Below are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

Omnibus test	α per comparison and α per family of comparisons
Contrast or comparison analysis	Captain's age
Planned (<i>a priori</i>) vs (<i>a posteriori</i>) comparisons	Sub-design analysis
Orthogonal (independent) comparisons	Trend analysis (linear, quadratic, cubic)
Family of comparisons	

12.16 New notations

Below are the new notations introduced in this chapter. Test yourself on their meaning.

ψ and C_a
 $\alpha[PC]$ and $\alpha[PF]$
 SS_ψ , df_ψ , MS_ψ , and F_ψ
 SS_ψ , df_ψ , MS_ψ , and F_ψ

12.17 Key formulas of the chapter

Over the page are the main formulas introduced in this chapter: try to go through them and understand what they mean.

Independence of two contrasts:

$$\sum_{a=1}^A C_{a,i} C_{a,j} = 0$$

α per family:

$$\alpha(PF) = 1 - [1 - \alpha(PC)]^C$$

$$\begin{aligned} SS_{\psi} &= \frac{S(\sum C_a M_a)^2}{\sum C_a^2} \\ &= \frac{\widehat{S_{\psi}}^2}{\sum C_a^2} \\ &= \frac{(\sum C_a Y_a)^2}{S \sum C_a^2} \\ &= R_{Y \cdot \psi}^2 SS_{\text{total}} \end{aligned}$$

12.18 Key questions of the chapter

Below are some questions about the content of this chapter. All the answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- * When would you use a contrast analysis?
- * What is the maximum size of a family of orthogonal contrasts?
- * Why do we need to differentiate between α per family and α per comparison?
- * Is it possible to obtain a significant contrast with an *F omnibus* not significant?
- * When do we want to test for trend?

13

ANOVA, one factor: planned non-orthogonal comparisons

13.1 Introduction

We have seen in the previous chapter that orthogonal comparisons are relatively straightforward. Each comparison can be evaluated on its own. The only problem (often ignored) is to correct for the increase in Type I error resulting from multiple tests. The simplicity of orthogonal comparisons parallels the simplicity of multiple orthogonal regression (because they are essentially the same technique).

Non-orthogonal comparisons are more complex. The main problem is to assess the importance of a given comparison concurrently with the other comparisons of the set. There are currently two (main) approaches to this problem. The classical approach corrects for multiple statistical tests (e.g. using a Šidák or Bonferroni correction), but essentially evaluates each contrast as if it were coming from a set of orthogonal contrasts. The multiple regression (or modern) approach evaluates each contrast as a predictor from a set of non-orthogonal predictors and estimates its *specific* contribution to the explanation of the dependent variable. In fact, these two approaches correspond to different questions asked about the data coming from an experiment. The classical approach evaluates each comparison for itself, the multiple approach evaluates each comparison as a member of a set of comparisons and estimates the specific contribution of each comparison in this set. When the set of comparisons is an orthogonal set, these two approaches are completely equivalent. Therefore, when designing an experiment, it is always good practice, whenever possible, to translate the experimental hypotheses into a set of orthogonal comparisons. A lot of interesting experiments can be designed that way. However (and unfortunately), some research hypotheses or sets of research hypotheses cannot be implemented as a set of orthogonal comparisons (and this is why we need to learn about non-orthogonal comparisons!).

Some problems are created by the use of multiple non-orthogonal comparisons. Recall that the most important one is that the greater the number of comparisons, the greater the risk of rejecting H_0 when it is actually true (a Type I error). The general strategy adopted by the classical approach to take this problem into account depends upon the type of comparisons to be performed.

Here two cases will be examined:

- The general case of a set of non-orthogonal comparisons. This corresponds to the *Šidàk* and *Bonferroni* tests.
- The particular case of the comparison of several experimental groups with one control group. This is known as the *Dunnett* test.

Recall that only the evaluation of the probability associated with F_{comp} changes with the different tests. Formalization of the comparisons and computation of the F test follow the procedure presented in the previous chapter.

When two contrasts are non-orthogonal, the prediction corresponding to a contrast overlaps with the prediction made by the other one. This can be a problem if the experiment was designed to test *rival* hypotheses, with each hypothesis being translated into a comparison. In this case, we want to evaluate the *specific* explanatory power of each comparison (i.e. theory). The classical approach does not handle this situation in a coherent manner because the explanatory power (i.e. the coefficient of correlation) of each comparison is evaluated as if it were coming from a set of orthogonal comparisons. The multiple regression approach gives a more coherent framework by using semi-partial coefficients of correlation to evaluate the specific contribution of each comparison.

In what follows, we first examine the classical approach and then we detail the multiple regression approach.

13.2 The classical approach

13.2.1 Šidàk and Bonferroni, Boole, Dunn tests

In the previous chapter it was shown that when several orthogonal comparisons are performed, the $\alpha[PC]$ level necessary to attain a given $\alpha[PF]$ is given by the Šidàk equation:

$$\alpha[PC] = 1 - (1 - \alpha[PF])^{1/C} .$$

This formula, however, is derived assuming *independence* or *orthogonality* of the comparisons. What happens when the comparisons do not constitute an orthogonal family? Then, the equality gives a lower bound for $\alpha[PC]$ (cf. Šidàk, 1967; Games, 1977). So, instead of having the previous equality, the following inequality, called the Šidàk inequality, holds:

$$\alpha[PF] \leq 1 - (1 - \alpha[PC])^C .$$

This inequality gives the maximum possible value of $\alpha[PF]$, hence we know that the real value of $\alpha[PF]$ will always be smaller than $1 - (1 - \alpha[PC])^C$.

As previously, we can approximate the Šidàk inequality by the *Bonferroni*, *Boole*, *Dunn* inequality.

The Bonferroni inequality relates $\alpha[PC]$ to $\alpha[PF]$ as

$$\alpha[PF] < C\alpha[PC] .$$

Šidák and Bonferroni, Boole, Dunn are linked to each other by the inequality

$$\alpha[PF] \leq 1 - (1 - \alpha[PC])^C < C\alpha[PC] .$$

As we have seen, the numerical values given by the Šidák and the Bonferroni approaches are, in general, very close to each other.

The Šidák or the Bonferroni, Boole, Dunn inequalities are used to find a correction on $\alpha[PC]$ in order to keep $\alpha[PF]$ fixed. The general idea of the procedure is to correct $\alpha[PC]$ in order to obtain the overall $\alpha[PF]$ for the experiment. By deciding that the family is the unit for evaluating Type I error, the inequalities give an approximation for each $\alpha[PC]$. The formula used to evaluate the alpha level for each comparison using the Šidák inequality is

$$\alpha[PC] \approx 1 - (1 - \alpha[PF])^{1/C} .$$

This is a conservative approximation, because the following inequality holds:

$$\alpha[PC] \geq 1 - (1 - \alpha[PF])^{1/C} .$$

The formula used to evaluate the alpha level for each comparison using Bonferroni, Boole, Dunn inequality would be

$$\alpha[PC] \approx \frac{\alpha[PF]}{C} .$$

By using these approximations, the statistical test will be a *conservative* one. That is to say, the real value of $\alpha[PF]$ will always be smaller than the approximation we use. For example, suppose you want to perform four non-orthogonal comparisons, and that you want to limit the risk of making at least one Type I error to an overall value of $\alpha[PF] = .05$. Using the Šidák correction, you will consider that any comparison of the family reaches significance if the probability associated with it is smaller than:

$$\alpha[PC] = 1 - (1 - \alpha[PF])^{1/C} = 1 - (1 - .05)^{1/4} = .0127 .$$

(This is again a change from the usual .05 and .01 values!)

If you do not have access to a computer running a program computing the probability associated with the criterion, you can use Table 5 (page 506) and 6 (page 509) in the Appendix. These tables give the values of $F_{\text{critical Šidák}}$ to be used for testing the null hypothesis for a family of contrasts. These tables can be used only to test comparisons with 1 degree of freedom (i.e. contrasts).

13.2.2 Splitting up $\alpha[PF]$ with unequal slices

Compared with Šidák, the Bonferroni, Boole, Dunn is very easy to compute. An additional advantage of the Bonferroni, Boole, Dunn rule over the Šidák is that it gives experimenters the possibility of making an unequal allocation of the whole $\alpha[PF]$. This works because when using the Bonferroni, Boole, Dunn approximation, $\alpha[PF]$ is the sum of the individual $\alpha[PC]$:

$$\alpha[PF] \approx C\alpha[PC] = \underbrace{\alpha[PC] + \alpha[PC] + \cdots + \alpha[PC]}_{C \text{ times}} .$$

If some comparisons are judged more important *a priori* than some others, it is possible to allocate unequally $\alpha[PF]$ (cf. Rosenthal and Rosnow, 1985). For example, suppose we

have three comparisons that we want to test with an overall $\alpha[PF] = .05$, and we think that the first comparison is the most important of the set. Then we can decide to test it with $\alpha[PC] = .04$, and share the remaining value $.01 = .05 - .04$ between the last 2 comparisons, which will be tested each with a value of $\alpha[PC] = .005$. The overall Type I error for the family is equal to $\alpha[PF] = .04 + .005 + .005 = .05$ which was indeed the value we set beforehand. It should be emphasized, however, that the (subjective) importance of the comparisons and the unequal allocation of the individual $\alpha[PC]$ should be decided *a priori* for this approach to be statistically valid. An unequal allocation of the $\alpha[PC]$ can also be achieved using the Šidák inequality, but it is slightly more tricky to handle and gives very close results to the additive allocation of the Bonferroni, Boole, Dunn inequality.

13.2.3 Bonferonni *et al.*: an example

An example will help to review this section. Let us go back to Bransford’s ‘Romeo and Juliet’ experiment (cf. Chapter 8, Section 8.7, page 157). The following table gives the different experimental conditions:

Context before	Partial context	Context after	Without context

Suppose that Bransford had build his experiment to test *a priori* four research hypotheses:

- 1. The presence of any context has an effect.
- 2. The context given after the story has an effect.
- 3. The context given before has an effect stronger than any other condition.
- 4. The partial context condition differs from the ‘context before’ condition.

These hypotheses can easily be translated into a set of contrasts given in the following table.

	Context before	Partial context	Context after	Without context
ψ_1	1	1	1	−3
ψ_2	0	0	1	−1
ψ_3	3	−1	−1	−1
ψ_4	1	−1	0	0

As you can check, this family of contrasts is not composed of orthogonal contrasts (actually, the number of contrasts guarantees that they cannot be independent. Do you remember why?).

If $\alpha[PF]$ is set to the value .05, this will lead to testing each contrast with the $\alpha[PC]$ level:

$$\alpha[PC] = 1 - .95^{1/4} = .0127 \text{ .}$$

If you want to use the critical values method, the table gives for $\nu_2 = 16$ (this is the number of degrees of freedom of $MS_{S(A)}$), $\alpha[PF] = .05$, and $C = 4$ the value $F_{\text{critical Šidák}} = 7.86$ (this is simply the critical value of the standard Fisher F with 1 and 16 degrees of freedom and with $\alpha = \alpha[PC] = .0127$).

As an exercise, let us compute the sums of squares for the contrasts. Recall that $MS_{S(A)} = 2.35$, and $S = 5$. The experimental means are given in the following table.

	Context before	Partial context	Context after	Without context
$M_{a\cdot}$	7.00	4.20	3.20	3.00

For the first contrast:

$$SS_{\psi_1} = 12.15, \quad F_{\psi_1} = \frac{MS_{\psi_1}}{MS_{S(A)}} = \frac{12.15}{2.35} = 5.17.$$

The value 5.17 is smaller than $F_{\text{critical } \check{\text{S}}\text{id}\check{\text{a}}\text{k}} = 7.86$, hence the null hypothesis cannot be rejected for that contrast. *Note* that the null hypothesis would have been rejected if this first contrast were the only comparison made, or if the comparisons were orthogonal.

For the second contrast:

$$SS_{\psi_2} = 0.1, \quad F_{\psi_2} = 0.04.$$

The value 0.04 is smaller than $F_{\text{critical } \check{\text{S}}\text{id}\check{\text{a}}\text{k}} = 7.86$, hence the null hypothesis cannot be rejected for this contrast.

For the third contrast:

$$SS_{\psi_3} = 46.8, \quad F_{\psi_3} = 19.92.$$

The value 19.92 is larger than $F_{\text{critical } \check{\text{S}}\text{id}\check{\text{a}}\text{k}} = 7.86$, hence the null hypothesis is rejected for this contrast.

For the fourth contrast:

$$SS_{\psi_4} = 19.6, \quad F_{\psi_4} = 8.34.$$

The value 8.34 is larger than $F_{\text{critical } \check{\text{S}}\text{id}\check{\text{a}}\text{k}} = 7.86$, hence the null hypothesis is rejected for this contrast.

13.2.4 Comparing all experimental groups with the same control group: Dunnett's test

When an experimental design involves a control group and several experimental groups, quite often the only comparisons of interest for the experimenter are to oppose each of the experimental groups to the control group. This, in particular, is the case when several experiments are run at the same time for the only purpose of showing the effect of each experiment independently of the others. The practice of using only one common control group is more economical than running a separate control group for each experimental condition.

These comparisons are clearly not orthogonal to each other. Actually, the correlation between any pair of contrasts is equal to .5. (Can you prove it?) This makes it possible to compute the real value of $\alpha[PF]$ rather than using the $\check{\text{S}}\text{id}\check{\text{a}}\text{k}$ approximation. The critical values are given in Table 8 (page 513) in the Appendix. This table gives the value of $F_{\text{critical Dunnett}}$ as a function of $\alpha[PF]$, of the number of groups A (i.e. including the control group), and of the number of degrees of freedom of $MS_{S(A)}$. Note, in passing, that the critical values for the Dunnett test, when $A = 2$, are the same as the standard Fisher distribution with 1 and $A(S - 1)$ degrees of freedom (why?).

Experimental group	SS_{comp}	F_{comp}	Decision
Context before	40.00	17.02	Reject H_o
Partial context	3.60	1.53	Suspend judgment
Context after	0.10	.04	Suspend judgment

Table 13.1 Dunnett’s test for ‘Romeo and Juliet’.

If you compare the critical values for Dunnett and Šidák, you will, indeed, remark that it is always easier to reject the null hypothesis with Dunnett than with Šidák or with Bonferroni, Boole, Dunn. (Why? If you do not know, re-read the previous chapter about *a priori* comparisons.)

For example, suppose that for Bransford’s ‘Romeo and Juliet’ the family of comparisons opposes each context group to the ‘no context’ group. The table for the Dunnett test gives a critical value of 6.71 for $\alpha[PF] = .05$, whereas the table for the Šidák test gives a value of 7.10 for the same $\alpha[PF] = .05$ level.

As an illustration, Table 13.1 gives the F values for an analysis contrasting each experimental group with the control group.

13.3 Multiple regression: the return!

We have mentioned in Chapter 9 that analysis of variance and multiple regression are equivalent if we use as many predictors for the multiple regression analysis as the number of degrees of freedom of the independent variable. An obvious choice for the predictors is to use a set of coefficients corresponding to a set of contrasts. Doing so makes contrast analysis a particular case of multiple regression analysis. Regression analysis, in return, helps to solve some of the problems associated with the use of non-orthogonal contrasts: it suffices to use multiple regression and semi-partial coefficients of correlation to analyze non-orthogonal contrasts. In this section, we illustrate the multiple regression analysis approach of the Bransford experiment (i.e. ‘Romeo and Juliet’; cf. Chapter 8, Section 8.7, page 157) with a set of orthogonal contrasts and with a set of non-orthogonal contrasts.

The main points of this section are summarized below:

- Analyzing an experiment with a set of contrasts or with multiple regression is completely equivalent.
- Life is easier with orthogonal multiple regression and with orthogonal contrasts.
- Even though it is not (yet!) traditionally encouraged, correcting the alpha level to take into account the problem of multiple tests is a good idea.
- When dealing with *a priori* non-orthogonal contrasts, evaluating them with semi-partial coefficient of correlation is often a good policy.

13.3.1 Multiple regression: orthogonal contrasts for Romeo and Juliet

Let us go back once again to Bransford’s ‘Romeo and Juliet’ experiment (cf. Chapter 8, Section 8.7, page 157). The independent variable is the ‘type of context’ with four levels: context before, partial context, context after, without context.

Experimental condition			
Context before	Partial context	Context after	Without context
5	5	2	3
9	4	4	3
8	3	5	2
4	5	4	4
9	4	1	3

Table 13.2 The data from a replication of Bransford's 'Romeo and Juliet' experiment. $M.. = 4.35$.

Contrast	Experimental groups			
	1	2	3	4
ψ_1	1	1	-1	-1
ψ_2	1	-1	0	0
ψ_3	0	0	1	-1

Table 13.3 An arbitrary set of orthogonal contrasts for analyzing 'Romeo and Juliet'.

A set of data from a replication of this experiment is given in Table 13.2.

In order to analyze these data with a multiple regression approach we can use any arbitrary set of contrasts as long as they satisfy the following constraints:

1. There are as many contrasts as the independent variable has degrees of freedom.
2. The set of contrasts is not multicollinear (cf. Chapter 6, Section 6.7.2, p. 126ff.). That is, no contrast can be obtained by combining the other contrasts.¹

In this section, we illustrate the lessons of Chapter 6 stating that orthogonal multiple regression is easier than non-orthogonal multiple regression. Remember that the main advantage of orthogonal multiple regression (over non-orthogonal regression) is to give the parameters for the multiple regression directly from the (simple) regression of each of the predictors with the dependent variable. Therefore we decide to use the set of contrasts given in Table 13.3.

When analyzing 'Romeo and Juliet' with the orthogonal multiple regression approach, the first step is to compute a set of coefficients of correlation between the dependent variable and the coefficients of each of the contrasts. The quantities that we need are given in Table 13.4. From this table we can compute 3 coefficients of correlation (one per contrast). As an exercise, check that you obtain the same values using the traditional formula of Chapter 12. Note that we introduce a new notation, $r_{Y \cdot C_{a,i}}^2$, for the coefficient of correlation between the dependent

¹ The technical synonym for non-multicollinear is *linearly independent*.

Y	y	y^2	$C_{a,1}$	$C_{a,1}^2$	$C_{a,1} \times y$	$C_{a,2}$	$C_{a,2}^2$	$C_{a,2} \times y$	$C_{a,3}$	$C_{a,3}^2$	$C_{a,3} \times y$
5	0.65	0.42	1	1	0.65	1	1	0.65	0	0	0.00
9	4.65	21.62	1	1	4.65	1	1	4.65	0	0	0.00
8	3.65	13.32	1	1	3.65	1	1	3.65	0	0	0.00
4	-0.35	0.12	1	1	-0.35	1	1	-0.35	0	0	-0.00
9	4.65	21.62	1	1	4.65	1	1	4.65	0	0	0.00
5	0.65	0.42	1	1	0.65	-1	1	-0.65	0	0	0.00
4	-0.35	0.12	1	1	-0.35	-1	1	0.35	0	0	-0.00
3	-1.35	1.82	1	1	-1.35	-1	1	1.35	0	0	-0.00
5	0.65	0.42	1	1	0.65	-1	1	-0.65	0	0	0.00
4	-0.35	0.12	1	1	-0.35	-1	1	0.35	0	0	-0.00
2	-2.35	5.52	-1	1	2.35	0	0	-0.00	1	1	-2.35
4	-0.35	0.12	-1	1	0.35	0	0	-0.00	1	1	-0.35
5	0.65	0.42	-1	1	-0.65	0	0	0.00	1	1	0.65
4	-0.35	0.12	-1	1	0.35	0	0	-0.00	1	1	-0.35
1	-3.35	11.22	-1	1	3.35	0	0	-0.00	1	1	-3.35
3	-1.35	1.82	-1	1	1.35	0	0	-0.00	-1	1	1.35
3	-1.35	1.82	-1	1	1.35	0	0	-0.00	-1	1	1.35
2	-2.35	5.52	-1	1	2.35	0	0	-0.00	-1	1	2.35
4	-0.35	0.12	-1	1	0.35	0	0	-0.00	-1	1	0.35
3	-1.35	1.82	-1	1	1.35	0	0	-0.00	-1	1	1.35
Σ	87	88.55	0	20	25.00	0	10	14.00	0	10	1.00
		SS_Y		$SS_{C_{a,1}}$	$SCP_{Y_{C_{a,1}}}$		$SS_{C_{a,2}}$	$SCP_{Y_{C_{a,2}}}$		$SS_{C_{a,3}}$	$SCP_{Y_{C_{a,3}}}$

Table 13.4 Quantities needed to compute the ANOVA for the ‘Romeo and Juliet’ example as multiple regression with a set of orthogonal contrasts. The following abbreviation is used in the table: $y = Y - M_Y = Y_{as} - M_{..}$.

variable and the set of C_a coefficients of the i th contrast. Likewise, $SS_{C_{a,i}}$ denotes the sums of squares of the C_a coefficients of the i th contrast.

- For ψ_1 :

$$r_{Y \cdot C_{a,1}}^2 = \frac{(SCP_{Y \cdot C_{a,1}})^2}{SS_Y SS_{C_{a,1}}} = \frac{25^2}{88.55 \times 20} \approx .3529 .$$

- For ψ_2 :

$$r_{Y \cdot C_{a,2}}^2 = \frac{(SCP_{Y \cdot C_{a,2}})^2}{SS_Y SS_{C_{a,2}}} = \frac{14^2}{88.55 \times 10} \approx .2213 .$$

- For ψ_3 :

$$r_{Y \cdot C_{a,3}}^2 = \frac{(SCP_{Y \cdot C_{a,3}})^2}{SS_Y SS_{C_{a,3}}} = \frac{1^2}{88.55 \times 10} \approx .0011 .$$

The multiple coefficient of correlation between the set of orthogonal contrasts and the dependent variable is denoted $R_{Y \cdot A}^2$ (it could have been denoted $R_{Y \cdot C_{a,1}C_{a,2}C_{a,3}}^2$ or $R_{Y \cdot \psi_1\psi_2\psi_3}^2$ also). Because the contrasts are all orthogonal to each other, the multiple coefficient of correlation is computed as a simple sum:

$$R_{Y \cdot A}^2 = r_{Y \cdot C_{a,1}}^2 + r_{Y \cdot C_{a,2}}^2 + r_{Y \cdot C_{a,3}}^2 = .3529 + .2213 + .0011 = .5754 \quad (13.1)$$

(which is, indeed, the same value as found previously). The omnibus F test can be computed directly from the multiple coefficient of correlation:

$$F_{Y \cdot A} = \frac{R_{Y \cdot A}^2}{1 - R_{Y \cdot A}^2} \times \frac{df_{\text{residual}}}{df_{\text{regression}}} = \frac{.5754}{1 - .5754} \times \frac{16}{3} = 7.227 , \quad (13.2)$$

and once again this is the same value as previously found.

Each contrast can be evaluated separately with an F test (cf. Chapter 5, Section 5.6, p. 100ff.)

- For the first contrast, the F ratio is:

$$F_{Y \cdot C_{a,1}} = \frac{r_{Y \cdot C_{a,1}}^2}{1 - R_{Y \cdot A}^2} \times \frac{df_{\text{residual}}}{df_{\text{contrast}}} = \frac{.3529}{1 - .5754} \times \frac{16}{1} = 13.30 . \quad (13.3)$$

Under the null hypothesis, this F ratio follows a Fisher distribution with $\nu_1 = 1$ and $\nu_2 = 16$ degrees of freedom. Using a value of $\alpha = .05$, the critical value for F is equal to 4.49. Because the computed value is larger than the critical value, we reject the null hypothesis and conclude that this contrast is significantly different from zero.

Because the first contrast opposes the first two groups to the last two groups, we conclude that the average of the first two groups differs significantly from the average of the last two groups.

If we decide to take into account that we are testing a family of three contrasts, we need to use the Šidák procedure. The critical value is then equal to 7.10 (from Table 5, page 506, in the Appendix). We can still reject the null hypothesis, even taking into account the size of the family of contrasts.

- For the second contrast, the F ratio is:

$$F_{Y \cdot C_{a,2}} = \frac{r_{Y \cdot C_{a,2}}^2}{1 - R_{Y \cdot A}^2} \times \frac{df_{\text{residual}}}{df_{\text{contrast}}} = \frac{.2213}{1 - .5754} \times \frac{16}{1} = 8.34 . \quad (13.4)$$

Under the null hypothesis, this F ratio follows a Fisher distribution with $\nu_1 = 1$ and $\nu_2 = 16$ degrees of freedom. Using a value of $\alpha = .05$, the critical value for F is equal to 4.49. Because the computed value is larger than the critical value, we reject the null hypothesis and conclude that this contrast is significantly different from zero. We will also reject the null hypothesis using the Šidák procedure.

Because the second contrast opposes the first group to the second, we conclude that the first group differs significantly from the second group.

- For the third contrast, the F ratio is:

$$F_{Y \cdot C_{a,3}} = \frac{r_{Y \cdot C_{a,3}}^2}{1 - R_{Y \cdot A}^2} \times \frac{df_{\text{residual}}}{df_{\text{contrast}}} = \frac{.0011}{1 - .5754} \times \frac{16}{1} = 0.04. \quad (13.5)$$

Under the null hypothesis, this F ratio follows a Fisher distribution with $\nu_1 = 1$ and $\nu_2 = 16$ degrees of freedom. Using a value of $\alpha = .05$, the critical value for F is equal to 4.49. Because the computed value is not larger than the critical value, we cannot reject the null hypothesis, and, therefore, we cannot conclude that this contrast is significantly different from zero.

Because the third contrast opposes the third group to the fourth group, this indicates that we cannot conclude that these two groups differ significantly.

We would have reached exactly the same conclusions using the standard contrast analysis described in Chapter 12. So far, we have seen that using orthogonal multiple regression, analysis of variance, or multiple orthogonal contrasts is essentially the same thing. Multiple regression makes it clear that each contrast is a *prediction* of the results from an *a priori* theoretical point of view. Next we will look at the non-orthogonal point of view and we realize that the multiple regression approach gives a more rational point of view than the traditional ANOVA approach described previously in this chapter.

13.3.2 Multiple regression vs classical approach: non-orthogonal contrasts

The classical ANOVA approach to *a priori* non-orthogonal contrasts is to compute a coefficient of correlation $r_{Y \cdot C_a}$ for each contrast and to compute an F ratio as if the contrast was the only contrast of the family, or equivalently as if the family of contrasts was an orthogonal family. The size of the family of contrasts is taken into account only when adjusting the alpha level for multiple comparisons (e.g. using the Šidák correction). A problem with this approach is that it may not correspond exactly to what an experimenter really wants to test. Most of the time, when experimenters are concerned with *a priori* non-orthogonal comparisons, each comparison represents a prediction from a given theory. The goal of the experiment is, in general, to decide which one (or which ones) of the theories can explain the data best. In other words, the experiment is designed to eliminate some theories by showing that they cannot predict what the other theory (or theories) can predict. Therefore experimenters are interested in what each theory can specifically explain. In other words when dealing with *a priori* non-orthogonal comparisons, what the experimenter wants to evaluate are *semi-partial coefficients of correlation* because they express the specific effect of a variable. Within this framework, the multiple regression approach for non-orthogonal predictors fits naturally. The main idea when analyzing non-orthogonal contrasts is simply to consider each contrast as an independent variable in a non-orthogonal multiple regression analyzing the dependent variable.

13.3.3 Multiple regression: non-orthogonal contrasts for Romeo and Juliet

Suppose (for the beauty of the argument) that the ‘Romeo and Juliet’ experiment was, in fact, designed to test three theories. Each of these theories is expressed as a contrast.

1. *Bransford’s* theory implies that only the subjects from the context-before group should be able to integrate the story with their long-term knowledge. Therefore this group should do better than all the other groups, which should perform equivalently. This is equivalent to the following contrast:

$$\psi_1 = 3 \times \mu_1 - 1 \times \mu_2 - 1 \times \mu_3 - 1 \times \mu_4. \quad (13.6)$$

2. The *imagery* theory would predict (at least at the time the experiment was designed) that any concrete context presented during learning will improve learning. Therefore groups 1 and 2 should do better than the other groups. This is equivalent to the following contrast:

$$\psi_2 = 1 \times \mu_1 + 1 \times \mu_2 - 1 \times \mu_3 - 1 \times \mu_4. \quad (13.7)$$

3. The *retrieval cue* theory would predict that the context acts during the retrieval phase (as opposed to Bransford’s theory which states that the context acts during the *encoding* phase. So, the precise timing of the presentation of the context at encoding should be irrelevant). Therefore groups 1 and 3, having been given the context during the encoding session (either immediately before or after hearing the story) should do better than the other groups. This is equivalent to the following contrast:

$$\psi_3 = 1 \times \mu_1 - 1 \times \mu_2 + 1 \times \mu_3 - 1 \times \mu_4. \quad (13.8)$$

Table 13.5 gives the set of non-orthogonal contrasts used for the multiple regression approach. The first step, for analyzing ‘Romeo and Juliet’ using non-orthogonal multiple regression, is to compute a set of coefficients of correlation between the dependent variable and each of the contrasts. The quantities that we need are given in Table 13.6. From this table we can compute three coefficients of correlation (one per contrast). As an exercise, check that you obtain the same values using the traditional formula of Chapter 12.

- For ψ_1 :

$$r_{Y \cdot C_{a,1}}^2 = \frac{(SCP_{Y \cdot C_{a,1}})^2}{SS_Y SS_{C_{a,1}}} = \frac{53^2}{88.55 \times 60} \approx .5287. \quad (13.9)$$

- For ψ_2 :

$$r_{Y \cdot C_{a,2}}^2 = \frac{(SCP_{Y \cdot C_{a,2}})^2}{SS_Y SS_{C_{a,2}}} = \frac{25^2}{88.55 \times 20} \approx .3529. \quad (13.10)$$

Contrast	Experimental groups			
	1	2	3	4
ψ_1	3	−1	−1	−1
ψ_2	1	1	−1	−1
ψ_3	1	−1	1	−1

Table 13.5 A set of non-orthogonal contrasts for analyzing ‘Romeo and Juliet’. The first contrast corresponds to Bransford’s theory. The second contrast corresponds to the imagery theory. The third contrast corresponds to the retrieval cue theory.

<i>Y</i>	<i>y</i>	<i>y</i> ²	<i>C</i> _{<i>a</i>,1}	<i>C</i> ² _{<i>a</i>,1}	<i>C</i> _{<i>a</i>,1} × <i>y</i>	<i>C</i> _{<i>a</i>,2}	<i>C</i> ² _{<i>a</i>,2}	<i>C</i> _{<i>a</i>,2} × <i>y</i>	<i>C</i> _{<i>a</i>,3}	<i>C</i> ² _{<i>a</i>,3}	<i>C</i> _{<i>a</i>,3} × <i>y</i>
5	0.65	0.42	3	9	1.95	1	1	0.65	1	1	0.65
9	4.65	21.62	3	9	13.95	1	1	4.65	1	1	4.65
8	3.65	13.32	3	9	10.95	1	1	3.65	1	1	3.65
4	-0.35	0.12	3	9	-1.05	1	1	-0.35	1	1	-0.35
9	4.65	21.62	3	9	13.95	1	1	4.65	1	1	4.65
5	0.65	0.42	-1	1	-0.65	1	1	0.65	-1	1	-0.65
4	-0.35	0.12	-1	1	0.35	1	1	-0.35	-1	1	0.35
3	-1.35	1.82	-1	1	1.35	1	1	-1.35	-1	1	1.35
5	0.65	0.42	-1	1	-0.65	1	1	0.65	-1	1	-0.65
4	-0.35	0.12	-1	1	0.35	1	1	-0.35	-1	1	0.35
2	-2.35	5.52	-1	1	2.35	-1	1	2.35	1	1	-2.35
4	-0.35	0.12	-1	1	0.35	-1	1	0.35	1	1	-0.35
5	0.65	0.42	-1	1	-0.65	-1	1	-0.65	1	1	0.65
4	-0.35	0.12	-1	1	0.35	-1	1	0.35	1	1	-0.35
1	-3.35	11.22	-1	1	3.35	-1	1	3.35	1	1	-3.35
3	-1.35	1.82	-1	1	1.35	-1	1	1.35	-1	1	1.35
3	-1.35	1.82	-1	1	1.35	-1	1	1.35	-1	1	1.35
2	-2.35	5.52	-1	1	2.35	-1	1	2.35	-1	1	2.35
4	-0.35	0.12	-1	1	0.35	-1	1	0.35	-1	1	0.35
3	-1.35	1.82	-1	1	1.35	-1	1	1.35	-1	1	1.35
87	0.00	88.55	0	60	53.00	0	20	25.00	0	20	15.00
		<i>SS_Y</i>		<i>SS_{C_a,1}</i>	<i>SCP_{YC_a,1}</i>		<i>SS_{C_a,2}</i>	<i>SCP_{YC_a,2}</i>		<i>SS_{C_a,3}</i>	<i>SCP_{YC_a,3}</i>

Table 13.6 Quantities needed to compute the ANOVA for the ‘Romeo and Juliet’ example as multiple regression with a set of non-orthogonal contrasts. The following abbreviation is used in the table: $y = Y - M_Y = Y_{as} - M_{..}$.

- For ψ_3 :

$$r_{Y \cdot C_{a,3}}^2 = \frac{(SCP_{Y \cdot C_{a,3}})^2}{SS_Y SS_{C_{a,3}}} = \frac{15^2}{88.55 \times 20} \approx .1270 . \quad (13.11)$$

The traditional approach would then compute the F ratio for each contrast and evaluate its significance with a Šidák correction. We would obtain the following values:

- For the first contrast, the F ratio is:

$$F_{Y \cdot C_{a,1}} = \frac{r_{Y \cdot C_{a,1}}^2}{1 - R_{Y \cdot A}^2} \times \frac{df_{\text{residual}}}{df_{\text{contrast}}} = \frac{.5287}{1 - .5754} \times \frac{16}{1} = 19.92 . \quad (13.12)$$

- For the second contrast, the F ratio is:

$$F_{Y \cdot C_{a,2}} = \frac{r_{Y \cdot C_{a,2}}^2}{1 - R_{Y \cdot A}^2} \times \frac{df_{\text{residual}}}{df_{\text{contrast}}} = \frac{.3529}{1 - .5754} \times \frac{16}{1} = 13.30 . \quad (13.13)$$

- For the third contrast, the F ratio is:

$$F_{Y \cdot C_{a,3}} = \frac{r_{Y \cdot C_{a,3}}^2}{1 - R_{Y \cdot A}^2} \times \frac{df_{\text{residual}}}{df_{\text{contrast}}} = \frac{.1270}{1 - .5754} \times \frac{16}{1} = 4.79 . \quad (13.14)$$

Under the null hypothesis, each of these F ratios follows a Fisher distribution with $\nu_1 = 1$ and $\nu_2 = 16$ degrees of freedom. Using a value of $\alpha = .05$, the critical value for F is equal to 4.49. Because the computed value of each F ratio is larger than the critical value, we reject the null hypothesis and conclude that each contrast is significantly different from zero. If we use the Šidák correction for multiple comparisons, the critical value (from Table 5, page 506, in the Appendix) is equal to 7.10, and we can reject the null hypothesis only for contrasts 1 and 2.

The multiple regression approach will compute the *semi-partial coefficients of correlation*. Because we are dealing with a set of three contrasts, computing the semi-partial coefficients of correlation is actually a rather long and cumbersome process. In fact, it is so cumbersome that we won't attempt to illustrate it here.² It is a task best accomplished with a statistical computer package. We used such a package to compute the semi-partial coefficients of correlation used below. For each contrast, we give also the value of the F ratio obtained with the formula adapted from Equation 6.28:

$$F_{Y \cdot C_{a,i} | C_{a,k} C_{a,\ell}} = \frac{r_{Y \cdot C_{a,i} | C_{a,k} C_{a,\ell}}^2}{1 - R_{Y \cdot A}^2} \times (df_{\text{residual}}) . \quad (13.15)$$

We obtain the following values:

- For the first contrast:

$$r_{Y \cdot C_{a,1} | C_{a,2} C_{a,3}}^2 = .0954 ,$$

$$F_{Y \cdot C_{a,1} | C_{a,2} C_{a,3}} = \frac{r_{Y \cdot C_{a,1} | C_{a,2} C_{a,3}}^2}{1 - R_{Y \cdot A}^2} \times (df_{\text{residual}}) = \frac{.0954}{1 - .5754} \times 16 = 3.59 .$$

- For the second contrast:

$$r_{Y \cdot C_{a,2} | C_{a,1} C_{a,3}}^2 = .0406 ,$$

$$F_{Y \cdot C_{a,2} | C_{a,1} C_{a,3}} = \frac{r_{Y \cdot C_{a,2} | C_{a,1} C_{a,3}}^2}{1 - R_{Y \cdot A}^2} \times (df_{\text{residual}}) = \frac{.0406}{1 - .5754} \times 16 = 1.53 .$$

² The courageous reader may use the formulas given in Chapter 6.

- For the third contrast:

$$r_{Y \cdot C_{a,3}|C_{a,1}C_{a,2}}^2 = .0011 ,$$

$$F_{Y \cdot C_{a,3}|C_{a,1}C_{a,2}} = \frac{r_{Y \cdot C_{a,3}|C_{a,1}C_{a,2}}^2}{1 - R_{Y \cdot A}^2} \times (df_{\text{residual}}) = \frac{.0011}{1 - .5754} \times 16 = 0.04 .$$

These F ratios follow a Fisher distribution with $\nu_1 = 1$ and $\nu_2 = 16$ degrees of freedom. From the table of critical values in Table 2 of the Appendix on page 499, we find that the critical value is equal to 4.49 for an α level of .05. None of these F ratios is larger than the critical value. Hence we cannot reject the null hypothesis for any of these contrasts! Comparing this conclusion with the conclusions reached from the classical analysis of the contrasts (i.e. using the $r_{Y \cdot C_a}^2$ coefficients, cf. page 265) highlights the drastic differences between these approaches. Combining these two analyses, we would reach the conclusion that each theory can somewhat agree with the data, but that none of them can specifically predict the results better than a combination of the other competing theories (at least this is the conclusion we would reach with this number of subjects per group!).

Chapter summary

13.4 Key notions of the chapter

Below are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

Family of comparisons	Šidák test
Non-orthogonal (non-independent) comparisons	Dunnett test
α per comparison	The classical approach corrects for multiple comparisons.
α per family of comparisons	The modern approach evaluates the <i>specific</i> contribution of each contrast using semi-partial correlation (i.e. multiple non-orthogonal regression).
Šidák or Bonferroni, Boole, Dunn inequality	

13.5 New notations

Below are the new notations introduced in this chapter. Test yourself on their meaning.

$F_{\text{critical Šidák}}$; $F_{\text{critical Dunnett}}$

$$F_{Y \cdot C_{a,i}|C_{a,k}C_{a,\ell}} = \frac{r_{Y \cdot C_{a,i}|C_{a,k}C_{a,\ell}}^2}{1 - R_{Y \cdot A}^2} \times (df_{\text{residual}}) . \quad (13.16)$$

13.6 Key formulas of the chapter

Below are the main formulas introduced in this chapter: try to go through them and understand what they mean.

$$\alpha[PF] \leq 1 - (1 - \alpha[PC])^C$$

$$\alpha[PC] \approx 1 - (1 - \alpha[PF])^{1/C}$$

$$\alpha[PF] \approx C\alpha[PC] = \underbrace{\alpha[PC] + \alpha[PC] + \dots + \alpha[PC]}_{C \text{ times}}$$

13.7 Key questions of the chapter

Below are some questions about the content of this chapter. All the answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- ★ Why do we need to differentiate between α per family and α per comparison?
- ★ When does statistical tradition regard correction of $\alpha[PC]$ as mandatory?
- ★ What are the ways to correct $\alpha[PC]$ and the advantages of each?
- ★ When is the Dunnett test applied?
- ★ When are there differences between the classical approach and the modern approach?
- ★ When are the classical approach and the modern approach equivalent?

14

ANOVA, one factor: *post hoc* or *a posteriori* analyses

14.1 Introduction

Post hoc analyses are performed after the data have been collected, or in other words, after the fact. When looking at the results of an experiment, you can find an unexpected pattern. If this pattern of results suggests some interesting hypothesis, then you want to be sure that it is not a fluke. This is the aim of *post hoc* (also called *a posteriori*) comparisons. The main problem with *post hoc* comparisons involves the size of the family of possible comparisons. In particular, because the number of possible comparisons grows very quickly as a function of A (the number of levels of the independent variable) it is unrealistic to use procedures involving the Šidák or Bonferroni, Boole, Dunn inequalities. Even when the set of contrasts to be performed consists only of 'pairwise' comparisons which are obtained by comparing pairs of means, the number of comparisons becomes too big to make the use of Šidák or Bonferroni, Boole, Dunn inequalities practical. This is illustrated in Tables 14.1 and 14.2.

In both cases (all possible contrasts, or all pairwise comparisons) the Šidák inequality will make the statistical tests so conservative that it will be almost impossible to reject the null hypothesis. In general, the approach used is to try to find a procedure more sensitive (i.e. more powerful) without having an excessive number of 'false alarms'. Several techniques have been proposed to that aim, and a single volume will not suffice to treat them all (cf. however, Miller, 1981; and Hochberg and Tamhane, 1987). Two main cases will be examined here:

- How to evaluate all the possible contrasts. This is known as *Scheffé's* test.
- The specific problem of pairwise comparisons. Here we will see three different tests: *Tukey*, *Newman-Keuls*, and *Duncan*.

A	Number of possible contrasts	$\alpha[PF]$	$\alpha[PC]$
2	1	.050	.050000
3	6	.265	.008512
4	25	.723	.002049
5	90	.990	.000569
6	301	.999	.000170
7	966	≈ 1.000	.000053
8	3025	≈ 1.000	.000017
9	9330	≈ 1.000	.000005
10	28501	≈ 1.000	.000002

Table 14.1 $\alpha[PF]$ and $\alpha[PC]$ as a function of the number of possible contrasts that can be performed for different values of A (number of groups). The number of possible contrasts (i.e. pairwise comparisons) is given by the formula: $1 + \{[(3^A - 1)/2] - 2^A\}$. The column labeled A gives the number of groups, next to it is the number of possible contrasts that can be performed. The column $\alpha[PF]$ gives the probability of making at least one Type I error when each of the contrasts is tested at the .05 level if the Šidák approximation is used: $\alpha[PF] \approx 1 - .95^C$ where C is the number of contrasts. The column $\alpha[PC]$ gives the probability to use to test each contrast (i.e. $\alpha[PC]$) in order to have $\alpha[PF] = .05$, if the Šidák approximation is used: $\alpha[PC] \approx 1 - .95^{1/C}$.

A	Number of pairwise comparisons	$\alpha[PF]$	$\alpha[PC]$
2	1	.0500	.0500
3	3	.1426	.0170
4	6	.2649	.0085
5	10	.4013	.0051
6	15	.5367	.0034
7	21	.6594	.0024
8	28	.7621	.0018
9	36	.8422	.0014
10	45	.9006	.0011

Table 14.2 $\alpha[PF]$ and $\alpha[PC]$ as a function of the number of possible pairwise comparisons that can be performed for different values of A (number of groups). The number of possible contrasts is given by the formula: $\frac{1}{2}A(A - 1)$. The column labeled A gives the number of groups, next to it is the number of possible pairwise comparisons. The column $\alpha[PF]$ gives the probability of making at least one Type I error when each of the contrasts is tested at the .05 level if the Šidák approximation is used: $\alpha[PF] \approx 1 - .95^C$. The column $\alpha[PC]$ gives the probability to use to test each contrast (i.e. $\alpha[PC]$) in order to have $\alpha[PF] = .05$, if the Šidák approximation is used: $\alpha[PC] \approx 1 - .95^{1/C}$.

14.2 Scheffé's test: all possible contrasts

14.2.1 Justification and general idea

When a comparison is decided *a posteriori*, the family of comparisons is composed of all the possible comparisons even if they are not explicitly made. Indeed, because we choose the comparisons to be made *a posteriori*, this implies that we have made implicitly and judged uninteresting all the possible comparisons that have not been made. Hence, whatever the number of comparisons actually performed, the family is composed of all the possible contrasts.

Scheffé's test was devised in order to be able to test all the possible contrasts *a posteriori* while maintaining the overall Type I error level for the family at a reasonable level, as well as trying to have a relatively powerful test. In other words, the Scheffé test is conservative. This happens because a given critical value for the Scheffé test is always larger than the corresponding critical value for other, more powerful, tests. Therefore, when the Scheffé test rejects the null hypothesis, more powerful tests also reject the null hypothesis. Also and conversely, there are cases where the Scheffé test will fail to reject the null hypothesis while more powerful tests will reject it. There is never a case where the most powerful test will not reject H_0 and Scheffé will reject H_0 , in other words, there is no discrepancy possible between Scheffé and the other more powerful tests. So, to sum up: when using Scheffé, the null hypothesis is rejected less often than it should be, which is equivalent to decreasing power and thus increasing the size of β , the probability of making a Type II error. The general idea of Scheffé (1953, 1959) is remarkably simple and intelligent at the same time.

It starts with the obvious observation that the omnibus F ratio tests the null hypothesis that all population means are equal which implies that all possible contrasts are zero. Now, the other point to take into account is that the test on the contrast with the largest sum of squares is equivalent to testing all the other contrasts at once. This is because the failure to reject the null hypothesis for the largest contrast implies obviously the failure to reject the null hypothesis for any smaller contrast. The general idea, then, is to devise a test such that no discrepant statistical decision can occur. By that Scheffé meant to avoid the case for which the omnibus test would fail to reject the null hypothesis, but one *a posteriori* contrast would be declared significant. The simplest way to achieve that aim is to make the test on the largest contrast equivalent to the omnibus test, and then to test each contrast as if it were the largest one. Doing so makes sure that it will be *impossible* to have a discrepant decision.

The largest value for the contrast sum of squares is actually equal to the sum of squares of $\mathcal{A}$. It is obtained for the specific case of C_a values:

$$C_a = (M_{a\cdot} - M_{\cdot\cdot}) .$$

This is, indeed, a contrast because the sum of the $M_{a\cdot} - M_{\cdot\cdot}$ is equal to zero. This contrast cannot be orthogonal to any other contrast, because the sum of squares of two orthogonal contrasts should add up to the sum of squares of $\mathcal{A}$, and that would lead to a contradiction. With this specific set of C_a values, it is shown below that the sum of squares of that contrast is equal to the sum of squares of $\mathcal{A}$.

$$\begin{aligned} SS_{\psi} &= \frac{S(\sum C_a M_{a\cdot})^2}{\sum C_a^2} \\ &= \frac{S[\sum (M_{a\cdot} - M_{\cdot\cdot}) M_{a\cdot}]^2}{\sum (M_{a\cdot} - M_{\cdot\cdot})^2} \quad [\text{substitute } (M_{a\cdot} - M_{\cdot\cdot}) \text{ for } C_a] \quad (14.1) \end{aligned}$$

The form of the numerator of this expression can be changed as follows:

$$\begin{aligned}
 S[\sum (M_{a.} - M_{..})M_{a.}]^2 &= S[\sum (M_{a.}^2 - M_{..}M_{a.})]^2 \quad (\text{multiply}) \\
 &= S(\sum M_{a.}^2 - \sum M_{..}M_{a.})^2 \quad (\text{distribute the } \sum \text{ sign}) \\
 &= S(\sum M_{a.}^2 - M_{..} \sum M_{a.})^2 \quad (M_{..} \text{ is a constant over } a) \\
 &= S(\sum M_{a.}^2 - M_{..}AM_{..})^2 \quad (\text{because } \sum M_{a.} = AM_{..}) \\
 &= S(\sum M_{a.}^2 - AM_{..}^2)^2 \quad (\text{multiply}) \\
 &= S(\sum M_{a.}^2 - 2AM_{..}^2 + AM_{..}^2)^2 \quad (\text{substitute}) \\
 &= S(\sum M_{a.}^2 - 2 \sum M_{..}^2 + \sum M_{..}^2)^2 \\
 &\quad (\text{because } AM_{..}^2 = \sum M_{..}^2) \\
 &= S[\sum (M_{a.}^2 - 2M_{..}^2 + M_{..}^2)]^2 \quad (\sum \text{ is distributive}) \\
 &= S[\sum (M_{a.} - M_{..})^2]^2. \quad (14.2)
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 SS_{\psi} &= \frac{S[\sum (M_{a.} - M_{..})M_{a.}]^2}{\sum (M_{a.} - M_{..})^2} \\
 &= \frac{S[\sum (M_{a.} - M_{..})^2]^2}{\sum (M_{a.} - M_{..})^2} \\
 &= S \sum (M_{a.} - M_{..})^2 \\
 &= SS_A. \quad (14.3)
 \end{aligned}$$

Consequently, the ratio F_{ψ} for this contrast is given by

$$F_{\psi} = \frac{SS_{\psi}}{MS_{S(A)}} = \frac{SS_A}{MS_{S(A)}} = \frac{(A-1)MS_A}{MS_{S(A)}} = (A-1)F_{\text{omnibus}}.$$

Et voilà!

We have now shown that the largest possible sum of squares for a contrast is equal to the experimental sum of squares, and that the F_{ψ} for this contrast is equal to $(A-1)$ times the omnibus F_{omnibus} .

In what follows, we use the notation $F_{\text{critical,omnibus}}$ for the critical value for F_{omnibus} . To have the test of the maximum contrast equivalent to the *omnibus* test, we have to reject the null hypothesis in the same conditions for the contrast and for the F_{omnibus} . Now, to reject H_0 -*omnibus* is equivalent to having:

$$F_{\text{omnibus}} > F_{\text{critical,omnibus}}.$$

Multiplying both terms of the inequality by $(A-1)$, we find

$$(A-1)F_{\text{omnibus}} > (A-1)F_{\text{critical,omnibus}}.$$

But for the maximum contrast

$$(A-1)F_{\text{omnibus}} = F_{\psi}.$$

Hence, H_0 is rejected when

$$F_\psi > (A - 1)F_{\text{omnibus}} .$$

Consequently, according to Scheffé, the critical value to test all the possible contrasts is defined as

$$F_{\text{critical,Scheffé}} = (A - 1)F_{\text{critical,omnibus}}$$

with

$$\nu_1 = A - 1 \quad \text{and} \quad \nu_2 = A(S - 1)$$

degrees of freedom.

14.2.2 An example: Scheffé test for Romeo and Juliet

We will use, once again, Bransford *et al.*'s 'Romeo and Juliet' experiment (cf. Chapter 8, Section 8.7, page 157). The following table gives the different experimental conditions:

Context before	Partial context	Context after	Without context

The error mean square for this experiment is $MS_{S(A)} = 2.35$; and $S = 5$. Here are the values of the experimental means (note that the means have been reordered from the largest to the smallest):

	Context before	Partial context	Context after	Without context
M_a	7.00	4.20	3.20	3.00

Suppose now that the experimenters wanted to test the following contrasts *after* having collected the data.

	Context before	Partial context	Context after	Without context
ψ_1	1	1	1	-3
ψ_2	0	0	1	-1
ψ_3	3	-1	-1	-1
ψ_4	1	-1	0	0

The critical value for $\alpha[PF] = .05$ is given by

$$F_{\text{critical,Scheffé}} = (A - 1)F_{\text{critical,omnibus}} = (4 - 1) \times 3.24 = 9.72$$

with $\nu_1 = A - 1 = 3$ and $\nu_2 = A(S - 1) = 16$.

Warning. Did you find the correct critical value? A common mistake is to use $\nu_1 = 1$ to find the critical value. Make sure that you have used $\nu_1 = A - 1 = 3$.

The results of the Scheffé procedure for the family are given in the following table:

Comparison	SS_ψ	F_ψ	Decision	$\Pr(F_{\text{Scheffé}})$
ψ_1	12.15	5.17	<i>ns</i>	.201599
ψ_2	0.10	0.04	<i>ns</i>	$F < 1$
ψ_3	46.82	19.92	reject H_0	.004019
ψ_4	19.60	8.34	<i>ns</i>	.074800

If you use a computer program that gives the probability associated with a Fisher F ratio, you can use other α levels than the usual .05 or .01. An equivalent way of implementing the Scheffé test is to divide the criterion F_ψ by $(A - 1)$, and to assume that this ratio follows a Fisher distribution with $\nu_1 = A - 1$ and $\nu_2 = A(S - 1)$.

For example, suppose you want to compute the probability associated with the value of the criterion $F_{\psi_3} = 19.92$ obtained for a contrast in a design with $A = 4$ and $S = 5$ using the Scheffé procedure. This is equivalent to computing the ‘corrected’ value

$$F = \frac{F_\psi}{A - 1} = \frac{19.92}{3} = 6.64 ,$$

and to finding its probability assuming it is a Fisher distribution with $\nu_1 = A - 1 = 3$ and $\nu_2 = A(S - 1) = 16$. The probability associated with this value is equal to .004019 as indicated in the previous table.

14.3 Pairwise comparisons

Very often a family of comparisons is actually composed only of the pairwise comparisons. A survey conducted by Jaccard *et al.* (1984) found that this is the case in almost two thirds of the papers using multiple comparisons published in APA journals. Actually, the specific problem of pairwise comparisons is one of the most popular in the statistical literature, and will be only alluded to here (cf. Scheffé, 1953; Winer, 1971; Games and Howell, 1976; Ury, 1976; Tamhane, 1977; Kirk, 1982; Games *et al.*, 1982; Jaccard *et al.*, 1984; Zwirk and Marascuilo, 1984, and many others).

In this section we will just give an overview of the three main methods. The first two use a new distribution called the ‘Studentized range’ or a pseudo- F equivalent distribution called the ‘ F range’. These are the Tukey test and the Newman–Keuls test. The Newman–Keuls test is a sequential test whose aim is to have greater power than the Tukey test. It is the clear winner as far as popularity is concerned.

14.3.1 Tukey test

The Tukey test uses a distribution derived by Gosset. Gosset is, in general, known under the name of Student (yes! Just like Student’s t -test). He used a pseudonym because his main employment was to apply statistical principles to improving beer brewing (at Guinness’s), and he wanted his scientific work to appear independent of this professional activity (some

historians of statistics suggest that he did so in order not to give a hint to competitors of what was going on at the Guinness breweries).

Student derived a distribution called Student's q or the Studentized range. We will use a slightly modified version of his distribution called F_{range} or F_{range} . F_{range} is derived from q by the transformation

$$F_{\text{range}} = \frac{q^2}{2}.$$

The range of two values is the number of scores falling between these two scores including those two scores. For example, suppose we have 4 groups, the range is 4. Another example: the range of the series of values 2, 4, 14, 25, 60 is 5 (there are 5 scores in the series).

14.3.1.1 Digression: what is F_{range} ?

Suppose that several samples are taken from the same population. Call the difference between the smallest mean and the largest mean the 'maximum pairwise difference'. The sum of squares of the contrast corresponding to the 'maximum pairwise difference' is noted

$$SS_{\psi \cdot \text{max}}.$$

Because this is a contrast, the sum of squares is also a mean square noted

$$MS_{\psi \cdot \text{max}}.$$

Now the criterion corresponding to the maximum pairwise difference is given by

$$F_{\psi \cdot \text{max}} = \frac{MS_{\psi \cdot \text{max}}}{MS_{\text{error}}}.$$

Student derived the distribution of that statistical index when the null hypothesis is true, and when the technical assumptions of normality and homoscedasticity hold. F_{range} corresponds to that distribution. This is a set of distributions which depends upon two parameters: the number of degrees of freedom of the denominator (i.e. in the MS_{error}) and the range of the comparison (i.e. the number of samples taken from the population). Table 9 (page 514) in the Appendix gives the critical values of F_{range} as a function of these parameters and of the $\alpha[PF]$ level.

You have probably realized that—at least in principle—the test using F_{range} and the standard analysis of variance both test the same null hypothesis. Most of the time, these two approaches agree about the decision whether to reject or not the null hypothesis. It is possible, however, to find discrepant cases (cf. Winer, 1971), because these two procedures use slightly different information. F_{range} uses only the extreme means, while the analysis of variance uses all the means. In general, the analysis of variance approach is more sensitive than F_{range} . If you look in the table of the critical values for F_{range} , you will see that the critical values get larger as the range increases. The critical value corresponding to the number of experimental groups A is called $F_{\text{critical, Tukey}}$.

When the largest difference is tested using $F_{\text{critical, Tukey}}$ as the critical value, the test gives the correct p value. Hence the value $\alpha[PC]$ given for that comparison is correct. Now, keeping the same critical value for all the other pairwise comparisons will give a conservative test. This is the case because rejecting the null hypothesis for any smallest difference will imply that the largest difference has already been judged significant. This is the same approach for pairwise comparisons that the Scheffé test uses for all possible contrasts.

14.3.1.2 An example: Tukey test for Romeo and Juliet

For an example, we will again use Bransford *et al.*'s Romeo and Juliet. Recall that

$$MS_{S(A)} = 2.35; \quad S = 5,$$

and that the experimental results were:

	Context before	Partial context	Context after	Without context
$M_{a.}$	7.00	4.20	3.20	3.00

The pairwise difference between means can be given in a table:

	$M_1.$	$M_2.$	$M_3.$	$M_4.$
$M_1.$		2.80	3.80	4.00
$M_2.$			1.00	1.20
$M_3.$				0.20

From those differences, the sum of squares can easily be computed. In the particular case of pairwise comparisons, the formula for the computation of the sum of squares of group a and a' reduces to

$$\begin{aligned} SS_{\psi} &= \frac{\widehat{S\psi}^2}{\sum C_a^2} \\ &= \frac{S(M_{a.} - M_{a'.})^2}{2}. \end{aligned} \quad (14.4)$$

For example, to compute the sum of squares of the pairwise comparison opposing group 1 to group 2, the formula gives

$$\begin{aligned} SS_{\psi} &= \frac{\widehat{S\psi}^2}{\sum C_a^2} \\ &= \frac{1}{2} S(M_1. - M_2.)^2 \\ &= \frac{5 \times 2.8^2}{2} = 19.6. \end{aligned} \quad (14.5)$$

The sum of squares can be presented in a table similar to the pairwise differences table.

	$M_1.$	$M_2.$	$M_3.$	$M_4.$
$M_1.$		19.60	36.10	40.00
$M_2.$			2.50	3.60
$M_3.$				0.10

The values for $F_{\text{critical, Tukey}}$ given by the table are:

$$8.20 \quad \text{for} \quad \alpha[PF] = .05$$

$$13.47 \quad \text{for} \quad \alpha[PF] = .01 \quad (14.6)$$

The results of the computation of the different F ratio for the pairwise comparisons are given in the following table where the sign * indicates a difference significant at the .05 level, and ** indicates a difference significant at the .01 level.

	$M_1.$	$M_2.$	$M_3.$	$M_4.$
$M_1.$		8.34*	15.36**	17.02**
$M_2.$			1.06**	1.53**
$M_3.$				0.04**

Tukey's test is clearly a conservative test. Several approaches have been devised in order to produce a more sensitive test. The most popular alternative (but not the 'safest') is the Newman-Keuls test. This is a sequential test which changes the critical value for each comparison by taking the range as being the actual number of means involved in the comparison, and by trying to avoid discrepant statistical decisions.

14.3.2 The Newman-Keuls test

Essentially, the Newman-Keuls test consists of a sequential test in which the critical value depends on the range of each pair of means. To make the explanation easier, we will suppose that the means are ordered from the smallest to the largest. Hence $M_1.$ is the smallest mean, and $M_4.$ is the largest mean.

The Newman-Keuls test starts like the Tukey test: the largest difference between two means is selected. The range of this difference is A . The null hypothesis is tested for that pair of means using F_{range} following exactly the same procedure as for the Tukey test. If the null hypothesis cannot be rejected, then the test stops here, because not rejecting the null hypothesis for the largest difference implies not rejecting the null hypothesis for any other difference.

If the null hypothesis is rejected for the largest difference, then the two differences with a range of $A - 1$ are examined. They will be tested with a critical value of F_{range} selected for a range of $A - 1$. When the null hypothesis cannot be rejected for a given difference, none of the differences embedded in that difference will be tested. If the null hypothesis can be rejected for a difference, then the procedure is re-iterated for a range of $A - 2$. The procedure is iterated until all the differences have been tested or declared non-significant by implication.

This procedure avoids discordant decisions such as deciding that a difference between two means is not significant and at the same time finding significant a difference of two means included in the previous one. An example of a discordant decision is declaring $M_2. - M_5.$ non-significant, but at the same time finding $M_3. - M_4.$ significant (recall that the means are ordered from smallest to largest, $M_1.$ is the smallest, $M_5.$ is the largest). This is because if the smaller difference $M_3. - M_4.$ is declared significantly greater than zero, then the largest difference $M_2. - M_5.$ should also be greater than zero (if you have the impression of being in the 'no context' condition of Bransford *et al.*, do not panic, the picture is coming soon ...).

It requires some experience to determine which comparisons are implied by other comparisons. To make it easier, you can have a look at Figure 14.1 which describes the structure of implication for a set of five means numbered from 1 (the smallest) to 5 (the largest). The pairwise comparisons implied by another comparison are obtained by following the

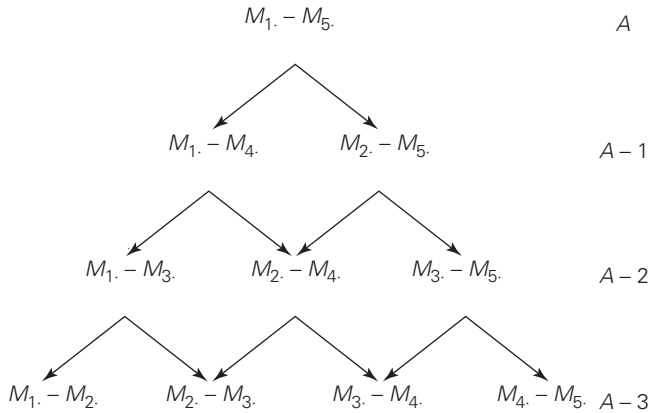


Figure 14.1 Structure of implication of the pairwise comparisons when $A=5$ for Newman-Keuls test. Means are numbered from 1 (the smallest) to 5 (the largest). The pairwise comparisons implied by another comparison are obtained by following the arrows. When the null hypothesis cannot be rejected for one pairwise comparison, then all the comparisons included in it can be crossed out in order to omit them from testing.

arrows. When the null hypothesis cannot be rejected for one pairwise comparison, then all the comparisons included in it can be crossed out in order not to test them.

14.3.3 An example: taking off ...

An example will help to describe the use of Figure 14.1. In an experiment on eyewitness testimony, Loftus and Palmer (1974, already presented in Chapter 10, page 206) tested the influence of the wording of a question on the answers given by eyewitnesses. They presented a film of a multiple car crash to 50 subjects (10 per group). After seeing the film, subjects were asked to answer a number of specific questions. Among these questions, one question about the speed of the car was presented with five different versions:

- 'HIT': About how fast were the cars going when they *hit* each other?
- 'SMASH': About how fast were the cars going when they *smashed* into each other?
- 'COLLIDE': About how fast were the cars going when they *collided* with each other?
- 'BUMP': About how fast were the cars going when they *bumped* into each other?
- 'CONTACT': About how fast were the cars going when they *contacted* each other?

The mean speed estimation by subjects for each version is given in the following table:

	Experimental group				
	Contact M_1	Hit M_2	Bump M_3	Collide M_4	Smash M_5
M_a	30.00	35.00	38.00	41.00	46.00

$$S = 10; \quad MS_{S(A)} = 80.00.$$

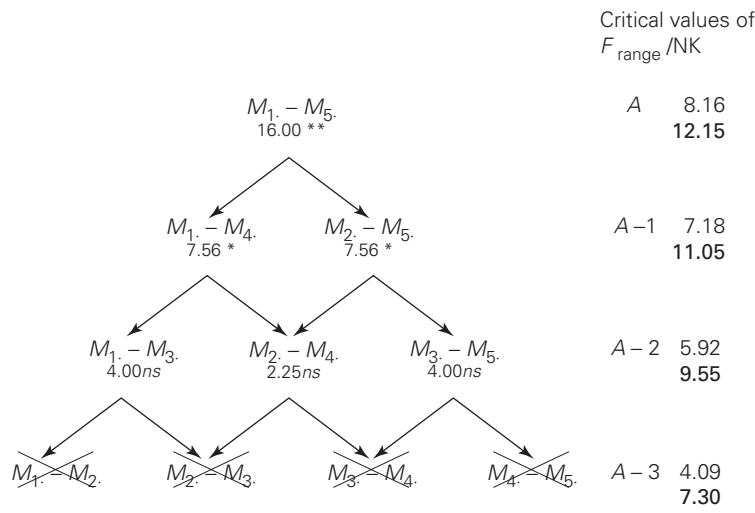


Figure 14.2 Newman–Keuls test for the data from a replication of Loftus and Palmer (1974). The number below each range is the F_{comp} for that range.

The F ratios are given in the following table:

	Experimental group				
	Contact	Hit	Bump	Collide	Smash
Contact	—	1.56	4.00	7.56*	16.00**
Hit		—	0.56	2.25*	7.56*
Bump			—	0.56*	4.00**
Collide				—	11.56**
Smash					—

As illustrated in Figure 14.2, the first step in the Newman–Keuls test is to order the group means from the smallest to the largest. The test starts by evaluating the difference between $M_1.$ and $M_5.$ (i.e. ‘contact’ and ‘smash’). With a range of 5, and for 45 degrees of freedom (actually 40 is the closest value found in the table) the critical value of the F_{range} is 8.16. This value is denoted $F_{\text{range, critical}}(5)$. The ratio $F_{\text{comp}} = 16.00$ is larger than the critical value, and H_0 is rejected for the largest pair. Now we can proceed to test the differences with a range of 4, namely the differences ($M_1. - M_4.$) and ($M_2. - M_5.$). With a range of 4, and for 45 degrees of freedom (actually 40) the critical value of F_{range} is 7.18. Both differences are declared significant at the .05 level. We then proceed to test the comparisons with a range of 3. The differences ($M_1. - M_3.$) and ($M_3. - M_5.$), both with a value of $F_{\text{comp}} = 4.00$, are declared non-significant. Further, the differences ($M_2. - M_4.$) is also declared non-significant ($F_{\text{comp}} = 2.25$). Hence, all comparisons implied by these differences should be crossed out. In other words, we do not test any differences with a range of $A - 3$ [($M_1. - M_2.$), ($M_2. - M_3.$), ($M_3. - M_4.$), and ($M_4. - M_5.$)]. Because we have already tested the comparisons with a range of 3 and found them to be non-significant, any comparisons with a range of 2 will consequently be declared non-significant as they are implied or included in the range of 3 (the test has been performed implicitly).

The results of the tests are often presented with the values of the pairwise difference between the means and with stars indicating the significance level (one star meaning significant at the .05 level, 2 stars meaning significant at the .01 level). The results of the Newman–Keuls test are presented hereunder following that convention. The value at the intersection between a column and a row is obtained by subtracting the column mean from the row mean. For example, the value 5.00 corresponds to the difference between M_4 and M_5 .

	Experimental group				
	M_1 . Contact 30	M_2 . Hit 35	M_3 . Bump 38	M_4 . Collide 41	M_5 . Smash 46
M_1 . = 30 Contact	0.00	5.00 <i>ns</i>	8.00 <i>ns</i>	11.00*	16.00**
M_2 . = 35 Hit		0.00	3.00 <i>ns</i>	6.00 <i>ns</i>	11.00*
M_3 . = 38 Bump			0.00	3.00 <i>ns</i>	8.00 <i>ns</i>
M_4 . = 41 Collide				0.00	5.00 <i>ns</i>
M_5 . = 46 Smash					0.00

There is another conventional way of presenting the results in a table. The means are ordered, and letters are put below them, so that two means with the same letter do not differ significantly.

M_1 . = 30	M_2 . = 35	M_3 . = 38	M_4 . = 41	M_5 . = 46
A	A	A	B	B
		B		

As an exercise, you can try to perform a Newman–Keuls test on the data from Bransford *et al.*; the results are given in Figure 14.3.

14.3.4 Duncan test

The Duncan test follows the same general sequential procedure as the Newman–Keuls test. The only difference is that the critical values come from a different table (Table 10, page 515, in the Appendix). Newman–Keuls uses the F_{range} distribution, Duncan uses the standard Fisher distribution with a Šidák correction for multiple comparisons.

In general, to compare 2 means separated by a range of E , it suffices to correct $\alpha[PC]$ as:

$$\alpha[PC] = 1 - (1 - \alpha[PF])^{1/(E-1)}.$$

This is the Šidák inequality when the order of the comparisons is taken into account.

For example, to compare 4 means ranked from the smallest to the largest with $\alpha[PF] = .01$, $\alpha[PC]$ is given by:

$$\begin{aligned}\alpha[PC] &= 1 - (1 - \alpha[PF])^{1/(E-1)} \\ &= 1 - (1 - .01)^{1/3} = .0033\end{aligned}\quad (14.7)$$

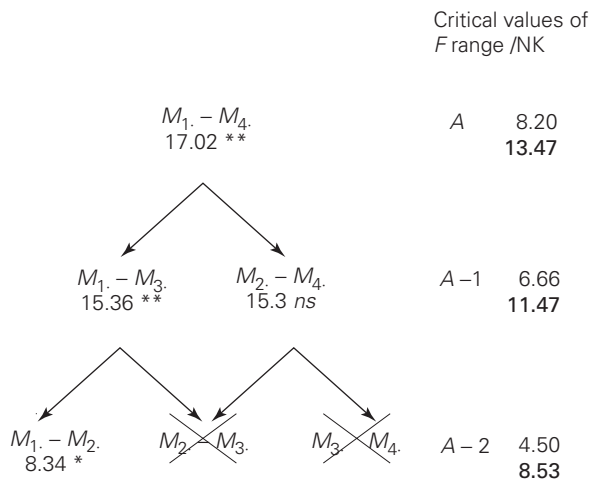


Figure 14.3 Newman–Keuls test for a replication of the Bransford *et al.* ‘Romeo and Juliet’ experiment.

with a range of 3, $\alpha[PC]$ becomes:

$$\alpha[PC] = 1 - (1 - .01)^{1/2} = .0050 .$$

Table 10 in the Appendix on page 515 was computed following this procedure.

Chapter summary

14.4 Key notions of the chapter

Below are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

<i>A posteriori</i> comparison	Pairwise comparisons
Scheffé’s test	Tukey, Newman–Keuls, and Duncan tests
Conservative test	<i>F</i> -range

14.5 New notations

Below are the new notations introduced in this chapter. Test yourself on their meaning.

F_{range}	$F_{\text{critical,Tukey}}$
$F_{\text{critical,Scheffé}}$	$F_{\text{critical,NK}}$

14.6 Key questions of the chapter

Below are two questions about the content of this chapter. Both answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- ★ When using a Scheffé test is it possible to have at the same time a contrast significant and an F -omnibus non-significant ?
- ★ Why is the Newman-Keuls test more powerful than the Tukey test?

15

More on experimental design: multi-factorial designs

15.1 Introduction

In a one-factor (or uni-factorial) design, symbolized $S(A)$, the effect of the independent variable A is determined by holding constant or controlling all other possible factors. Control is often achieved by randomization, which is typically used to assign subjects to the experimental groups. Variations in the dependent variable can therefore be attributed to the independent variable (supposing that independent variables that are confounded, facetious, or perverse are not at play). There are situations, however, where it is best to assess the effects of more than one independent variable to answer specific research questions. We have already seen in Chapter 1 that by increasing the number of independent variables in an experimental design, we increase the external or ecological validity of the experiment. Experimental designs involving more than one factor or independent variable are generally referred to as *multi-factorial designs*.

Using multi-factorial designs makes experiments more realistic. It allows us to study the effects of each independent variable, and, above all, of their interactions. By interaction we mean a result in which *the effects of one independent variable are different, depending on the levels of the other independent variable*. For example, for lists of five nonsense syllables, the independent variable ‘number of times the list is read’ has no effect on the dependent variable ‘number of syllables recalled’ (memorization is perfect after only one time through the list). By contrast, when the list is 20 syllables long there is a definite effect of the independent variable ‘number of times the list is read’ on the dependent variable. (Learning increases with greater practice.) We would say that there is an interaction between the independent variable ‘number of syllables in the list’ and the independent variable ‘number of times the list is read’ (or vice versa—interactions are symmetrical). The effect of one independent variable depends on the level of the other.

The use of factorial designs, as opposed to uni-factorial designs, can often be defended on theoretical grounds. Many experiments have been carried out to 'test' or support one theoretical model or another, and many theories can produce successful explanations for the effects of one independent variable. (We describe this in terms of 'main effects'; each independent variable corresponds to each main effect.) But a theory that successfully predicts an interaction is a more successful theory, because it is generally more difficult to predict an interaction than to predict a main effect.

Rubin *et al.* (1979; cf. Shoben, 1982) give a good example of the use of an interaction between independent variables for theoretical purposes. Rubin *et al.* wished to evaluate certain predictions made by Taft and Foster (1975, 1976), who believed that when subjects hear words with prefixes ('insensitive', 'untiring', 'apathetic', etc.) they decompose them into two parts to understand them: the prefix and the root. Then they check their internal dictionary for the meanings of the components and their combination. If the prefix and the root are acceptable, but not their combination (as in 'intiring'), then the subject searches the internal dictionary for the complete word. This model predicts longer response times for the recognition of words that seem to divide into prefix and root but in reality do not (for example, 'insult') than for words that really do divide that way (for example, 'indirect'). This design involves one independent variable—one factor. (What are the independent variable and dependent variable?) But Rubin *et al.* added another independent variable. In certain cases the words to be recognized were surrounded by composite words consisting of prefix + root; in other cases the surrounding words were not decomposable in that way. (These two cases are levels of what independent variable?) The effect on response time of decomposability of target words was clearly replicated in the context of decomposable words, but not in the context of non-decomposable words. These results led Rubin *et al.* to interpret the subjects' behavior in terms of different strategies. When the subjects expected to encounter composite words, their behavior followed the model of Taft and Foster. By contrast, when they expected to encounter non-decomposable words they tended not to use that strategy. It is interesting to note that the results of this experiment would not be correctly interpretable if there had been only one experimental factor. Moreover, if the authors had used only the condition involving composite words, their results would not have ruled out the operation of a parasite independent variable such as familiarity, or concreteness, etc. But such simple alternative explanations cannot be defended in the face of the interaction that Rubin *et al.* found. This example helps us understand the central place accorded in psychology to multi-factorial design and to the concept of interaction (cf. Tulving, 1983).

15.2 Notation of experimental designs

In order to describe and analyze experiments, the first step is to isolate the factors (i.e. independent variables). When this is done, the second step is to describe the relations between variables in order to obtain the *formula* describing the design. This step is very

important and should be done when designing the experiment (i.e. before starting to collect the data). There are two main reasons for doing so:

- If it is possible to write down the formula of a design then it is always possible to analyze the results. In other words, if one cannot write down the formula of a design, it is highly likely that the results from the experiment will be difficult to analyze or even will not be analyzable at all.
- Statistical programs use the formula of the design in order to perform the data analysis.

There are two fundamental relationships for describing experimental designs: we have already seen (in Chapter 8) the first one which is called the *nesting* relationship [as used to write the design: $\mathcal{S}(\mathcal{A})$]. We shall now introduce the second one: the *crossing* relationship. Together these two relations are used to describe a large number of experimental designs. In fact, any design that can be written with these two relations can be analyzed with the tools given in this book. Before starting to describe these relations, we need to clarify a point of notation about the experimental factors (or independent variables).

Experimental factors are described using upper case calligraphic letters like $\mathcal{A}$, $\mathcal{B}$ with the exception of the letters $\mathcal{E}$ and $\mathcal{Y}$ used for denoting the error and the dependent variable respectively. The letter $\mathcal{X}$ is used also to denote an arbitrary independent variable. The letter $\mathcal{S}$ is used to denote a specific factor, namely the subject factor. This factor should always be present when describing an experimental design.¹

The levels of a factor are denoted using lower case letters with the corresponding number of the level used as a subscript. For example the second level of factor $\mathcal{A}$ is denoted a_2 . The total number of levels of a factor is denoted by the same letter as the factor but in italic. For example, the total number of levels of $\mathcal{A}$ is A .

15.2.1 Nested factors

Recall that a factor is nested in a second factor *when each level of the first factor appears only in a single level of the second*. For example, suppose we select three experimental groups from one school, and three groups from another school. Then each level of the factor ‘group’ appears in only a given school, and the factor ‘group’ is nested in the factor ‘school’. When $\mathcal{A}$ is nested in $\mathcal{B}$ we write $\mathcal{A}(\mathcal{B})$, read ‘ $\mathcal{A}$ nested in (or under) $\mathcal{B}$ ’. $\mathcal{A}$ is the nested factor, and $\mathcal{B}$ the nesting factor. Incidentally, you will sometimes see other notations such as $\mathcal{A} < \mathcal{B} >$, $\mathcal{A}[\mathcal{B}]$, or even $\mathcal{A}/\mathcal{B}$.

One factor can be nested in a factor that is itself nested in a third factor. For example, if we test ten subjects in each of three schools in each of two cities, then ‘subjects’ is nested in ‘school’ which in turn is nested in ‘city’. This design would be written $\mathcal{S}(\mathcal{A}(\mathcal{B}))$, where $\mathcal{A}$ is the factor ‘school’ and $\mathcal{B}$ the factor ‘city’. When one factor is nested in a factor that is itself nested in another factor, the first factor is nested in both of the others. (Why? If the answer is not immediately evident, note that in the previous example a given subject appears in only one experimental group, which itself appears in only one school.) More formally, we would say that the nesting relationship is *transitive* for the designs described in this book.

¹ This is true insofar as we always use subjects in psychology but the subject factor does not always represent human subjects—it could be animals, words, houses, cities, etc.

15.2.2 Crossed factors

Two factors are completely crossed when each level of one appears in conjunction with each level of the other. For example, in an experiment on learning, $\mathcal{A}$ represents ‘time of day’ with three levels (morning, afternoon, evening), and $\mathcal{B}$ the factor ‘reward’ with two levels (without and with reward). $\mathcal{A}$ and $\mathcal{B}$ are crossed if the experiment includes the following conditions:

$$\begin{array}{ll}
 a_1, b_1 : & \text{morning, without reward} & a_1, b_2 : & \text{morning, with reward} \\
 a_2, b_1 : & \text{afternoon, without reward} & a_2, b_2 : & \text{afternoon, with reward} \\
 a_3, b_1 : & \text{evening, without reward} & a_3, b_2 : & \text{evening, with reward}
 \end{array}
 \tag{15.1}$$

The crossing relationship is symbolized by the sign $\times$. Thus $\mathcal{A}$ crossed with $\mathcal{B}$ is written $\mathcal{A} \times \mathcal{B}$. The number of levels of $\mathcal{A} \times \mathcal{B}$ equals the number of levels of $\mathcal{A}$ times the number of levels of $\mathcal{B}$, where the sign $\times$ means multiply. Note that the order of writing $\mathcal{A}$ and $\mathcal{B}$ doesn’t matter: $\mathcal{A} \times \mathcal{B}$ defines the same set of levels as $\mathcal{B} \times \mathcal{A}$.

Clearly we could cross several factors. Suppose, for example, we added another factor to the previous design: ‘type of material’, with two levels (verbal and pictorial). Then the crossing of all three factors $\mathcal{A} \times \mathcal{B} \times \mathcal{C}$ would give us $3 \times 2 \times 2 = 12$ experimental conditions:

$$\begin{array}{cccc}
 a_1, b_1, c_1 & a_1, b_1, c_2 & a_1, b_2, c_1 & a_1, b_2, c_2 \\
 a_2, b_1, c_1 & a_2, b_1, c_2 & a_2, b_2, c_1 & a_2, b_2, c_2 \\
 a_3, b_1, c_1 & a_3, b_1, c_2 & a_3, b_2, c_1 & a_3, b_2, c_2
 \end{array}$$

Special Case: The factor ‘subjects’ ($\mathcal{S}$) can be crossed with one or more other factors. We then say that the experimental design involves *repeated measures*. We could speak also of *matched*, or *related* or even *yoked* samples. (Strictly speaking, the term repeated is used when $\mathcal{S}$ is crossed with *all* the experimental factors.) In the preceding example, if each subject serves in all 12 conditions, the subject factor $\mathcal{S}$ would be crossed with the whole set of other factors, and the design would be written $\mathcal{S} \times \mathcal{A} \times \mathcal{B} \times \mathcal{C}$ (read ‘ $\mathcal{S}$ cross $\mathcal{A}$ cross $\mathcal{B}$ cross $\mathcal{C}$ ’ or ‘ $\mathcal{S}$ times $\mathcal{A}$ times $\mathcal{B}$ times $\mathcal{C}$ ’). Note the conventions of putting the $\mathcal{S}$ factor first, and the others in alphabetical order.

15.3 Writing down experimental designs

To write the formula for a design, you start by identifying the factors and representing them by letters. Then you specify the relations of crossing and nesting among the factors.

15.3.1 Some examples

1. We want to evaluate the effect of the sex of the subjects on their performance in a foot race. We take two groups, the first composed of ten boys and the second of ten girls. We measure the time each subject takes to run 500 meters (the dependent variable). We have two factors. The first (trivially) is the factor ‘subjects’ represented by $\mathcal{S}$. The second is the sex of the subjects, with two levels (boys and girls) and denoted by $\mathcal{A}$. A given subject only appears under one level of $\mathcal{A}$ (that is, each subject is either a boy or a girl), and so $\mathcal{S}$ is nested in $\mathcal{A}$. The design is thus symbolized $\mathcal{S}(\mathcal{A})$, read ‘ $\mathcal{S}$ nested in $\mathcal{A}$,’ or just ‘ $\mathcal{S}$ in $\mathcal{A}$.’

In this example, like almost every case in this book, each group contains the same number of observations. You would say that the experimental design is *balanced*. Because of this balance in number of observations the design can be symbolized ' $S_{10}(\mathcal{A}_2)$ '.

When, as in this case, the factor $\mathcal{S}$ is nested in all the other experimental factors, it can be omitted from the description of the design. Thus $S(\mathcal{A})$ can be written $\mathcal{A}$, $S(\mathcal{A} \times \mathcal{B})$ can be written $\mathcal{A} \times \mathcal{B}$, and $S(\mathcal{A}(\mathcal{B}))$ can be written $\mathcal{A}(\mathcal{B})$, etc. These designs are called 'designs with completely independent measures'.

2. We show ten different faces to each of 20 students. We ask each student to rate their liking for each face. (This rating is the dependent variable.) Here we have two factors: 'subjects' $\mathcal{S}$ and 'faces' $\mathcal{A}$. Each subject rates all ten faces, and so $\mathcal{S}$ and $\mathcal{A}$ are crossed. We symbolize the design $\mathcal{S} \times \mathcal{A}$ or $S_{20} \times \mathcal{A}_{10}$.

3. Three groups of ten children each, of ages 8, 9, and 10, learn four lists of words. This experiment has three factors: 'subjects' $\mathcal{S}$, 'age' $\mathcal{A}$, and 'list to learn' $\mathcal{B}$. Each child is found in only one age group, and thus $\mathcal{S}$ is nested in $\mathcal{A}$. Each child learns all four lists, and so $\mathcal{S}$ is crossed with $\mathcal{B}$, and $\mathcal{A}$ is crossed with $\mathcal{B}$ as well. The design is written $S(\mathcal{A}) \times \mathcal{B}$.

4. In each of five cities we select three schools. In each school 10 third-graders and 10 sixth-graders play one or the other of two versions of a new video game. The dependent variable is the time they spend playing the game. We have five factors: 'subjects' $\mathcal{S}$, 'class' $\mathcal{A}$, 'schools' $\mathcal{B}$, 'cities' $\mathcal{C}$, and 'versions of the game' $\mathcal{D}$. Subjects are nested in classes which are nested in schools which are nested in cities. Subjects are nested in game versions, since each child plays only one version. But version is crossed with cities, schools, and classes. We write this design $S(\mathcal{A}(\mathcal{B}(\mathcal{C})) \times \mathcal{D})$ or $S_{10}(\mathcal{A}_2(\mathcal{B}_3(\mathcal{C}_5)) \times \mathcal{D}_2)$.

Notice that the formula is read from left to right. A given factor is nested in all the factors contained between a parenthesis that opens immediately to its right and the corresponding closing parenthesis. Sometimes what is contained within the parentheses is called the *nest* of the factor. Thus the nest of factor $\mathcal{S}$ includes factors $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$; the nest of factor $\mathcal{A}$ includes factors $\mathcal{B}$ and $\mathcal{C}$. A nested factor is crossed with factors outside the nest if a $\times$ separates them. $\mathcal{A}$ is crossed with $\mathcal{D}$; $\mathcal{B}$ is crossed with $\mathcal{D}$; but $\mathcal{S}$ is not crossed with $\mathcal{D}$.

If each child were to play with both versions of the game, then the experimental design would be written $S(\mathcal{A}(\mathcal{B}(\mathcal{C}))) \times \mathcal{D}$.

5. 120 students are each exposed to one of three prose passages (science fiction, popular science, or statistics). Half the subjects read the text in question, while half listen to a recording of it. Each student is exposed to only one text, which is presented four times. In half the cases the experimenter tells the subjects that the experiment is concerned with word-for-word memory for the text, while the other half of the subjects are told that it concerns memory for the meaning. After each presentation subjects are asked to write down what they recall from the text, and those responses are scored for correct content (the dependent variable). There are five factors: 'subjects' $\mathcal{S}$, 'type of text' $\mathcal{A}$, 'instructions' $\mathcal{B}$ (word-for-word versus meaning), 'presentation' $\mathcal{C}$ (reading versus listening), and 'order of presentation' $\mathcal{D}$ (first through fourth time through the text). The design is written $S(\mathcal{A} \times \mathcal{B} \times \mathcal{C}) \times \mathcal{D}$.

15.4 Basic experimental designs

Certain experimental designs can be thought of as fundamental designs from which more complex designs can be constructed. We have already examined in the first part of this book the statistical analysis of experiments involving a single independent variable, and in which

subjects contribute to only one observation apiece (that is, they serve in only one group in the design). We described these designs as *one-factor independent measures designs*, or *one-factor between-subjects designs*, and we saw that following the convention indicated above they can be written as $S(\mathcal{A})$ or simply $\mathcal{A}$.

In the remainder of this book we shall start by first examining designs with two independent variables in which the observations are independent (that is, different subjects serve in the different groups). Those are written $S(\mathcal{A} \times \mathcal{B})$, or simply $\mathcal{A} \times \mathcal{B}$. The letter $\mathcal{A}$ represents the first independent variable and $\mathcal{B}$ the second. Such designs are described as *two-factor independent measures designs*, or *two-factor between-subjects designs*.

Then we shall consider experiments with one independent variable, but in which the *same subjects* serve in all the different experimental groups. We describe these designs as *one-factor repeated measures designs*, or *one-factor within-subjects designs* and we write them $S \times \mathcal{A}$; that is, subjects are *crossed* with conditions.

After that we shall examine the case of two independent variables in which the same subjects serve in all the experimental conditions of the experiment (i.e. the conditions defined by the crossing of $\mathcal{A}$ and $\mathcal{B}$). These designs are referred to as *two-factors repeated measures designs*, or *two-factors within-subjects designs*. They are written $S \times \mathcal{A} \times \mathcal{B}$.

Finally, when there are two independent variables, we may want to repeat the measures for one independent variable but not for the other. For example, suppose one independent variable is ‘age of subjects’ $\mathcal{A}$ and the other is ‘task difficulty’ $\mathcal{B}$. Each independent variable has only two levels: the subjects are ‘young’ or ‘old’, and the tasks are ‘difficult’ or ‘easy’, giving us four experimental conditions. We can, if we wish, use the same subjects in the two conditions of $\mathcal{B}$, ‘difficult’ and ‘easy’, but we cannot have a subject serve in both the ‘young’ and the ‘old’ conditions of $\mathcal{A}$ simultaneously. We say that the measurements are only partially repeated, or repeated on one factor. These designs are described as $S(\mathcal{A}) \times \mathcal{B}$, read ‘ S nested in $\mathcal{A}$, by $\mathcal{B}$ ’. We sometimes also say that $\mathcal{A}$ is a *between-subjects* factor and that $\mathcal{B}$ is a *within-subjects* factor and term this design a *mixed* design. In some literature on statistics these designs are given the evocative name *split-plot*.² (Note that the variable in parentheses in the design is the one for which the measurements are not repeated. Did you notice that variable $\mathcal{A}$ here is a classificatory or tag variable?)

15.5 Control factors and factors of interest

Among experiments that bring into play several independent variables, it is possible to distinguish two types of factors: *factors of interest* and *control factors*. The experiment is designed to disclose the effects of the factors of interest on the dependent variable. *Control factors* are included in the experiment to increase its effectiveness, to take into explicit account potential sources of error, or because we suspect that they affect the dependent variable but we think that their effect is secondary or uninteresting.

For example, suppose we wish to study the effect of the length of phrases on children’s comprehension of text. Specifically, we believe that short phrases will be better understood than long phrases. We construct two texts, one written with short phrases and the other with

² Sometimes misspelled—even in the best books—as ‘split-splot’ (as for example, Abdi, 1987; or Chanquoy, 2005).

long phrases. Forty boys and girls of age 10 read one of the two texts, and respond to a comprehension test. The grade on the test is the dependent variable. The children come from five different schools. We use two mixed classes from each school. The design includes five factors: A 'phrase length', B 'sex', C 'class within school', D 'school', and S 'subjects'. The design is written $S(A \times B \times C(D))$.

This experiment is designed to gather evidence concerning the effect of phrase length on the understanding of texts. *Phrase length*, therefore, is the factor of interest. The other factors in the design are control factors, also called 'secondary factors' or 'pseudo-factors'. Taking them into account helps us refine the experimental results. For example, we might find that the students of one school (in general) understand the texts less well than students of another school, but that the advantage of short phrases is the same (relatively speaking) in both schools. In that case, taking the factor 'school' into account helps bring out the effects of the factor 'phrase length' more clearly.

A problem that sometimes arises is that of the *interaction* of a control factor with a factor of interest. In this example, suppose that the children in one school understand texts with short phrases better than texts with long phrases, but that in another school the reverse is the case. Thus the results in terms of the factor of interest depend on the level of one of the control factors—the two factors *interact*. Such a result—often an unpleasant surprise—is always difficult to interpret. This is all the more true because control factors are often classificatory variables. On this basis we might suspect that the interaction arises from the confounding of some other variable with the control variable(s). However, this type of situation can also lead to an unexpected discovery when the variable(s) responsible for the interaction are brought to light. Among the most frequently encountered control factors are:

- The factor 'groups' when several groups are used in the same condition.
- The factor 'experimenter' when several experimenters conduct different sessions in the experiment.
- The factor 'testing order' for different orders in which different groups perform the tasks.
- The factor 'time of day'.
- The factor 'subject', and all those factors describing characteristics of subjects (sex, age, religion, ethnic group, etc.).

Since these control factors are in general of secondary importance, we could equally well eliminate them by maintaining them at constant levels throughout the experiment. For example, by using just one experimenter we could eliminate the need to account for the factor 'experimenter', and in testing children from just one school we could eliminate the effect of 'school', etc. Nevertheless, we thereby run the risk that the effect we do observe (or fail to observe) might have been linked to the level of the control factor we included (or failed to include). Therefore, taking explicit account of control factors (as opposed to holding them constant) increases the *ecological validity* of the experiment; that is, it improves its *generalizability*.

Let us also note here that the distinction between control factors and factors of interest is imposed by the experimenter as a function of the aims he or she has in mind. This distinction is not absolute. One experiment's control factors may be another experiment's factors of interest.

Chapter summary

15.6 Key notions of the chapter

Below are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

One versus several independent variables.	Independent measures or between-subjects designs.
Interaction among independent variables.	Repeated measures or within-subject designs.
Crossed and nested relationships between factors.	Mixed or split-plot designs. Control factors and factors of interest.

15.7 Key questions of the chapter

Below are some questions about the content of this chapter. All the answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- ★ What are the advantages of using a multi-factorial design as opposed to a uni-factorial design?
- ★ What is the difference between a repeated (within-subjects) and an independent (between-subject) design?
- ★ Why would you use control factors in an experiment?

16

ANOVA, two factors: $A \times B$ or $S(A \times B)$

16.1 Introduction

Until now we have been considering experimental designs involving only one factor (i.e. one independent variable). However, it often (perhaps usually) happens that the questions posed in psychology require the assessment of the effects of more than one factor at a time. For example, let us take a phenomenon observed by Treisman (1986). In Figure 16.1A the Q is easy to spot amidst the Os, but in Figure 16.1B the T is difficult to spot among the Ls. According to Treisman's theory, letters distinguished by different features (such as the slanted line possessed by the Q that the Os lack) can be sorted out rapidly by an automatic 'pre-attentive' process, whereas letters sharing features (such as the vertical and horizontal lines of Ts and Ls) must be sorted out by a more effortful process requiring attention to focus on them one by one. This theory makes the prediction that adding more distractor letters to the display will slow subjects down in responding when the target letter (T) shares all its features with the distractors (Ls), but not when the target (Q) has a feature distinguishing it from the distractors (Os). Thus responses should be slower when the subject is picking out the T in Figure 16.1B than in Figure 16.1D, while it should take just about as long to pick out the Q in Figure 16.1A as in Figure 16.1C. However, to verify this with an experiment we need a design in which we can manipulate two independent variables at once: the degree of shared features of targets and distractors, and the number of distractors in the display.

Note that in the previous example one of the independent variables (number of distractors) has a different effect depending on the value of the other independent variable (shared features). Increasing the number of distractors slows target detection when the target and distractors share all their features, but not otherwise. This is what is called an *interaction*, and we will consider interactions at some length after we have introduced the basic framework for designs with two independent variables.

As we have seen in the previous chapter, designs with several independent variables are often called *factorial* or *multi-factorial* designs. Factorial designs

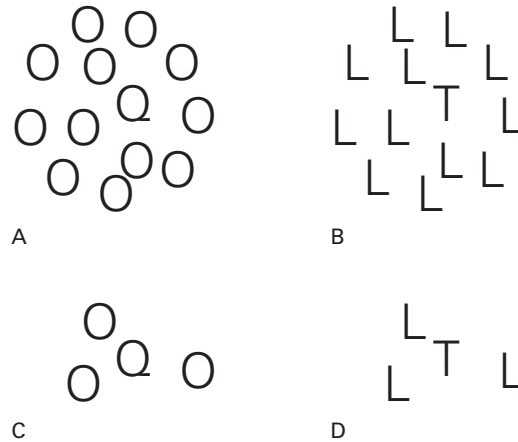


Figure 16.1 Illustration of Treisman (1986) experiment.

are used very frequently in psychology as well as in other sciences. They are preferred to uni-factorial designs for several reasons:

- Multi-factorial designs are better tests of theories. As we just saw in the case of Treisman's theory of attention, the theory required us to assess the effects of the two factors of shared features and number of distractors simultaneously.
- Multi-factorial designs have, potentially, a greater ecological validity. Few factors that affect human behavior operate in isolation, and most operate in different ways depending on the real-world context the experiment is modeling. To take a simple example, the intelligibility of speech over a telephone varies with both the amount of background noise and the filtering effects of the phone system on various frequencies of sound wave. Therefore we might wish to assess the effects of those two factors in the same study (Miller and Nicely, 1965).
- Multi-factorial designs can explicitly take into account the effect of an 'interfering' independent variable. Therefore, they constitute a way of controlling parasite independent variables (see discussion about control factors in the previous chapter).

Designs with two factors in which subjects are measured once—in just one combination of conditions—are symbolized by the notation $S(\mathcal{A} \times \mathcal{B})$, where $\mathcal{A}$ represents the first independent variable and $\mathcal{B}$ the second one. The factor S is nested in the complete design and is sometimes omitted from the notation of the design by authors who, therefore, prefer the notation $\mathcal{A} \times \mathcal{B}$ instead. Recall that 'nested' means that different groups of subjects experience different combinations of the $\mathcal{A}$ and $\mathcal{B}$ variables or, to put it in an equivalent way, that a given subject is exposed to *only one* experimental condition.

Actually, designs with two factors are built from the one-factor designs. The concepts and techniques used previously remain valid. There is only one new

notion (but a very important one): the *interaction* between the independent variables.

It is very often simpler and more practical to have a *balanced* design (i.e. the same number of subjects in each experimental group) than to allow different sizes of groups. In the case of multi-factorial designs, this requirement is even more important than in the case of uni-factorial designs. As we shall see later, this condition is necessary for the ‘slicing’ of the experimental sum of squares to work correctly. As the notation $\mathcal{S}(\mathcal{A} \times \mathcal{B})$ indicates, each subject is assigned to one and only one experimental condition. For the design to provide for valid generalization of the results, the assignment of subjects to experimental groups should be random.

16.2 Organization of a two-factor design: $\mathcal{A} \times \mathcal{B}$

In a uni-factorial design, each experimental condition corresponds to a level of the independent variable. For a two-factor design, each experimental condition corresponds to a specific *combination* of levels of the two independent variables. For example, suppose that we are dealing with a design in which factor $\mathcal{A}$ has four levels and factor $\mathcal{B}$ has three levels. ‘Crossing’ (i.e. multiplying) these two factors gives $4 \times 3 = 12$ experimental groups (if you remember set theory: the groups are defined by the Cartesian product of the levels of the two factors). The levels of $\mathcal{A}$ are denoted a_1, a_2, a_3 and a_4 . The levels of $\mathcal{B}$ are denoted b_1, b_2 and b_3 . The experimental groups are indicated by combining levels of $\mathcal{A}$ and $\mathcal{B}$: $a_1b_1, a_1b_2, a_1b_3, a_2b_1, a_2b_2, \text{etc.}$

The pattern of experimental groups can be displayed in a table such as the following:

Levels of $\mathcal{B}$	Levels of $\mathcal{A}$			
	a_1	a_2	a_3	a_4
b_1	a_1b_1	a_2b_1	a_3b_1	a_4b_1
b_2	a_1b_2	a_2b_2	a_3b_2	a_4b_2
b_3	a_1b_3	a_2b_3	a_3b_3	a_4b_3

Experimental designs are often named from the number of levels of their factors. Here the design is obtained by crossing the 4-level factor $\mathcal{A}$ with the 3-level factor $\mathcal{B}$. It can be symbolized as a 4×3 design (read ‘4 by 3 design’). Obviously, the number of experimental groups is obtained by computing the multiplication indicated by the $\times$ sign.

To be a bit more concrete, consider a replication of an experiment by Tulving and Pearlstone (1965). Subjects were asked to learn lists of 12, 24 or 48 words (factor $\mathcal{A}$ with three levels). These words can be put in pairs by categories (for example, apple and orange can be grouped as ‘fruits’). Subjects were asked to learn these words, and the category name was shown at the same time as the words were presented. Subjects were told that they did not have to learn the category names. After a very short time, subjects were asked to recall the words. At that time, half of the subjects were given the list of the category names, and the other half had to recall the words without the list of categories (factor $\mathcal{B}$ with two levels). The dependent variable is the number of words recalled by each subject. We will look at the results later on.

The six experimental conditions used in this experiment are summarized in the following table.

Levels of $\mathcal{B}$	Levels of $\mathcal{A}$		
	a_1 12 words	a_2 24 words	a_3 48 words
b_1 Cued recall	$a_1 b_1$ 12 words cued recall	$a_2 b_1$ 24 words cued recall	$a_3 b_1$ 48 words cued recall
b_2 Free recall	$a_1 b_2$ 12 words free recall	$a_2 b_2$ 24 words free recall	$a_3 b_2$ 48 words free recall

Another way to look at this experiment is to interpret it as two one-factor designs with three levels each (obviously, we could also consider this experiment as being made of three two-level designs). Each of those designs corresponds to a level of the independent variable $\mathcal{B}$. This is illustrated in the following table.

Sub-design 1: Cued recall			
Cued recall b_1	12 words a_1	24 words a_2	48 words a_3

Sub-design 2: Free recall			
Free recall b_2	12 words a_1	24 words a_2	48 words a_3

16.2.1 Notations

The notation for a two-factor design is a straightforward generalization of the notations used previously. The only important change is that three subscripts are now needed to index a given score.

- The number of levels of factor $\mathcal{A}$ and $\mathcal{B}$ are denoted, respectively by A and B .
- The number of subjects per group is denoted S (a group is defined as one ab combination, there are $A \times B$ groups).
- The score of subject s in the group ab is denoted Y_{abs} .
- The implicit sum system used previously generalizes to the present case. Specifically:

$$Y_{ab.} = \sum_{s=1}^S Y_{abs} : \text{sum of the scores for experimental group } ab$$

$$M_{ab.} = \frac{1}{S} Y_{ab.} : \text{mean for experimental group } ab$$

$$Y_{a..} = \sum_{s=1}^S \sum_{b=1}^B Y_{abs} : \text{sum of scores in condition } a$$

$$M_{a..} = \frac{1}{BS} Y_{a..} : \text{mean for condition } a$$

$$Y_{.b.} = \sum_{s=1}^S \sum_{a=1}^A Y_{abs} : \text{sum of scores in condition } b$$

$$M_{.b.} = \frac{1}{AS} Y_{.b.} : \text{mean for condition } b$$

$$Y_{...} = \sum_{s=1}^S \sum_{a=1}^A \sum_{b=1}^B Y_{abs} : \text{sum of the scores for all the}$$

subjects of all the groups

$$M_{...} = \frac{1}{ABS} Y_{...} : \text{grand mean}$$

16.3 Main effects and interaction

Remember that in a one-factor design the total sum of squares is partitioned into two parts:

- The first part reflects the effect of the independent variable, as it creates differences *between* experimental groups. It corresponds to the source of variation $\mathcal{A}$.
- The second part reflects the experimental error *within* groups. It corresponds to the source of variation $S(\mathcal{A})$.

In a one-factor design, the experimental sum of squares and the sum of squares of $\mathcal{A}$ are synonymous. In a two-factor design, the experimental sum of squares represents the variability of all the groups as if they were representing only one factor. The aim of the analysis, now, is to find out what part the independent variables play in that sum of squares. This is found by subdividing the between-group sum of squares into three parts:

- One reflecting the effect of the independent variable $\mathcal{A}$.
- One reflecting the effect of the independent variable $\mathcal{B}$.
- One reflecting the effect of the interaction between the two variables $\mathcal{A}$ and $\mathcal{B}$. This source of variation is denoted $\mathcal{AB}$.

We will say, likewise, that the experimental sum of squares is partitioned into three sources of variation: two main effects (i.e. $\mathcal{A}$ and $\mathcal{B}$) and the interaction $\mathcal{AB}$. We will see later on how to make that partition. But first, let us detail the notions of main effect and interaction.

16.3.1 Main effects

In order to introduce the notions of main effects and interaction it is easier to start with an example. Table 16.1 presents the results of a hypothetical experiment involving a 3×3 design. The independent variable $\mathcal{A}$, and the independent variable $\mathcal{B}$ have three levels each, so crossing them defines 9 experimental groups. The mean of each experimental group (i.e. the $M_{ab.}$ values) is given in the body of the table. The column and the row margins give the

Factor $\mathcal{B}$	Factor $\mathcal{A}$			Means
	a_1	a_2	a_3	
b_1	7.5	6.0	4.5	6.0
b_2	7.0	5.5	4.0	5.5
b_3	3.5	2.0	0.5	2.0
Means	6.0	4.5	3.0	4.5

Table 16.1 Means of a fictitious experiment with main effects and no interaction (compare with Table 16.2).

values of the means for each level of $\mathcal{A}$ and $\mathcal{B}$. Because the means $M_{a..}$ values and $M_{.b.}$ values are in the ‘margins’ of the table, they are often referred to as the *marginal* means.

The column and row margins can be computed by averaging the values in the rows and columns they represent. For example, the mean for the third level of $\mathcal{B}$ is computed as $M_{.3.} = 2.0 = (3.5 + 2.0 + 0.5)/3$. The *main effect* of $\mathcal{A}$ corresponds to the set of numbers along the bottom margin. Notice that the effect of $\mathcal{A}$ is evaluated by averaging over all levels of $\mathcal{B}$, which is equivalent to ignoring the effect of $\mathcal{B}$ when evaluating $\mathcal{A}$. So, the effect of $\mathcal{A}$ is seen through the set of means:

$$M_{1..} = 6.0, \quad M_{2..} = 4.5, \quad M_{3..} = 3.0.$$

And the main effect of $\mathcal{B}$ is seen through the set of means:

$$M_{.1.} = 6.0, \quad M_{.2.} = 5.5, \quad M_{.3.} = 2.0.$$

16.3.2 Interaction

When we compute the marginal means to show the main effect of one independent variable, the effect of the other independent variable is ignored. In contrast, the interaction reveals the conjoint effects of both independent variables. This very important notion is specific to multi-factorial designs. The two following examples introduce this notion. The first involves two variables without any interaction, whereas the second one shows an interaction between the independent variables.

16.3.3 Example without interaction

In general, it is a good idea to represent the results of a multi-factorial experiment graphically. Figure 16.2A shows the results of the experiment described in the previous section. The levels of factor $\mathcal{A}$ are represented on the abscissa (the horizontal, or x -axis), and the values of the dependent variable are represented on the ordinate (the vertical, or y -axis). The mean for each group is represented by a circle, and the circles representing the groups in the same level of the independent variable $\mathcal{B}$ are joined by a line. This approach is equivalent to representing a two-factor design as several one-factor designs (one per level of $\mathcal{B}$). Each of those one-factor designs is called a *sub-design* of the two-factor design. The experimental effect of a sub-design is called a *simple effect*. For example, the *simple effect* of $\mathcal{A}$ for level 1 of factor $\mathcal{B}$ is expressed

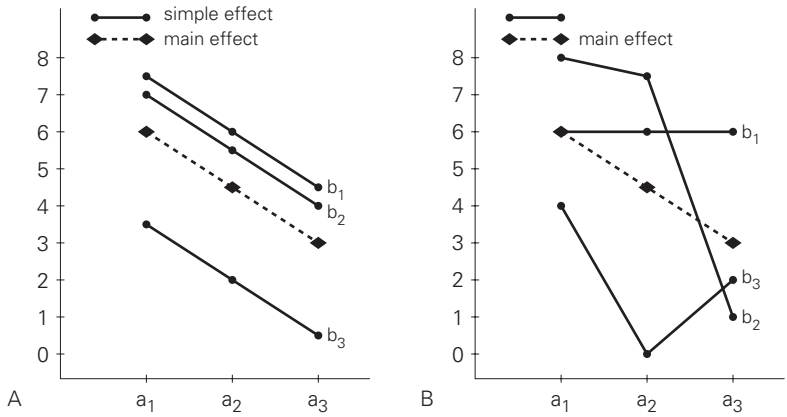


Figure 16.2 Two examples of two-factor designs: (A) without interaction (corresponding to the means of Table 16.1); (B) with interaction (corresponding to the means of Table 16.2). The y-axis represents the values of a hypothetical dependent variable.

by the differences between the experimental groups a_1b_1 , a_2b_1 , and a_3b_1 .¹ These three groups constitute a sub-design of the two-factor design $S(\mathcal{A} \times \mathcal{B})$. The main effect of factor $\mathcal{A}$ is also drawn on Figure 16.2A with diamonds and a dashed line. It corresponds to the values of the means $M_{1..}$, $M_{2..}$, and $M_{3..}$ represented by diamonds.

The lines depicting the three simple effects of factor $\mathcal{A}$ for (or conditional on) the different levels of factor $\mathcal{B}$ have the same general shape. Each group in the a_1 condition always has a mean with a value of 1.5 units more than the group in the a_2 condition, and 3.0 units more than the group in the a_3 condition. This is true whatever the level of $\mathcal{B}$. As a consequence, the curves describing the three simple effects are parallel. The three solid curves for the simple effects are averaged in the overall curve depicting the main effect of $\mathcal{A}$ (dashed line). Conversely, in this example we could have plotted the effects of $\mathcal{B}$ at different levels of $\mathcal{A}$ to see the simple effects of $\mathcal{B}$, which again would add up to a main effect of $\mathcal{B}$.

Another way of grasping this point can be obtained by looking at the group means and the marginal means in Table 16.1. Note that the mean of any group can be computed precisely knowing only the marginal means and the grand mean. The main effects are simply *additive* in a case such as this where there is no interaction. For example, the mean of the group in the a_1b_1 condition is computed by adding the effect of a_1 ($M_{1..} - M_{...}$) and the effect of b_1 ($M_{.1} - M_{...}$) to the grand mean (i.e. $7.5 = 4.5 + 1.5 + 1.5$). As an exercise, you can try to rebuild the table from the marginal means.

16.3.4 Example with interaction

Table 16.2 shows the results of another hypothetical experiment. The main effects (i.e. the margins of the table) are the same as those described previously in Table 16.1.

The results are plotted in Figure 16.2B. As can be seen by comparison with Figure 16.2A, the main effect of $\mathcal{A}$ is the same in both cases. To see the main effect of $\mathcal{B}$, you would need to plot the effects of $\mathcal{A}$ as a function of $\mathcal{B}$, or to check the values of the marginal means.

¹ Strictly speaking, we should say: ‘the simple effect of factor $\mathcal{A}$ conditional on the first level of $\mathcal{B}$ ’.

Factor $\mathcal{B}$	Factor $\mathcal{A}$			Means
	a_1	a_2	a_3	
b_1	6.0	6.0	6.0	6.0
b_2	8.0	7.5	1.0	5.5
b_3	4.0	0.0	2.0	2.0
Means	6.0	4.5	3.0	4.5

Table 16.2 Means of a fictitious experiment with main effects and an interaction (compare with Table 16.1).

The important point is that the curves describing the effects of $\mathcal{A}$ for the different levels of $\mathcal{B}$ are no longer parallel. That means that the *effect of $\mathcal{B}$ depends upon the levels of $\mathcal{A}$* . In other words, to know the effects of $\mathcal{A}$, a knowledge of $\mathcal{B}$ is necessary. This is what defines an *interaction* between two factors.

16.3.5 More about the interaction

The concept of interaction can be defined many ways. The concept is important enough for us to have a look at several of them:

- When $\mathcal{A}$ and $\mathcal{B}$ interact, the simple effects of one factor are different from each other, and the curves describing them are not parallel.
- There is an interaction if the effect of one independent variable depends upon the level of the other independent variable.
- There is an interaction if the main effect of an independent variable is not completely informative about the effects of that independent variable.
- There is an interaction when the group means (i.e. the $M_{ab\cdot}$ values) cannot be computed from the marginal means (i.e. the $M_{a\cdot\cdot}$ values and the $M_{\cdot b\cdot}$ values) and the grand mean (i.e. $M_{\dots}$).

16.4 Partitioning the experimental sum of squares

In a one-factor design the total sum of squares was partitioned into two parts: the first reflecting the effect of the independent variable $\mathcal{A}$ (SS_A or SS_{between}), and the other reflecting the error ($SS_{S(A)}$ or SS_{within}). The same general strategy is used here. The only difference is that the experimental sum of squares itself is going to be partitioned. In an $\mathcal{S}(\mathcal{A} \times \mathcal{B})$ design, the effect of the experiment is shown by the fact that the different groups corresponding to the different experimental conditions vary from each other in outcome. This is the *experimental* variability, and is formalized by an *experimental* sum of squares. The groups can differ because of the main effect of $\mathcal{A}$, the main effect of $\mathcal{B}$, or an interaction between $\mathcal{A}$ and $\mathcal{B}$. Each of these possible sources has a sum of squares expressing its effect. In following the computation of these effects bear in mind the relationship of within- and between-group deviations:

$$\begin{array}{rcl}
 Y_{abs} - M_{\dots} & = & (Y_{abs} - M_{ab\cdot}) + (M_{ab\cdot} - M_{\dots}) \\
 \text{deviation from} & \text{within-group} & \text{between-group} \\
 \text{the grand mean} & \text{deviation} & \text{deviation} .
 \end{array}$$

The between-group deviation itself can be decomposed into three parts:

$$\begin{aligned}
 (M_{ab.} - M_{...}) &= (M_{a..} - M_{...}) + (M_{.b.} - M_{...}) \\
 \text{between-group} &\quad \text{main effect} \quad \text{main effect} \\
 \text{deviation} &\quad \text{of } \mathcal{A} \quad \text{of } \mathcal{B} \\
 &+ (M_{ab.} - M_{a..} - M_{.b.} + M_{...}) \\
 &\quad \text{interaction between } \mathcal{A} \text{ and } \mathcal{B} .
 \end{aligned}$$

The result is obtained by dropping the parentheses and by simplifying. The main effects express themselves through the marginal means. The interaction can be interpreted as what is left of the experimental variability after the main effects have been taken into account (i.e. subtracted from the between-group variability). To make that point clearer, we can rewrite the previous equation as

$$\begin{aligned}
 (M_{ab.} - M_{a..} - M_{.b.} + M_{...}) &= (M_{ab.} - M_{...}) \\
 \text{interaction between} &\quad \text{between-group} \\
 \mathcal{A} \text{ and } \mathcal{B} &\quad \text{deviation} \\
 &- (M_{a..} - M_{...}) - (M_{.b.} - M_{...}) \\
 &\quad \text{effect of } \mathcal{A} \quad \text{effect of } \mathcal{B} .
 \end{aligned}$$

In order to compute the between-group sum of squares, the deviations are squared and summed. So the experimental sum of squares is computed as

$$SS_{\text{between}} = S \sum (M_{ab.} - M_{...})^2 .$$

By plugging the three-part decomposition of the between-group deviation into the previous equation, the sum of the squares between groups becomes

$$\begin{aligned}
 S \sum (M_{ab.} - M_{...})^2 &= S \sum [(M_{a..} - M_{...}) + (M_{.b.} - M_{...}) \\
 &\quad + (M_{ab.} - M_{a..} - M_{.b.} + M_{...})]^2 .
 \end{aligned}$$

This expression is like the old-timer $(a + b + c)^2$. It will expand into a series of square terms $(a^2 + b^2 + c^2)$ and a series of rectangular terms $(2ab + 2ac + 2bc)$. Using the same general technique used in Chapter 7 to partition the total sum of squares, all the rectangular terms can be shown to vanish *when the design is balanced* (recall that a design is balanced when there is the same number of observations per experimental condition). The proof is left as an exercise. The experimental sum of squares can be written as

$$\begin{aligned}
 S \sum (M_{ab.} - M_{...})^2 &= BS \sum (M_{a..} - M_{...})^2 + AS \sum (M_{.b.} - M_{...})^2 \\
 &\quad + S \sum (M_{ab.} - M_{a..} - M_{.b.} + M_{...})^2 .
 \end{aligned}$$

Therefore,

$$SS_{\text{between}} = SS_{\text{effect of } \mathcal{A}} + SS_{\text{effect of } \mathcal{B}} + SS_{\text{interaction } \mathcal{AB}} ,$$

or

$$SS_{\text{between}} = SS_{\text{experimental}} = SS_{\mathcal{A}} + SS_{\mathcal{B}} + SS_{\mathcal{AB}} .$$

(Note that a new notation, $\mathcal{AB}$, is used as a subscript to denote the interaction between $\mathcal{A}$ and $\mathcal{B}$.)

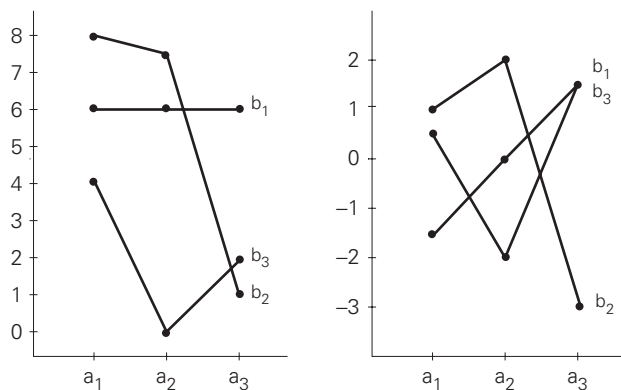


Figure 16.3 (Left) Plot of the experimental means ($M_{ab.}$). (Right) Plot of the pure interaction ($M_{ab.} - M_{a..} - M_{b.} + M_{...}$).

16.4.1 Plotting the pure interaction

As seen previously, a graph of the experimental means provides a good way of detecting an interaction by eye-balling the data. Those means, however, express all three sources of possible experimental effects. This can be illustrated by rewriting the equation of the between-group deviation as:

$$M_{ab.} = M_{...} + \underbrace{(M_{a..} - M_{...})}_{\text{effect of } \mathcal{A}} + \underbrace{(M_{b.} - M_{...})}_{\text{effect of } \mathcal{B}} + \underbrace{(M_{ab.} - M_{a..} - M_{b.} + M_{...})}_{\text{effect of interaction } \mathcal{AB}}.$$

If the purpose is to study the interaction, it is easier to look at the interaction only, and to plot its components only, namely to plot the terms $(M_{ab.} - M_{a..} - M_{b.} + M_{...})$ instead of the composite experimental means $M_{ab.}$. This technique is illustrated in Figure 16.3.

16.5 Degrees of freedom and mean squares

Reminder: In a one-factor design the strategy used to show an effect of the independent variable on the dependent variable is to compute the sum of squares between groups, then to divide it by its degrees of freedom in order to obtain a mean square. The division of the between-group mean square by the within-group mean square gives an F ratio which is, in turn, used to evaluate the likelihood that H_0 is true.

The same general procedure is used for a two-factor design. The only difference is that there are three sources of effects that can be tested in a two-factor experiment, namely two main effects for $\mathcal{A}$ and $\mathcal{B}$, and one interaction effect ($\mathcal{AB}$). There is an additional source of effects corresponding to the subjects nested in the design [$\mathcal{S}(\mathcal{AB})$]. The formulas for the computation of the sums of squares for these four sources have been seen already. Therefore the only problem is to find their degrees of freedom and then compute the mean squares.

The sum of squares for $\mathcal{A}$ is computed from the deviations of the A marginal means (the $M_{a..}$ values) to the grand mean. The grand mean is the mean of the $M_{a..}$ means. Hence, the sum of squares of $\mathcal{A}$ has $A - 1$ degrees of freedom. A similar argument gives $B - 1$ degrees

of freedom for the sum of squares of $\mathcal{B}$. What about the interaction $\mathcal{AB}$? The examination of Table 16.2 shows that the marginal means (the $M_{a..}$ values and the $M_{.b.}$ values) can be obtained from the experimental means (the $M_{ab.}$ values). Hence, in a given column, only $B - 1$ means are free to vary (because the last one has the value needed to make the mean of those means equal to the marginal mean). Likewise, in a given row, only $A - 1$ means are free to vary. Extending this principle to the complete table indicates that the total number of degrees of freedom for the interaction is $(A - 1)(B - 1)$.

The within-group sum of squares [corresponding to the source $\mathcal{S}(\mathcal{AB})$] is computed with the deviations of the subjects' scores from the mean of their group. In any given group there are $S - 1$ independent scores (why?). There are $A \times B$ groups, with $S - 1$ degrees of freedom each. Consequently the within-group sum of squares has $AB(S - 1)$ degrees of freedom. As a check, it is easy to verify that the degrees of freedom of these four sums of squares add up to the total number of degrees of freedom of the experiment:

$$(A - 1) + (B - 1) + (A - 1)(B - 1) + AB(S - 1) = ABS - 1.$$

Dividing each sum of squares by its number of degrees of freedom gives four mean squares:

$$\begin{aligned} MS_A &= \frac{SS_A}{df_A} = \frac{SS_A}{(A - 1)} \\ MS_B &= \frac{SS_B}{df_B} = \frac{SS_B}{(B - 1)} \\ MS_{AB} &= \frac{SS_{AB}}{df_{AB}} = \frac{SS_{AB}}{(A - 1)(B - 1)} \\ MS_{S(\mathcal{AB})} &= \frac{SS_{S(\mathcal{AB})}}{df_{S(\mathcal{AB})}} = \frac{SS_{S(\mathcal{AB})}}{AB(S - 1)}. \end{aligned} \quad (16.1)$$

These mean squares are now used to compute three F ratios: One for source $\mathcal{A}$, one for source $\mathcal{B}$, and one for source $\mathcal{AB}$. The F ratios are computed as follows:

To evaluate the effect of $\mathcal{A}$:

$$F_A = \frac{MS_A}{MS_{S(\mathcal{AB})}}.$$

This ratio will follow (when the null hypothesis is true) a Fisher distribution with $(A - 1)$ and $AB(S - 1)$ degrees of freedom.

To evaluate the effect of $\mathcal{B}$:

$$F_B = \frac{MS_B}{MS_{S(\mathcal{AB})}}.$$

This ratio will follow (when the null hypothesis is true) a Fisher distribution with $(B - 1)$ and $AB(S - 1)$ degrees of freedom.

To evaluate the effect of $\mathcal{AB}$:

$$F_{AB} = \frac{MS_{AB}}{MS_{S(\mathcal{AB})}}.$$

This ratio will follow (when the null hypothesis is true) a Fisher distribution with $(A - 1)(B - 1)$ and $AB(S - 1)$ degrees of freedom.

The rationale behind this test is the same as for the one-factor design. Under the null hypothesis both the numerator and denominator are estimations of the error.

16.6 The score model (Model I) and the sums of squares

Just as in the ANOVA for a one-factor design, a subject's score can be decomposed so that each potential source of effect is represented by a component. In a two-factor design the score of a subject in experimental group ab is decomposed in the following manner.

$$Y_{abs} = \mu... + \alpha_a + \beta_b + \alpha\beta_{ab} + e_{s(ab)}$$

with α_a : the effect of the a th level of $\mathcal{A}$

β_b : the effect of the b th level of $\mathcal{B}$

$\alpha\beta_{ab}$: the effect of the a th level of $\mathcal{A}$ and the b th level of $\mathcal{B}$

$e_{s(ab)}$: the error associated with subject s in the condition ab

$\mu...$: the population mean

When $\mathcal{A}$ and $\mathcal{B}$ are fixed factors the following conditions are imposed on the parameters of the score model (see Chapter 10).

- The sum of the α_a is 0 : $\sum \alpha_a = 0$.
- The sum of the β_b is 0 : $\sum \beta_b = 0$.
- The sum over the index a as well as the index b of $\alpha\beta_{ab}$ is 0 :

$$\sum_a \alpha\beta_{ab} = \sum_b \alpha\beta_{ab} = 0.$$

Also, conditions for the error must be specified in addition to those above:

- The expected value of the error is zero. Its intensity is independent of the experimental groups. Its distribution is normal.
- Briefly: the error is $\mathcal{N}(0, \sigma_e^2)$.

The different elements of the score model can be estimated from a sample:

$$\text{est}\{\alpha_a\} = (M_{a\cdot} - M_{...})$$

$$\text{est}\{\beta_b\} = (M_{\cdot b} - M_{...})$$

$$\text{est}\{\alpha\beta_{ab}\} = (M_{ab\cdot} - M_{a\cdot\cdot} - M_{\cdot b} + M_{...})$$

$$\text{est}\{e_{s(ab)}\} = (Y_{abs} - M_{ab\cdot}).$$

As always, the parameter estimations of the score model are closely associated with the sums of squares. Precisely:

$$SS_A = BS \sum \text{est}\{\alpha_a\}^2$$

$$SS_B = AS \sum \text{est}\{\beta_b\}^2$$

$$SS_{AB} = S \sum \text{est}\{\alpha\beta_{ab}\}^2$$

$$SS_{S(AB)} = \sum \text{est}\{e_{s(ab)}\}^2.$$

Likewise, the different quantities of the ANOVA can be expressed in terms of the score model as shown below:

$$M_{ab\cdot} = \mu... + \alpha_a + \beta_b + \alpha\beta_{ab} + \bar{e}_{\cdot(ab)},$$

with $\bar{e}_{(ab)}$ the mean of the error for the group ab (see Chapter 10 on the score model).

$$SS_A = BS \sum (M_{a..} - M_{...})^2 = BS \sum (\alpha_a + \bar{e}_{(a.)} - \bar{e}_{(..)})^2$$

$$SS_B = AS \sum (M_{.b.} - M_{...})^2 = AS \sum (\beta_b + \bar{e}_{(.b)} - \bar{e}_{(..)})^2$$

$$\begin{aligned} SS_{AB} &= S \sum (M_{ab.} - M_{a..} - M_{.b.} + M_{...})^2 \\ &= S \sum (\alpha\beta_{ab} - \bar{e}_{(a.)} + \bar{e}_{(..)} - \bar{e}_{(.b)} + e_{(ab)})^2 \end{aligned}$$

$$SS_{S(AB)} = \sum (Y_{abs} - M_{ab.})^2 = \sum (e_{s(ab)} - \bar{e}_{.ab})^2.$$

Here, each of the sums of squares of the effects is made of a term expressing the effect of the source and an error term. For example, the sum of squares of $\mathcal{A}$ consists of the values α_a , corresponding to the effect of $\mathcal{A}$, and terms in 'e' corresponding to the error. By applying the rules for mathematical expectation, the following values are found for the expected values of the different mean squares:

$$\mathbf{E}\{MS_A\} = \sigma_e^2 + BS\vartheta_a^2$$

$$\mathbf{E}\{MS_B\} = \sigma_e^2 + AS\vartheta_b^2$$

$$\mathbf{E}\{MS_{AB}\} = \sigma_e^2 + S\vartheta_{ab}^2$$

$$\mathbf{E}\{MS_{S(AB)}\} = \sigma_e^2.$$

with

$$\vartheta_a^2 = \sum \alpha_a^2 / (A - 1),$$

$$\vartheta_b^2 = \sum \beta_b^2 / (B - 1),$$

$$\vartheta_{ab}^2 = \sum \alpha\beta_{ab}^2 / (A - 1)(B - 1),$$

$$\sigma_e^2 = \mathbf{E}\{e_{s(ab)}^2\}.$$

The statistical test procedure previously described is directly applicable here. To test the existence of an effect corresponding to a particular source, simply assume the absence of this effect under the null hypothesis. Then find *under the null hypothesis* a mean square having the same expected value. Under the null hypothesis (and the 'validity assumptions') the ratio of these two mean squares follows a Fisher distribution. If the probability associated with this F ratio is small, the null hypothesis is rejected, and the alternative hypothesis of the existence of the effect is accepted.

The test of the effect of $\mathcal{A}$, the null hypothesis, is: there is no effect of $\mathcal{A}$ and so all α_a are zero and $BS\vartheta_a^2 = 0$.

Under the null hypothesis, note that MS_A and $MS_{S(AB)}$ have the same expected value. In other words, these two mean squares estimate the same thing: the error variance. As a consequence, the ratio:

$$F_A = \frac{MS_A}{MS_{S(AB)}}$$

follows a Fisher distribution under the null hypothesis with $(A - 1)$ and $AB(S - 1)$ degrees of freedom.

Similarly, the effect of $\mathcal{B}$ is tested by computing:

$$F_B = \frac{MS_B}{MS_{S(AB)}},$$

which follows a Fisher distribution under the null hypothesis with $(B - 1)$ and $AB(S - 1)$ degrees of freedom.

Finally, we calculate the F index for the interaction $\mathcal{AB}$ as

$$F_{AB} = \frac{MS_{AB}}{MS_{S(AB)}},$$

which follows a Fisher distribution under the null hypothesis with $(A - 1)(B - 1)$ and $AB(S - 1)$ degrees of freedom.

After all this: a nice example ...

16.7 Example: cute cued recall

Let us go back to the example used at the beginning of the chapter. The general story (from Tulving and Pearlstone, 1966) is that 60 subjects (10 per experimental group, so $S = 10$) are asked to memorize lists of 12, 24, or 48 words (factor $\mathcal{A}$ with 3 levels). These words can be clustered by pairs into categories (e.g. apple and orange are fruits). Subjects are asked to learn these words. During the learning phase, the experimenter shows the subjects the category names but tells them that they do not have to learn the names of the categories.

The test takes place immediately after the learning phase and involves two conditions. Half the subjects are shown the category names, whereas the other half are not (factor $\mathcal{B}$, with 2 levels: cued vs free recall). In this replication of the experiment, the dependent variable is the number of words recalled by each subject. The results are presented in Table 16.3.

Factor $\mathcal{B}$	Factor $\mathcal{A}$					
	a ₁ : 12 Words		a ₂ : 24 Words		a ₃ : 48 Words	
b ₁ Free recall	11	07	13	15	17	16
	09	12	18	13	20	23
	07	11	19	09	22	19
	11	10	13	08	13	20
	12	10	08	14	21	19
b ₂ Cued recall	12	10	13	14	32	30
	12	12	21	13	31	33
	07	10	20	14	27	25
	09	07	15	16	30	25
	09	12	17	07	29	28

Table 16.3 Results of a replication of the Tulving and Pearlstone experiment (1965). The dependent variable is the number of words recalled (see text for explanation).

If you consider the independent variables, you probably feel that the effect of the first one (i.e. list length) is trivial (because the greater the number of words learned, the greater the number of words that should be recalled, indeed). This means that the authors of the experiment were clearly interested in an *interaction* effect between the independent variables rather than in the main effects. The results of the experiment are displayed in the mean table (*not* a nasty one!) presented below.

Factor <i>B</i> means	Factor <i>A</i> means			
	a ₁ : 12 words	a ₂ : 24 words	a ₃ : 48 words	Margin
b ₁ : Free recall	10	13	19	14
b ₂ : Cued recall	10	15	29	18
Margin	10	14	24	Grand mean 16

The results are graphically displayed in Figure 16.4, which suggests the existence of a main effect of the list length, a main effect of the presentation of the cue, and also an interaction effect: *the facilitating effect of the cue is much stronger for the long lists than for the short lists*. The computation of the ANOVA confirms this interpretation. The details of the computations are shown below.

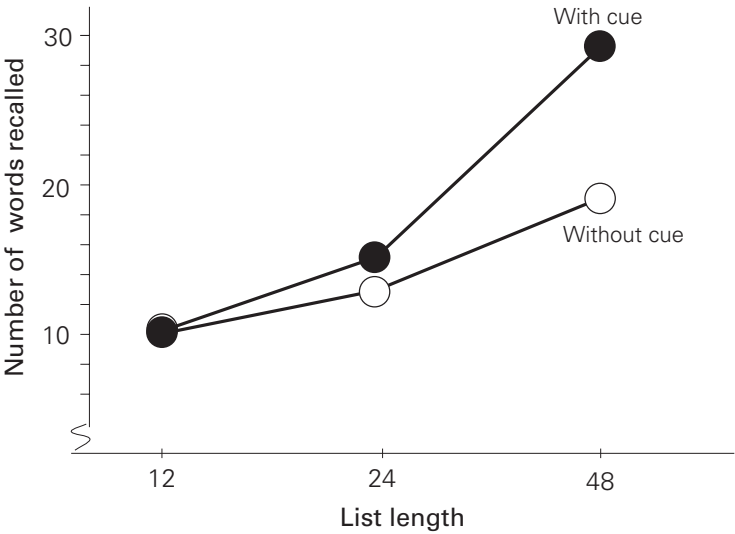


Figure 16.4 Mean number of words recalled with or without cue (factor *B*, with 2 levels) given at test as a function of list length (factor *A*, with 3 levels). Data from a replication of Tulving and Pearlstone (1965).

Try to do the computation on your own first, then check that you obtain the right results.

First we need to compute the different sums of squares.

Sum of squares for main effect of $\mathcal{A}$:

$$\begin{aligned}
 SS_A &= BS \sum_a (M_{a..} - M_{...})^2 \\
 &= 2 \times 10[(10.00 - 16.00)^2 + (14.00 - 16.00)^2 + (24.00 - 16.00)^2] \\
 &= 20[36.00 + 4.00 + 64.00] \\
 &= 2,080.00 .
 \end{aligned}$$

Sum of squares for main effect of $\mathcal{B}$:

$$\begin{aligned}
 SS_B &= AS \sum_b (M_{.b.} - M_{...})^2 \\
 &= 3 \times 10[(14.00 - 16.00)^2 + (18.00 - 16.00)^2] \\
 &= 30[4.00 + 4.00] \\
 &= 240.00 .
 \end{aligned}$$

Sum of squares for interaction between $\mathcal{A}$ and $\mathcal{B}$:

$$\begin{aligned}
 SS_{AB} &= S \sum_{ab} (M_{ab.} - M_{a..} - M_{.b.} + M_{...})^2 \\
 &= 10[(10.00 - 10.00 - 14.00 + 16.00)^2 + (10.00 - 10.00 - 18.00 + 16.00)^2 \\
 &\quad + (13.00 - 14.00 - 14.00 + 16.00)^2 + (15.00 - 14.00 - 18.00 + 16.00)^2 \\
 &\quad + (19.00 - 24.00 - 14.00 + 16.00)^2 + (29.00 - 24.00 - 18.00 + 16.00)^2] \\
 &= 10[4.00 + 4.00 + 1.00 + 1.00 + 9.00 + 9.00] \\
 &= 280.00 .
 \end{aligned}$$

Sum of squares within experimental group:

$$\begin{aligned}
 SS_{S(AB)} &= \sum_{abs} (Y_{abs} - M_{ab.})^2 \\
 &= (11.00 - 10.00)^2 + \dots + (10.00 - 10.00)^2 + (12.00 - 10.00)^2 + \dots + (12.00 - 10.00)^2 \\
 &\quad + (13.00 - 13.00)^2 + \dots + (14.00 - 13.00)^2 + (13.00 - 15.00)^2 + \dots + (7.00 - 15.00)^2 \\
 &\quad + (17.00 - 19.00)^2 + \dots + (19.00 - 19.00)^2 + (32.00 - 29.00)^2 + \dots + (28.00 - 29.00)^2 \\
 &= 1.00 + \dots + 0.00 + 4.00 + \dots + 4.00 \\
 &\quad + 0.00 + \dots + 1.00 + 4.00 + \dots + 64.00 \\
 &\quad + 4.00 + \dots + 0.00 + 9.00 + \dots + 1.00 \\
 &= 486.00 .
 \end{aligned} \tag{16.2}$$

We can now fill in the following table:

Source	df	SS	MS	F
$\mathcal{A}$	$A - 1$	SS_A	$\frac{SS_A}{(A - 1)}$	$\frac{MS_A}{MS_{S(AB)}}$
$\mathcal{B}$	$B - 1$	SS_B	$\frac{SS_B}{(B - 1)}$	$\frac{MS_B}{MS_{S(AB)}}$
$\mathcal{AB}$	$(A - 1)(B - 1)$	SS_{AB}	$\frac{SS_{AB}}{(A - 1)(B - 1)}$	$\frac{MS_{AB}}{MS_{S(AB)}}$
$S(\mathcal{AB})$	$AB(S - 1)$	$SS_{S(AB)}$	$\frac{SS_{S(AB)}}{AB(S - 1)}$	
Total	$ABS - 1$	SS_{total}		

which gives:

Source	df	SS	MS	F	$p(F)$
$\mathcal{A}$	2	2,080.00	1,040.00	115.56	<.000001
$\mathcal{B}$	1	240.00	240.00	26.67	.000007
$\mathcal{AB}$	2	280.00	140.00	15.56	.000008
$S(\mathcal{AB})$	54	486.00	9.00		
Total	59	3,086.00			

The ANOVA table shows a (trivial) effect of the list length, $F(2, 54) = 115.56$, $MS_e = 9.00$, $p < .01$ on the number of words recalled. Giving cues at test improves subject performance: $F(1, 54) = 26.67$, $MS_e = 9.00$, $p < .01$. More importantly, the interaction between list length and presence of cues at test is clearly significant: $F(2, 54) = 15.56$, $MS_e = 9.00$, $p < .01$. This interaction is essentially due to the fact that the facilitating effect of the cue is manifest only for long lists (48 words). As a consequence, an experiment designed to show the effect of giving cues at test that used short lists would probably report negative findings.

16.8 Score model II: $\mathcal{A}$ and $\mathcal{B}$ random factors

16.8.1 Introduction and review

The difference between fixed and random factors has been presented previously. When an experimental design has two factors, this distinction still holds. However, an additional consideration exists when there are two factors: one factor may be fixed and the other random. When this is the case, the design is sometimes referred to as Model III, or 'mixed model'. We will begin by examining the case where both factors are random (Model II).

First, the good news: the calculations needed to obtain the mean squares are the same for the different models. Now for the bad news: the value of the index F is not obtained in the

same manner for the different models. After presenting the new formulas to obtain the F value for Model II, we will justify these formulas by using the score model.

16.8.2 Calculating F when $\mathcal{A}$ and $\mathcal{B}$ are random factors

Recall that when $\mathcal{A}$ and $\mathcal{B}$ are fixed factors the three F ratios are calculated as follows:

- Main effect of $\mathcal{A}$

$$F_A = \frac{MS_A}{MS_{S(AB)}} ,$$

- Main effect of $\mathcal{B}$

$$F_B = \frac{MS_B}{MS_{S(AB)}} ,$$

- Interaction effect

$$F_{AB} = \frac{MS_{AB}}{MS_{S(AB)}} .$$

When $\mathcal{A}$ and $\mathcal{B}$ are both random factors, the computations of F ratios change. In order to test the main effects of $\mathcal{A}$ or $\mathcal{B}$, their mean squares are not divided by the mean square within groups ($MS_{S(AB)}$) but instead they are divided by the interaction mean square (MS_{AB}). However, the effect of the interaction is still found by dividing the mean square of the interaction by the mean square within groups (i.e. $MS_{S(AB)}$).

Thus, the new formulas are as follows:

- Main effect of $\mathcal{A}$

$$F_A = \frac{MS_A}{MS_{AB}} ,$$

which will follow (when the null hypothesis is true) a Fisher distribution with $\nu_1 = (A - 1)$ and $\nu_2 = (A - 1)(B - 1)$ degrees of freedom.

- Main effect of $\mathcal{B}$

$$F_B = \frac{MS_B}{MS_{AB}} ,$$

which will follow (when the null hypothesis is true) a Fisher distribution with $\nu_1 = (B - 1)$ and $\nu_2 = (A - 1)(B - 1)$ degrees of freedom.

- Interaction effect

$$F_{AB} = \frac{MS_{AB}}{MS_{S(AB)}} ,$$

which will follow (when the null hypothesis is true) a Fisher distribution with $\nu_1 = (A - 1)(B - 1)$ and $\nu_2 = AB(S - 1)$ degrees of freedom.

The following information from the score model provides the rationale for these different formulas.

16.8.3 Score model when $\mathcal{A}$ and $\mathcal{B}$ are random

When $\mathcal{A}$ and $\mathcal{B}$ are both random, the score model is written as follows:

$$Y_{abs} = \mu_{...} + a_a + b_b + ab_{ab} + e_{s(ab)}$$

with a_a : the effect of the a th level of $\mathcal{A}$
 b_b : the effect of the b th level of $\mathcal{B}$
 ab_{ab} : the effect of the interaction between the a th level of $\mathcal{A}$
and the b th level of $\mathcal{B}$
 $e_{s(ab)}$: the error associated to subject s in the condition ab
 $\mu \dots$: the population mean

The usual condition is imposed on the error term: $e_{s(ab)}$ is normally distributed with variance σ_e^2 , denoted $\mathcal{N}(0, \sigma_e^2)$. Because the a_a , b_b , and ab_{ab} are (like $e_{s(ab)}$) random factors, we impose analogous conditions on their distribution:

- a_a is $\mathcal{N}(0, \sigma_a^2)$,
- b_b is $\mathcal{N}(0, \sigma_b^2)$,
- ab_{ab} is $\mathcal{N}(0, \sigma_{ab}^2)$.

When the rules of expected values are applied, this score model gives the following expected values for the mean squares:

$$\mathbf{E}\{MS_A\} = \sigma_e^2 + S\sigma_{ab}^2 + BS\sigma_a^2,$$

$$\mathbf{E}\{MS_B\} = \sigma_e^2 + S\sigma_{ab}^2 + AS\sigma_b^2,$$

$$\mathbf{E}\{MS_{AB}\} = \sigma_e^2 + S\sigma_{ab}^2,$$

$$\mathbf{E}\{MS_{S(AB)}\} = \sigma_e^2.$$

Examination of these formulas shows that the null hypothesis: ‘The independent variable $\mathcal{A}$ has no effect on the dependent variable’ is equivalent to assuming

$$\sigma_a^2 = 0.$$

Therefore, if the null hypothesis is true, then MS_A and MS_{AB} estimate the same error variance, and the ratio $F_A = MS_A/MS_{AB}$ will follow a Fisher distribution with $\nu_1 = (A - 1)$ and $\nu_2 = (A - 1)(B - 1)$ degrees of freedom.

The other F ratio formulas can be justified in the same manner.

16.8.4 $\mathcal{A}$ and $\mathcal{B}$ random: an example

Suppose a psychologist is interested in analyzing a projective test consisting of ten pictures which subjects must interpret. The purpose of this particular test is to measure the anxiety level of the subjects. The psychologist wants to test two research hypotheses. The first hypothesis of interest is that the order of presentation of the pictures influences a subject’s score on the dependent variable. The second hypothesis posits that the person administering the test has an influence on the scores obtained by the subjects.

Because there are ten pictures, $10!$ presentation orders are possible (and there are a lot of them, because $10! = 3,628,800$). It seems unrealistic, and certainly not practical, to assign an experimental group to each possible order. Therefore, the experimenter decides to choose *randomly* five presentation orders from the $10!$ possibilities. These five orders, once chosen, become the five levels of the independent variable: ‘presentation order of the test pictures’.

The number of psychologists qualified to administer a projective test is too large to use them all and so the experimenter needed to select a sample from this population.

Because the experimenter had enough funding to pay only four qualified psychologists, the size of the sample was, *de facto*, set to four. These four psychologists were selected at random from a professional list and all of them accepted—given the proper retribution—to participate in the experiment. (Note that the experimenter assumes this list to be representative of the population of psychologists. Does this seem justified to you?) These four psychologists become the four levels of the independent variable ‘test administrator’.

The experimenter decides to assign two participants to each condition.

Before going any further, take a sheet of paper and fully detail the steps of a statistical test:

1. Express the statistical hypotheses.
2. Determine a statistical index.
3. Choose a level of significance.
4. Define the sampling distribution under the null hypothesis.
5. Determine the region of rejection and of suspension of judgment.
6. Compute and decide.

Here are the results obtained:

Order ($\mathcal{B}$)	Test administrators ($\mathcal{A}$)				Means
	1	2	3	4	
I	127	117	111	108	113
	121	109	111	100	
II	117	113	111	100	107
	109	113	101	92	
III	107	108	99	92	99
	101	104	91	90	
IV	98	95	95	87	91
	94	93	89	77	
V	97	96	89	89	90
	89	92	83	85	
Means	106	104	98	92	100

If you do all the computations, you will obtain the results presented in the following ANOVA table. Remember, the mean squares are calculated in the same manner as when both factors are fixed.

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Pr(F)</i>
$\mathcal{A}$	3	1,200.00	400.00	20.00**	.00008
$\mathcal{B}$	4	3,200.00	800.00	40.00**	.00000
$\mathcal{AB}$	12	240.00	20.00	1.00 <i>ns</i>	.48284
$S(\mathcal{AB})$	20	400.00	20.00		
Total	39	5,040.00			

Note that the calculated F for the interaction is equal to 1; so we cannot reject the null hypothesis, and simply indicate this with the *ns* in the table.

16.9 ANOVA $\mathcal{A} \times \mathcal{B}$ (Model III): one factor fixed, one factor random

In the Model III or ‘mixed model’ case, one independent variable is fixed and the other independent variable is random. In the following discussion we will assume the independent variable $\mathcal{A}$ is the *fixed* factor, and $\mathcal{B}$ is the *random* factor. This model is frequently encountered in psychology; in particular, when variable $\mathcal{B}$ represents a *parasitic variable* or a *control factor* taken into account to improve the effectiveness of the design. Sometimes the term *randomized block design* is used for a particular case of this model. A variation of this design is presented in the following chapter on $\mathcal{S} \times \mathcal{A}$ designs (where $\mathcal{S}$ is a random factor and $\mathcal{A}$ is a fixed factor).

Before justifying the score model, here are the calculation formulas for the F index. They differ from those for the other two models of the factorial design $\mathcal{S}(\mathcal{A} \times \mathcal{B})$.

$$F_A = \frac{MS_A}{MS_{AB}} \quad \text{when } \mathcal{A} \text{ is fixed}$$

is distributed under the null hypothesis following a Fisher distribution with $\nu_1 = (A - 1)$ and $\nu_2 = (A - 1)(B - 1)$ degrees of freedom.

$$F_B = \frac{MS_B}{MS_{S(AB)}} \quad \text{when } \mathcal{B} \text{ is random}$$

is distributed under the null hypothesis following a Fisher distribution with $\nu_1 = (B - 1)$ and $\nu_2 = AB(S - 1)$ degrees of freedom.

$$F_{AB} = \frac{MS_{AB}}{MS_{S(AB)}}$$

is distributed under the null hypothesis following a Fisher distribution with $\nu_1 = (A - 1)(B - 1)$ and $\nu_2 = AB(S - 1)$ degrees of freedom.

Warning: A common mistake is to follow a first intuition and think that the fixed factor should be evaluated by dividing MS_A by $MS_{S(AB)}$. Looking at the score model shows why this is a mistake.

16.9.1 Score model for $\mathcal{A} \times \mathcal{B}$ (Model III)

When $\mathcal{A}$ is fixed, and $\mathcal{B}$ is random, the score model is written:

$$Y_{abs} = \mu \dots + \alpha_a + b_b + \alpha b_{ab} + e_{s(ab)}$$

with α_a : the effect of the a th level of $\mathcal{A}$

b_b : the effect of the b th level of $\mathcal{B}$

αb_{ab} : the effect of the interaction between the a th level of $\mathcal{A}$
and the b th level of $\mathcal{B}$

$e_{s(ab)}$: the error associated to subject s in the condition ab

$\mu \dots$: the population mean

with the usual conditions:

- $\sum \alpha_a = 0, \quad \sum \alpha_a^2 = (A-1)\vartheta_a^2,$
- $\sum_a \alpha b_{ab} = 0,$
- ab_{ab} is $\mathcal{N}(0, \sigma_{ab}^2)$ for all $a,$
- b_b is $\mathcal{N}(0, \sigma_b^2),$
- $e_{s(ab)}$ is $\mathcal{N}(0, \sigma_e^2).$

These conditions, combined with the expectation derivation, yield the following values for the mean squares of the different sources of variation:

$$\mathbf{E}\{MS_A\} = \sigma_e^2 + S\sigma_{ab}^2 + BS\vartheta_a^2$$

$$\mathbf{E}\{MS_B\} = \sigma_e^2 + AS\sigma_b^2$$

$$\mathbf{E}\{MS_{AB}\} = \sigma_e^2 + S\sigma_{ab}^2$$

$$\mathbf{E}\{MS_{S(AB)}\} = \sigma_e^2$$

An examination of these equations justifies the formulas given above.

16.10 Index of effect size

In an $\mathcal{A} \times \mathcal{B}$ design, four sources of effect can be identified: $\mathcal{A}$, $\mathcal{B}$, $\mathcal{AB}$, and $\mathcal{S}(\mathcal{AB})$. The source $\mathcal{S}(\mathcal{AB})$ represents the error and $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{AB}$ represent the experimental effects. Three kinds of indices of effect size can be computed:

- *Global indices* express the importance of one source of variability in comparison with all the sources of variability (experimental or not).
- *Partial indices* express the importance of a source of variability in comparison with the sum of this source and the error.
- *Indices corresponding to the partition of the experimental effect* express the importance of a source of variability in comparison with the experimental source of variability.

16.10.1 Index R^2 'global'

For each source of variability a 'global' index R^2 can be computed. It expresses the part of the total sum of squares taken by the sum of squares of the source. The interpretation of these indices in terms of coefficients of correlation will be detailed in the next section. Because there are three sources of effects ($\mathcal{A}$, $\mathcal{B}$, and $\mathcal{AB}$) in a two-factor design, three global indices R^2 can be computed:

The first gives the proportion of variance in the dependent variable Y explained by the independent variable $\mathcal{A}$:

$$R_{Y.A}^2 = \frac{SS_A}{SS_{\text{total}}}.$$

The second gives the proportion of variance in the dependent variable Y explained by the independent variable $\mathcal{B}$:

$$R^2_{Y.B} = \frac{SS_B}{SS_{\text{total}}}.$$

The third gives the proportion of variance in the dependent variable Y explained by the interaction between $\mathcal{A}$ and $\mathcal{B}$:

$$R^2_{Y.AB} = \frac{SS_{AB}}{SS_{\text{total}}}.$$

The relationship between correlation and analysis of variance is presented in detail in the next section.

16.10.2 The regression point of view

As we have seen previously for the designs with one factor, the ANOVA can be interpreted in the general framework of regression and correlation. We will illustrate this approach with a small fictitious example illustrated in Table 16.4. In this example, we have $S = 3$ subjects per experimental condition, $A = 3$ and $B = 2$.

The general idea is the same as in the one-factor case. With one factor, the ANOVA model is equivalent to trying to predict the subjects' scores from the mean of the group to which they belong. If we use this procedure here, we need to remember that the experimental means themselves are composed of different effects. Specifically, $M_{ab.}$, the mean of the group corresponding to the a th level of $\mathcal{A}$ and the b th level of $\mathcal{B}$, can be decomposed as:

$$\begin{aligned} M_{ab.} = & M_{...} + (M_{a..} - M_{...}) + (M_{.b.} - M_{...}) \\ & \text{effect of } \mathcal{A} \quad \text{effect of } \mathcal{B} \\ & + (M_{ab.} - M_{a..} - M_{.b.} + M_{...}) \\ & \text{effect of interaction } \mathcal{AB}. \end{aligned}$$

In order to make the formula easier to read, we can abbreviate the notations for the different components of a mean:

- The effect of the a th level of $\mathcal{A}$ which is equal to $(M_{a..} - M_{...})$ is denoted: $\mathcal{A}_a$. This corresponds to $\text{est}\{\alpha_a\}$ in the score model.

Levels of $\mathcal{B}$	Levels of $\mathcal{A}$			Means
	a_1	a_2	a_3	
b_1	18	15	3	$M_{1.} = 11$
	15	12	6	
	12	12	6	
b_2	8	11	12	$M_{2.} = 9$
	9	9	8	
	10	7	7	
Means	$M_{1..} = 12$	$M_{2..} = 11$	$M_{3..} = 7$	$M_{...} = 10$

Table 16.4 Results of a fictitious example to illustrate an $S(\mathcal{A} \times \mathcal{B})$ design.

- The effect of the b th level of $\mathcal{B}$ which is equal to $(M_{.b.} - M_{...})$ is denoted: $\mathcal{B}_b$. This corresponds to $\text{est}\{\beta_b\}$ in the score model.
- The effect of the interaction between the a th level of $\mathcal{A}$ and the b th level of $\mathcal{B}$ which is $(M_{ab.} - M_{a..} - M_{.b.} + M_{...})$ is denoted: $\mathcal{AB}_{ab}$. This corresponds to $\text{est}\{\alpha\beta_{ab}\}$ in the score model.

With this new notation, the previous equation describing the components of the mean $M_{ab.}$ can be rewritten as

$$M_{ab.} = M_{...} + \mathcal{A}_a + \mathcal{B}_b + \mathcal{AB}_{ab} .$$

As an illustration, we will find the values of these components for the example given in Table 16.4.

$$\mathcal{A}_1 = M_{1..} - M_{...} = 12 - 10 = 2$$

$$\mathcal{A}_2 = M_{2..} - M_{...} = 11 - 10 = 1$$

$$\mathcal{A}_3 = M_{3..} - M_{...} = 7 - 10 = -3$$

$$\mathcal{B}_1 = M_{.1.} - M_{...} = 11 - 10 = 1$$

$$\mathcal{B}_2 = M_{.2.} - M_{...} = 9 - 10 = -1$$

$$\mathcal{AB}_{11} = M_{11.} - M_{1..} - M_{.1.} + M_{...} = 15 - 12 - 11 + 10 = 2$$

$$\mathcal{AB}_{12} = M_{12.} - M_{1..} - M_{.2.} + M_{...} = 9 - 12 - 9 + 10 = -2$$

$$\mathcal{AB}_{21} = M_{21.} - M_{2..} - M_{.1.} + M_{...} = 13 - 11 - 11 + 10 = 1$$

$$\mathcal{AB}_{22} = M_{22.} - M_{2..} - M_{.2.} + M_{...} = 9 - 11 - 9 + 10 = -1$$

$$\mathcal{AB}_{31} = M_{31.} - M_{3..} - M_{.1.} + M_{...} = 5 - 7 - 11 + 10 = -3$$

$$\mathcal{AB}_{32} = M_{32.} - M_{3..} - M_{.2.} + M_{...} = 9 - 7 - 9 + 10 = 3$$

If you look carefully at the components of the group mean you might notice the following relationship:

$$\sum_a \mathcal{A}_a = \sum_b \mathcal{B}_b = \sum_a \mathcal{AB}_{ab} = \sum_b \mathcal{AB}_{ab} = 0 .$$

Can you see why this relationship holds? Let us look at the first of these, and leave the rest to you as an exercise or pastime.

16.10.2.1 Digression: the sum equals zero

In this digression we want to show that

$$\sum_a \mathcal{A}_a = 0 .$$

We start by replacing $\mathcal{A}_a$ by its value in terms of the means, and then develop and simplify:

$$\sum_a \mathcal{A}_a = \sum_a (M_{a..} - M_{...}) .$$

The first step is to distribute the $\sum$ sign:

$$\sum_a \mathcal{A}_a = \sum_a M_{a..} - \sum_a M_{...} .$$

Because $M_{...}$ is a constant, $\sum_a M_{...} = AM_{...}$, hence

$$\sum_a \mathcal{A}_a = \sum_a M_{a..} - AM_{...} ,$$

and the grand mean $M_{...}$ can be obtained from the group mean as

$$M_{...} = \frac{1}{A} \sum_a M_{a..}$$

or

$$\sum_a M_{a..} = AM_{...}$$

Therefore we obtain

$$\begin{aligned} \sum_a \mathcal{A}_a &= AM_{...} - AM_{...} \\ &= 0 . \end{aligned}$$

Et voilà!

16.10.2.2 Back from the digression

Because each group mean is made of several components, in order to predict the subjects' scores by the mean of their group we will try to predict their scores from each of the components of the means. Hence, we will try to predict each score Y_{abs} from $\mathcal{A}_a$, $\mathcal{B}_b$ and $\mathcal{AB}_{ab}$. For each score, the difference between the predicted value and the actual value is the residual. It is noted $e_{s(ab)}$, hence the residual of the score of subject s in condition ab is

$$e_{s(ab)} = (Y_{abs} - M_{ab.}) .$$

An equivalent way to present the previous equation is to express each individual score as a sum of components. Namely,

$$\begin{aligned} Y_{abs} &= M_{ab.} + e_{s(ab)} \\ &= M_{...} + \mathcal{A}_a + \mathcal{B}_b + \mathcal{AB}_{ab} + e_{s(ab)} . \end{aligned}$$

In this equation the term $e_{s(ab)}$ (i.e. the residual) expresses the variability of the dependent variable that cannot be explained by the experimental manipulations, whereas the other components (i.e. $\mathcal{A}_a$, $\mathcal{B}_b$, and $\mathcal{AB}_{ab}$) express the variability that can be explained by the experimental manipulation (because $M_{...}$ is a constant it cannot explain any variability in the data, it appears only as a scaling factor). Recall that the decomposition of the score Y_{abs} into its components in the ANOVA is referred to as the *score model* in the experimental design literature (see Chapter 10).

We can now organize all the different quantities to compute the coefficients of correlation between the scores and the components of the means. This is done in Table 16.5.

We can compute four coefficients of correlation (one for $\mathcal{A}$, one for $\mathcal{B}$, one for $\mathcal{AB}$, and one for the residual $e_{s(ab)}$). Each of these coefficients expresses the *proportion* of the total variability (i.e. of the total sum of squares) that can be explained by the source in question. These are, indeed, the very same coefficients that we have defined previously.

	Y	y	y^2	A	A^2	B	B^2	AB	AB^2	$e_{s(ab)}$	$e_{s(ab)}^2$
	18	8	64	2	4	1	1	2	4	3	9
	15	5	25	2	4	1	1	2	4	0	0
	12	2	4	2	4	1	1	2	4	-3	9
	8	-2	4	2	4	-1	1	-2	4	-1	1
	9	-1	1	2	4	-1	1	-2	4	0	0
	10	0	0	2	4	-1	1	-2	4	1	1
	15	5	25	1	1	1	1	1	1	2	4
	12	2	4	1	1	1	1	1	1	-1	1
	12	2	4	1	1	1	1	1	1	-1	1
	11	1	1	1	1	-1	1	-1	1	2	4
	9	-1	1	1	1	-1	1	-1	1	0	0
	7	-3	9	1	1	-1	1	-1	1	-2	4
	3	-7	49	-3	9	1	1	-3	9	-2	4
	6	-4	16	-3	9	1	1	-3	9	1	1
	6	-4	16	-3	9	1	1	-3	9	1	1
	12	2	4	-3	9	-1	1	3	9	3	9
	8	-2	4	-3	9	-1	1	3	9	-1	1
	7	-3	9	-3	9	-1	1	3	9	-2	4
Σ	180	0	240	0	84	0	18	0	84	0	54
	SS_{total}			SS_A		SS_B		SS_{AB}		$SS_{S(AB)}$	

Table 16.5 Detail of the different quantities needed to compute an ANOVA with an $S(A \times B)$ as a regression/correlation problem. The following abbreviations are used: $y = (Y_{abs} - M_{...})$, $y^2 = (Y_{abs} - M_{...})^2$, $A = (M_{a..} - M_{...})$, $B = (M_{.b.} - M_{...})$, and $AB = (M_{ab.} - M_{a..} - M_{.b.} + M_{...})$.

Specifically, the four coefficients of correlation are:

$$R_{Y \cdot A}^2 = \frac{SS_A}{SS_{\text{total}}}, \quad (16.3)$$

which expresses the intensity of the effect of A ;

$$R_{Y \cdot B}^2 = \frac{SS_B}{SS_{\text{total}}}, \quad (16.4)$$

which expresses the intensity of the effect of B ;

$$R_{Y \cdot AB}^2 = \frac{SS_{AB}}{SS_{\text{total}}}, \quad (16.5)$$

which expresses the intensity of the effect of the interaction AB ;

$$R_{Y \cdot S(AB)}^2 = \frac{SS_{S(AB)}}{SS_{\text{total}}}, \quad (16.6)$$

which expresses the importance of the *residual* effect of $S(AB)$ (variability due to the ‘subject factor’, or individual differences of the subjects in one condition).

For our example, we find the following values:

$$R_{Y \cdot A}^2 = \frac{SS_A}{SS_{\text{total}}} = \frac{84}{240} = .350 \quad (16.7)$$

$$R_{Y \cdot B}^2 = \frac{SS_B}{SS_{\text{total}}} = \frac{18}{240} = .075 \quad (16.8)$$

$$R_{Y \cdot AB}^2 = \frac{SS_{AB}}{SS_{\text{total}}} = \frac{84}{240} = .350 \quad (16.9)$$

$$R_{Y \cdot S(AB)}^2 = \frac{SS_{S(AB)}}{SS_{\text{total}}} = \frac{54}{240} = .225 \quad (16.10)$$

Note that the sum of these coefficients of correlation is one.

$$R_{Y \cdot A}^2 + R_{Y \cdot B}^2 + R_{Y \cdot AB}^2 + R_{Y \cdot S(AB)}^2 = .350 + .075 + .350 + .225 = 1.$$

This is, indeed, always the case. (Can you show why? It is because $SS_A + SS_B + SS_{AB} + SS_{S(AB)} = SS_{\text{total}}$. If this still does not seem obvious, you may want to re-read this chapter.)

16.10.3 **F ratios and coefficients of correlation**

If we use the same elementary algebraic manipulations that we have used in the one-factor designs, we can re-express the F ratios in terms of the coefficients of correlation. We start by giving the formulas for F and then we will show the derivation for one of them.

If we want to express the different F ratios as a function of the coefficients of correlation, we find:

$$F_A = \frac{R_{Y \cdot A}^2}{R_{Y \cdot S(AB)}^2} \times \frac{df_{S(AB)}}{df_A},$$

whose sampling distribution under the null hypothesis is a Fisher's F distribution with $\nu_1 = (A - 1)$ and $\nu_2 = AB(S - 1)$ degrees of freedom;

$$F_B = \frac{R_{Y \cdot B}^2}{R_{Y \cdot S(AB)}^2} \times \frac{df_{S(AB)}}{df_B},$$

whose sampling distribution under the null hypothesis is a Fisher's F distribution with $\nu_1 = (B - 1)$ and $\nu_2 = AB(S - 1)$ degrees of freedom; and

$$F_{AB} = \frac{R_{Y \cdot AB}^2}{R_{Y \cdot S(AB)}^2} \times \frac{df_{S(AB)}}{df_{AB}},$$

whose sampling distribution under the null hypothesis is a Fisher's F distribution with $\nu_1 = (A - 1)(B - 1)$ and $\nu_2 = AB(S - 1)$ degrees of freedom. (Compare these formulas with the formula for one-factor designs: what do you conclude?)

For our example we find the following values:

$$F_A = \frac{R_{Y \cdot A}^2}{R_{Y \cdot S(AB)}^2} \times \frac{df_{S(AB)}}{df_A} = \frac{.350}{.225} \times \frac{12}{2} = 9.333$$

$$F_B = \frac{R_{Y \cdot B}^2}{R_{Y \cdot S(AB)}^2} \times \frac{df_{S(AB)}}{df_B} = \frac{.075}{.225} \times \frac{12}{1} = 4.000$$

$$F_{AB} = \frac{R_{Y \cdot AB}^2}{R_{Y \cdot S(AB)}^2} \times \frac{df_{S(AB)}}{df_{AB}} = \frac{.350}{.225} \times \frac{12}{2} = 9.333$$

which indeed give the same values as the alternative formulas involving the mean squares, as we can check by first computing the mean squares:

$$MS_A = \frac{SS_A}{df_A} = \frac{84}{2} = 42.00$$

$$MS_B = \frac{SS_B}{df_B} = \frac{18}{1} = 18.00$$

$$MS_{AB} = \frac{SS_{AB}}{df_{AB}} = \frac{84}{2} = 42.00$$

$$MS_{S(AB)} = \frac{SS_{S(AB)}}{df_{S(AB)}} = \frac{54}{12} = 4.50 .$$

Then, these mean squares give the following values for the F ratios:

$$F_A = \frac{MS_A}{MS_{S(AB)}} = \frac{42.00}{4.50} = 9.333$$

$$F_B = \frac{MS_B}{MS_{S(AB)}} = \frac{18.00}{4.50} = 4.000$$

$$F_{AB} = \frac{MS_{AB}}{MS_{S(AB)}} = \frac{42.00}{4.50} = 9.333$$

16.10.3.1 Digression: two equivalent ways of computing the F ratio

In order to prove that the two formulas for the F ratio are equivalent, we will start with the formula for F involving mean squares, and then develop it. If we take F_A as an example, we find:

$$\begin{aligned} F_A &= \frac{MS_A}{MS_{S(AB)}} \\ &= \frac{SS_A/df_A}{SS_{S(AB)}/df_{S(AB)}} \\ &= \frac{SS_A \times df_{S(AB)}}{SS_{S(AB)} \times df_A} \\ &= \frac{SS_A}{SS_{S(AB)}} \times \frac{df_{S(AB)}}{df_A} \\ &= \frac{SS_A/SS_{\text{total}}}{SS_{S(AB)}/SS_{\text{total}}} \times \frac{df_{S(AB)}}{df_A} \\ &= \frac{R^2_{Y \cdot A}}{R^2_{Y \cdot S(AB)}} \times \frac{df_{S(AB)}}{df_A} \end{aligned}$$

16.10.4 Index R^2 'partial'

The size of the effect of a source of variability can also be determined in comparison with the error [$S(AB)$]. It suffices to divide the sum of squares for the source by the sum of squares for the source and the sum of squares of error. These new indices are called 'partial indices' (strictly speaking they should be called 'semi-partial', but the distinction is not important here, for more details, see Chapter 6, Section 6.4.2, page 118). For the one-factor designs $S(A)$, the source A being the only source of variability, global and partial indices are equivalent.

For each source of variability a partial index R^2 can be computed. These indices can be interpreted as the square of a (semi) partial coefficient of correlation, that is, the correlation between the source of variability and the subject's scores after subtracting the other sources of experimental effect. This procedure is often said *to partial out the effect of the other sources*. Hence, to estimate the size of the effect of $\mathcal{A}$ in comparison with the error, we compute the following index:

$$R_{Y.A/B,AB}^2$$

This is read as ' R squared between Y and A after the effect of B and AB has been partialled out or eliminated'.

$$R_{Y.A/B,AB}^2 = \frac{SS_A}{SS_A + SS_{S(AB)}} .$$

The coefficient $R_{Y.A/B,AB}^2$ can also be obtained as the squared coefficient of correlation between

$$\mathcal{A}_a = (M_{a.} - M_{...})$$

and

$$Y_{abs} - (M_{ab.} - M_{a..} - M_{.b.} + M_{...}) - (M_{.b.} - M_{...}) .$$

The last term represents what is left of Y_{abs} when the effects of $\mathcal{B}$ and $\mathcal{AB}$ have been partialled out. This coefficient of correlation is called a 'semi-partial coefficient of correlation'.

In the same way, the partial index of the strength of the effect for the source $\mathcal{B}$ is given by

$$R_{Y.B/A,AB}^2 = \frac{SS_B}{SS_B + SS_{S(AB)}} ,$$

which can also be obtained as the squared coefficient of correlation between

$$\mathcal{B}_b = (M_{.b.} - M_{...})$$

and

$$Y_{abs} - (M_{ab.} - M_{a..} - M_{.b.} + M_{...}) - (M_{a..} - M_{...})$$

where the second term represents what is left of Y_{abs} when the effects of $\mathcal{A}$ and $\mathcal{AB}$ have been partialled out.

The partial index for the interaction is given by

$$R_{Y.AB/A,B}^2 = \frac{SS_{AB}}{SS_{AB} + SS_{S(AB)}} ,$$

which can also be obtained as the squared coefficient of correlation between

$$\mathcal{AB}_{ab} = (M_{ab.} - M_{a..} - M_{.b.} + M_{...})$$

and

$$Y_{abs} - (M_{a..} - M_{...}) - (M_{.b.} - M_{...}) ,$$

where the second term represents what is left of Y_{abs} when the main effects of $\mathcal{A}$ and $\mathcal{B}$ have been partialled out.

16.10.5 Partitioning the experimental effect

In this section we look at coefficients giving an indication of the importance of an experimental source of effect compared to the total experimental effect.

Remember that the sum of squares ‘between the experimental groups’

$$S \sum (M_{ab.} - M_{...})^2$$

represents the experimental effect. Because

$$S \sum (M_{ab.} - M_{...})^2 = SS_{\text{experimental}} = SS_A + SS_B + SS_{AB} ,$$

it is possible to express the importance of an experimental source in comparison with the whole experimental effect. It suffices to compute some new indices R^2 as follows:

$$R^2_{\text{experimental-source}} = \frac{SS_{\text{source}}}{SS_{\text{experimental}}} .$$

As usual, those indices can be interpreted as coefficients of correlation between the source of effect and the experimental means (instead of the scores).

The strength of the effect of $\mathcal{A}$ in comparison to the experimental effect is given by

$$R^2_{\text{experimental-}\mathcal{A}} = \frac{SS_A}{SS_A + SS_B + SS_{AB}} .$$

This value is the squared coefficient of correlation taken between

$$\mathcal{A}_a = (M_{a..} - M_{...}) \quad \text{and} \quad M_{ab.} .$$

The strength of the effect of $\mathcal{B}$ in comparison with the experimental effect is given by

$$R^2_{\text{experimental-}\mathcal{B}} = \frac{SS_B}{SS_A + SS_B + SS_{AB}} .$$

This value is the squared coefficient of correlation taken between

$$\mathcal{B}_b = (M_{.b.} - M_{...}) \quad \text{and} \quad M_{ab.} .$$

The strength of the effect of the interaction $\mathcal{AB}$ in comparison with the experimental effect is given by

$$R^2_{\text{experimental-}\mathcal{AB}} = \frac{SS_{AB}}{SS_A + SS_B + SS_{AB}} .$$

This value is the squared coefficient of correlation taken between

$$\mathcal{AB}_{ab} = (M_{ab.} - M_{a..} - M_{.b.} + M_{...}) \quad \text{and} \quad M_{ab.} .$$

16.11 Statistical assumptions and conditions of validity

The conditions of validity needed to derive the sampling distribution of F (the so-called distribution conditions) for $\mathcal{S}(\mathcal{A})$ designs are still required for $\mathcal{S}(\mathcal{A} \times \mathcal{B})$ designs. Specifically the assumptions of independence of the observations, normality and homoscedasticity are still assumed to be true. However, as long as the design is balanced, those conditions

can be somewhat violated without causing too much trouble for the validity of the statistical decisions.

16.12 Computational formulas

The results of the ‘cute cued recall’ experiment on the effect of context on memory are recalled below.

	Factor A			
Factor B	a ₁ : 12 words	a ₂ : 24 words	a ₃ : 48 words	Means
b ₁ : Free recall				
Mean	10.00	13.00	19.00	14.00
Total	100.00	130.00	190.00	420.00
b ₂ : Cued recall				
Mean	10.00	15.00	29.00	18.00
Total	100.00	150.00	290.00	540.00
Margin				
Mean	10.00	14.00	24.00	16.00
Total	200.00	280.00	480.00	960.00

We already saw how to compute the sums of squares corresponding to these data using the comprehension formulas. We will now see how to compute them using the ‘numbers in a square’ presented in Chapter 8.

The following routine shows how to calculate the different sums of squares needed to compute the ANOVA using 11 values labeled Q1 to Q11 and five numbers in a square:

A **B** **AB** **ABS** **1** .

Q1 = Y... = 11 + ... + 28 = 960.00

Q2 = **ABS** = ∑ Y²_{abs} = 11² + ... + 28² = 18,446.00

Q3 = **AB** = ∑ $\frac{Y^2_{ab.}}{S}$ = $\frac{100^2 + \dots + 290^2}{10}$ = 17,960.00

Q4 = **A** = ∑ $\frac{Y^2_{a..}}{BS}$ = $\frac{200^2 + 280^2 + 480^2}{20}$ = 17,440.00

Q5 = **B** = ∑ $\frac{Y^2_{b.}}{AS}$ = $\frac{420^2 + 540^2}{30}$ = 15,600.00

Q6 = **1** = $\frac{Y^2_{...}}{ABS}$ = $\frac{960^2}{60}$ = 15,360.00

Q7 = SS_{total} = **ABS** - **1** = 3,086.00

$$Q8 = SS_A = \boxed{A} - \boxed{1} = 2,080.00$$

$$Q9 = SS_B = \boxed{B} - \boxed{1} = 240.00$$

$$Q10 = SS_{AB} = \boxed{AB} - \boxed{A} - \boxed{B} + \boxed{1} = 280.00$$

$$Q11 = SS_{S(AB)} = \boxed{ABS} - \boxed{AB} = 486.00 \quad (16.11)$$

Note that due to rounding differences the results differ slightly from the ones we found with the comprehension formulas. From these results, Table 16.6 shows how to build the ANOVA table given in Table 16.7.

Source	df	SS	MS	F
A	$A - 1$	$Q8.$	$\frac{Q8.}{(A - 1)}$	$\frac{MS_A}{MS_{S(AB)}}$
B	$B - 1$	$Q9.$	$\frac{Q9.}{(B - 1)}$	$\frac{MS_B}{MS_{S(AB)}}$
AB	$(A - 1)(B - 1)$	$Q10.$	$\frac{Q10.}{(A - 1)(B - 1)}$	$\frac{MS_{AB}}{MS_{S(AB)}}$
$S(AB)$	$AB(S - 1)$	$Q11.$	$\frac{Q11.}{AB(S - 1)}$	
Total	$ABS - 1$	$Q7.$		

Table 16.6 How to build the ANOVA table from the 11 step computational routine.

Source	df	SS	MS	F	$p(F)$
A	2	2,080.00	1,040.00	115.56	<.000,001
B	1	240.00	240.00	26.67	.000,007
AB	2	280.00	140.00	15.56	.000,008
$S(AB)$	54	486.00	9.00		
Total	59	3,086.00			

Table 16.7 The ANOVA table for Tulving and Pearlstone (1965) experiment (the data are in Table 16.3).

16.13 Relationships between the names of the sources of variability, df and SS

For an $S(A)$ design, the names of the sources of variability defined for the ANOVA are sufficient to reconstitute the full ANOVA table. This property will always hold for the designs described in this book. Table 16.8 summarizes those relationships for an $S(A \times B)$ design.

Source	<i>df</i>	Expanded <i>df</i>	Comprehension formula	Computational formula
<i>A</i>	$A - 1$	$A - 1$	$BS \sum (M_{a..} - M_{...})^2$	$\boxed{A} - \boxed{1}$
<i>B</i>	$B - 1$	$B - 1$	$AS \sum (M_{.b.} - M_{...})^2$	$\boxed{B} - \boxed{1}$
<i>AB</i>	$(A - 1)(B - 1)$	$AB - A - B + 1$	$S \sum (M_{ab.} - M_{a..} - M_{.b.} + M_{...})^2$	$\boxed{AB} - \boxed{A} - \boxed{B} + \boxed{1}$
<i>S(AB)</i>	$AB(S - 1)$	$ABS - AB$	$\sum (Y_{abs} - M_{ab.})^2$	$\boxed{ABS} - \boxed{AB}$
Total	$ABS - 1$	$ABS - 1$	$\sum (Y_{abs} - M_{...})^2$	$\boxed{ABS} - \boxed{1}$

Table 16.8 Relationships between the names of the sources of variability, *df* and *SS*.

Chapter summary

16.14 Key notions of the chapter

Below are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

Interaction between factor <i>A</i> and factor <i>B</i> .	Index of strength of effect ‘global’, ‘partial’, and ‘partitioning the experimental sum of squares’.
Partition of the experimental sum of squares into three parts: main effect of <i>A</i> , main effect of <i>B</i> and interaction between <i>A</i> and <i>B</i> .	

16.15 New notations

Below are the new notations introduced in this chapter. Test yourself on their meaning.

$M_{ab.} = \frac{1}{S} Y_{ab.}; \quad M_{a..} = \frac{1}{BS} Y_{a..}; \quad M_{.b.} = \frac{1}{AS} Y_{.b.}$

$SS_{AB}; \quad df_{AB}; \quad MS_{AB}; \quad \boxed{ABS} \quad ; \quad \boxed{AB}$

$R^2_{Y \cdot A/B, AB} \quad R^2_{Y \cdot B/A, AB} \quad R^2_{Y \cdot AB/A, B}$

$R^2_{\text{experimental-source}} \quad R^2_{\text{experimental-A}}$

$R^2_{\text{experimental-B}} \quad R^2_{\text{experimental-AB}}$

16.16 Key formulas of the chapter

Below are the main formulas introduced in this chapter: try to go through them and understand what they mean.

$$S \sum (M_{ab.} - M_{...})^2 = BS \sum (M_{a..} - M_{...})^2 + AS \sum (M_{.b.} - M_{...})^2 \\ + S \sum (M_{ab.} - M_{a..} - M_{.b.} + M_{...})^2$$

$$SS_{\text{experimental}} = SS_A + SS_B + SS_{AB}$$

For A and B Fixed

$$F_A = \frac{MS_A}{MS_{S(AB)}} \quad F_B = \frac{MS_B}{MS_{S(AB)}} \quad F_{AB} = \frac{MS_{AB}}{MS_{S(AB)}}$$

For A Fixed and B Random

$$F_A = \frac{MS_A}{MS_{AB}} \quad F_B = \frac{MS_B}{MS_{S(AB)}} \quad F_{AB} = \frac{MS_{AB}}{MS_{S(AB)}}$$

For A and B Random

$$F_A = \frac{MS_A}{MS_{AB}} \quad F_B = \frac{MS_B}{MS_{AB}} \quad F_{AB} = \frac{MS_{AB}}{MS_{S(AB)}}$$

$$R^2_{Y.\text{source}} = \frac{SS_{\text{source}}}{SS_{\text{total}}}$$

$$R^2_{Y.\text{source} / \text{other source}} = \frac{SS_{\text{source}}}{SS_{\text{source}} + SS_{\text{error}}}$$

$$R^2_{\text{experimental-source}} = \frac{SS_{\text{source}}}{SS_{\text{experimental}}}$$

16.17 Key questions of the chapter

Below are some questions about the content of this chapter. All the answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- * Why are the factorial designs preferable to the $S(A)$ design? (3 reasons)
- * How do you define the interaction between two factors? (4 equivalent definitions)
- * What is a main effect? A simple effect?
- * What is a partial coefficient of correlation?
- * How do you interpret the R^2 indices in terms of correlation coefficient?
- * What are the distribution conditions for a $S(A \times B)$ design?
- * To plot the interaction, why is it preferable to plot the estimations of the interaction component rather than the experimental means?

17

Factorial designs and contrasts

17.1 Introduction

When dealing with factorial designs it is as important as with single factor designs to be able to perform a fine-grained analysis of the results. The same general tools already used for simple designs are still in order: contrast analyses and sub-designs allow for the evaluation of specific hypotheses. The general idea for the analysis remains the same: a prediction corresponds to a contrast. But some new problems may occur. A first issue deals with evaluating whether a contrast corresponds to a finer-grained partition of the standard¹ decomposition performed in an $S(\mathcal{A} \times \mathcal{B})$ design. A second issue has to do with the choice of the error terms for testing a contrast. For an $S(\mathcal{A} \times \mathcal{B})$ design with both factors fixed, this last problem does not exist because all contrasts or sub-design mean squares have the same expected value under the null hypothesis: namely the error (i.e. $\sigma_{S(AB)}^2$) as estimated by $MS_{S(AB)}$.

17.2 Vocabulary

When dealing with factorial designs, it is useful to distinguish between two types of source of variance: the *experimental sources* and the *non-experimental sources*. The experimental sources have names made up only of letters corresponding to experimental factors (i.e. $\mathcal{A}$, $\mathcal{B}$, $\mathcal{AB}$ for a two-factor design). The sum of the sums of squares for these sources corresponds to the experimental sum of squares. The non-experimental sources have names which are made up of at least one letter that does not correspond to an experimental source. In most cases, this means that the letter $\mathcal{S}$ (corresponding to the subject factor) is part of the name of a non-experimental source [e.g. $S(\mathcal{AB})$]. In the specific case of an $S(\mathcal{AB})$ design, the experimental sources are $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{AB}$, whereas the non-experimental source is $\mathcal{S}(\mathcal{AB})$.

¹ By standard decomposition we refer to the partition of the experimental sum of squares into main effects and interaction sums of squares. This is called by some authors the *canonical* decomposition.

17.3 Fine-grained partition of the standard decomposition

In a factorial design, the experimental sources of variance correspond to an orthogonal partition of the experimental sum of squares (obviously!). Therefore, their analysis is equivalent to the sub-design type of analysis described in Section 12.13, pages 246ff. This is equivalent to saying that each source of effect can be obtained as a sum of orthogonal contrasts (with as many contrasts as the source has degrees of freedom). This analysis can be seen as a hierarchical analysis with each source being further analyzed with a set of (orthogonal) contrasts. This is illustrated by Figure 17.1.

17.3.1 An example: back to cute cued recall

As an example, the $S(\mathcal{A} \times \mathcal{B})$ design of Tulving and Pearlstone (Section 16.7) with $S = 10$, $A = 3$, and $B = 2$ can be analyzed using 5 orthogonal contrasts (there are $3 \times 2 = 6$

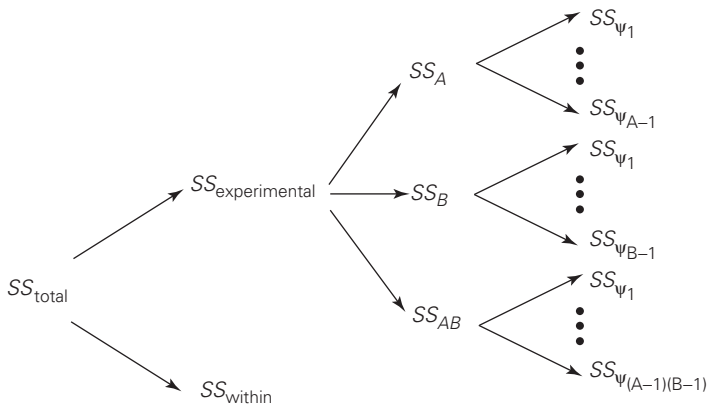


Figure 17.1 A fine-grained partition of the standard decomposition for a $S(\mathcal{A} \times \mathcal{B})$ design. The sources of effects in $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{AB}$ are further partitioned into a set of contrasts. Each source is decomposed into as many contrasts as the source has degrees of freedom.

	Experimental group						SS_ψ
	a_1b_1	a_1b_2	a_2b_1	a_2b_2	a_3b_1	a_3b_2	
M_{ab}	10	10	13	15	19	29	
ψ_1 [12 vs 24 & 48]	-2	-2	1	1	1	1	1,080.00
ψ_2 [24 vs 48]	0	0	1	1	-1	-1	1,000.00
ψ_3 [cue vs no cue]	1	-1	1	-1	1	-1	240.00
ψ_4 [interaction of ψ_1 and $\mathcal{B}$]	-2	2	1	-1	1	-1	120.00
ψ_5 [interaction of ψ_2 and $\mathcal{B}$]	0	0	1	-1	-1	1	160.00

Table 17.1 A possible family of orthogonal contrasts for analyzing the $S(\mathcal{A} \times \mathcal{B})$ of Tulving and Pearlstone (Chapter 12, Section 12.13, pages 246ff.). Contrasts 1 and 2 partition the sum of squares of $\mathcal{A}$. Contrast 3 expresses the sum of squares of $\mathcal{B}$. Contrasts 4 and 5 partition the sum of squares of $\mathcal{AB}$. The coefficients for contrasts 4 and 5 are obtained by multiplication of the coefficients of (respectively) ψ_1 by ψ_3 , and ψ_2 by ψ_3 .

experimental groups, hence the experimental sum of squares has 5 degrees of freedom). Table 17.1 gives a possible set of five orthogonal contrasts; as usual, other sets could be used.

17.3.1.1 The same old story: computing the sum of squares for a contrast

Computing the sum of squares for contrast in a factorial design follows the same logic seen for the one-factor design case. First, we compute the estimated value of the contrast as the weighted sum of the group means (with the weights being the coefficients of the contrast). We denote that estimation $\hat{\psi}$ (read ‘psi-hat’). The value of $\hat{\psi}$ is obtained by adapting the formula for a one-factor design contrast (see Equation 12.1, page 227) to accommodate a two-factor design:

$$\hat{\psi} = \sum_a \sum_b C_{ab} M_{ab}. \quad (17.1)$$

Then, the sum of squares for a contrast is obtained by adapting, once again, the equivalent formula from a one-factor design:

$$SS_{\psi} = \frac{S\hat{\psi}^2}{\sum C_{ab}^2}. \quad (17.2)$$

As an exercise, let us compute the sum of squares for Contrast 3 from Table 17.1. First, we compute the estimated value of this contrast:

$$\begin{aligned} \hat{\psi}_3 &= \sum_a \sum_b C_{ab} M_{ab} \\ &= (1 \times 10) + (-1 \times 10) + (1 \times 13) + (-1 \times 15) + (1 \times 19) + (-1 \times 29) \\ &= -12. \end{aligned}$$

Then the sum of squares of Contrast 3 is:

$$SS_{\psi_3} = \frac{S\hat{\psi}_3^2}{\sum C_{ab}^2} = \frac{10 \times (-12)^2}{6} = \frac{1440}{6} = 240.00.$$

17.3.1.2 Main effect contrasts

As can be seen in Table 17.1, the first 2 contrasts correspond to a partition of the sum of squares of $\mathcal{A}$. The first contrast, ψ_1 , opposes Condition a_1 to Conditions a_2 and a_3 ; the second contrast, ψ_2 , opposes Condition a_2 to a_3 . In these contrasts, the means for a given level of $\mathcal{A}$ have the *same* value for their contrast coefficients. For example, for ψ_1 the means of the first level of $\mathcal{A}$ are $M_{1,1,\cdot}$ and $M_{1,2,\cdot}$, and their contrast coefficients are equal to -2 . The third contrast ψ_3 corresponds to the main effect of $\mathcal{B}$. This contrast opposes the means of b_1 to the means of b_2 (i.e. the contrast coefficients for b_1 have the value $+1$, those for b_2 have the value -1).

Table 17.1 gives the values of SS_{ψ_1} and SS_{ψ_2} which can be computed with the same procedure used to compute the value of SS_{ψ_3} . Remember that the two contrasts ψ_1 and ψ_2 are orthogonal, and that $\mathcal{A}$ has 2 degrees of freedom. Therefore, the sum of SS_{ψ_1} and SS_{ψ_2} should be equal to the sum of squares of $\mathcal{A}$. And this is, indeed, the case:

$$SS_{\mathcal{A}} = 2,080 = 1,080 + 1,000 = SS_{\psi_1} + SS_{\psi_2}.$$

To recap, the sum of ψ_1 and ψ_2 corresponds to the sum of squares of the main effect of $\mathcal{A}$; the contrast ψ_3 corresponds to the main effect of $\mathcal{B}$. Because each of these contrasts partitions a main effect, these three contrasts are called *main effect* contrasts.

17.3.1.3 Interaction contrasts

Contrasts 4 and 5 from Table 17.1 give a partition of the interaction sum of squares. It is a bit more difficult to see why these contrasts are part of the interaction. The important information to remember comes from the score model: the sum of the interaction component is zero for each level of $\mathcal{A}$ and for each level of $\mathcal{B}$. An interaction contrast should also have that property.

Let us look at ψ_4 for an example. If we put the coefficients of ψ_4 in a 2 by 3 table, we can easily check that each row and each column sums to zero:

	a_1	a_2	a_3	$\sum_a C_{ab}$
b_1	$C_{1,1} = -2$	$C_{2,1} = 1$	$C_{3,1} = 1$	$\sum_a C_{a,1} = 0$
b_2	$C_{1,2} = 2$	$C_{2,2} = -1$	$C_{3,2} = -1$	$\sum_a C_{a,2} = 0$
$\sum_b C_{ab}$	$\sum_b C_{1,b} = 0$	$\sum_b C_{2,b} = 0$	$\sum_b C_{3,b} = 0$	

This property guarantees that this contrast is an interaction contrast. When we have already found a set of contrasts for the main effect of $\mathcal{A}$ and $\mathcal{B}$, the coefficients for a set of orthogonal interaction contrasts are easy to find: each interaction contrast can be obtained by multiplying the coefficients of the contrast on one main effect by the coefficients of the contrast on the other main effect. For example, the coefficients of Contrast 4 are obtained by multiplication of the coefficients of Contrasts 1 and 3. This is illustrated by the following table:

	$a_1 : C_{a_1} = -2$	$a_2 : C_{a_2} = 1$	$a_3 : C_{a_3} = 1$
b_1	$C_{1,1} =$	$C_{2,1} =$	$C_{3,1} =$
$C_{b_1} = 1$	$-2 \times 1 = -2$	$1 \times 1 = 1$	$1 \times 1 = 1$
b_2	$C_{1,2} =$	$C_{2,2} =$	$C_{3,2} =$
$C_{b_2} = -1$	$-2 \times -1 = 2$	$1 \times -1 = -1$	$1 \times -1 = -1$

Using the same procedure, the coefficients of Contrast 5 are obtained by multiplication of the coefficients of Contrasts 2 and 3. In brief,

$$C_{a,b;4} = C_{a,b;1} \times C_{a,b;3} \quad \text{and} \quad C_{a,b;5} = C_{a,b;2} \times C_{a,b;3}. \quad (17.3)$$

(with $C_{a,b;4}$ being the coefficients of Contrast 4).

Incidentally, the notation ‘ $\times$ ’ used to denote the interaction in terms like $\mathcal{A} \times \mathcal{B}$ comes from the fact that an interaction contrast is *always* obtained as the product of contrasts of the main effects involved. In general, if the main effect of $\mathcal{A}$ is partitioned into $(A - 1)$ contrasts (one per degree of freedom) and if the main effect of $\mathcal{B}$ is expressed with $(B - 1)$ contrasts, the $(A - 1)(B - 1)$ interaction contrasts are obtained by all the pairwise multiplications of one contrast for $\mathcal{A}$ with one contrast of $\mathcal{B}$.

17.3.1.4 Adding contrasts: sub-design analysis

Because the contrasts are orthogonal, their sums of squares as well as their degrees of freedom can be added to create a comparison with more than one degree of freedom (i.e. a subdesign analysis). This is particularly useful for analyzing the interaction of one contrast on a main effect of one-factor with another factor. As an example, suppose that in an $\mathcal{S}(\mathcal{A} \times \mathcal{B})$ design,

$\mathcal{A}$ represents several (say three) dosages of a drug (plus the usual placebo condition which makes a total of 4 levels for this factor); and that $\mathcal{B}$ represents the age of the subject (a tag variable obviously), with 3 levels (young, middle aged, a bit more aged). Suppose, also, that your favorite theory about the effect of the drug (i.e. factor $\mathcal{A}$) is translated as one contrast (so even though the factor $\mathcal{A}$ has 3 degrees of freedom, we are interested in testing only *one* contrast). This contrast partitions the main effect of $\mathcal{A}$.² In order to represent the interaction of the contrast on factor $\mathcal{A}$ with factor $\mathcal{B}$, the first step is to partition the effect of $\mathcal{B}$ into two contrasts (one per degree of freedom). Then two interaction contrasts expressing the interaction of the contrast on factor $\mathcal{A}$ will be computed by multiplication of the coefficients of this contrast with each contrast on $\mathcal{B}$. The sum of the sums of squares of these two interaction contrasts will then give the sum of squares of the interaction of the contrast (from factor $\mathcal{A}$) with factor $\mathcal{B}$. Because this sum of squares of interaction is obtained by the addition of the sums of squares of two orthogonal contrasts, it has 2 degrees of freedom. The interaction sum of squares has a total of $(A - 1)(B - 1) = 3 \times 2 = 6$ degrees of freedom. After subtraction of the interaction of the contrast of interest with $\mathcal{B}$, the residual of the interaction of $\mathcal{A} \times \mathcal{B}$ will have $6 - 2 = 4$ degrees of freedom.

17.4 Contrast analysis in lieu of the standard decomposition

In some cases, a prediction to be tested with a factorial design will correspond to a contrast that is not orthogonal to the standard decomposition. As an example, think of an $S(\mathcal{A} \times \mathcal{B})$ design with $A = 2$ and $B = 2$. Suppose that your pet theory predicts the outcome described in Figure 17.2. This corresponds to the contrast coefficients given in Table 17.2.

If we compare this contrast with the three contrasts described in Table 17.3, which express the usual decomposition of this design in term of main effects and interaction, we find that our ‘pet contrast’ is not orthogonal to any of these contrasts. Therefore it corresponds to a different partition of the experimental sum of squares. If this contrast is the only contrast of interest, then the best partition of the experimental sum of squares involves this contrast and the residual comparison with 2 degrees of freedom (i.e. the residual comparison sum of squares can be found by subtracting the sum of squares of the contrast of interest from the experimental sum of squares). Figure 17.3 illustrates this hierarchical partition: first the total sum of squares is partitioned into experimental sum of squares and within-group sum of squares. Then, the experimental sum of squares is, in turn, partitioned into the contrast sum of squares and the residual sum of squares.

If we want to test both this contrast and the standard decomposition, then we need to use the procedures tailored for the case of *a priori* non-orthogonal comparisons. Note that the problem of testing an *a priori* contrast that is not orthogonal to the standard decomposition can occur only when *both* factors $\mathcal{A}$ and $\mathcal{B}$ are *fixed* (because there is no way to predict the specific effect of a random factor *a priori* as we cannot even know *a priori* what levels of this factor will be selected).

² Implicitly, this contrast partitions the sum of squares of $\mathcal{A}$ into two comparisons: the first is the contrast of interest; the second comparison, with 2 degrees of freedom, is the residual of the sum of squares of $\mathcal{A}$ after subtraction of the sum of squares of the first contrast.

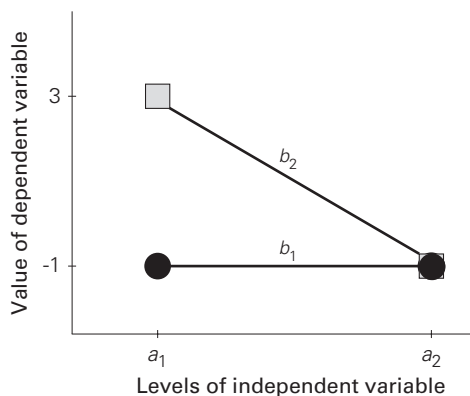


Figure 17.2 A prediction for a 2×2 design.

Experimental group				
	a_1b_1	a_1b_2	a_2b_1	a_2b_2
ψ	-1	3	-1	-1

Table 17.2 The coefficients of a contrast equivalent to the prediction described in Figure 17.2 for a 2×2 $S(\mathcal{A} \times \mathcal{B})$ design.

Experimental group				
	a_1b_1	a_1b_2	a_2b_1	a_2b_2
$\psi_1 : \mathcal{A}$	1	1	-1	-1
$\psi_2 : \mathcal{B}$	1	-1	1	-1
$\psi_3 : \mathcal{AB}$	1	-1	-1	1

Table 17.3 A set of three orthogonal contrasts equivalent to the canonical (standard) decomposition of a 2×2 $S(\mathcal{A} \times \mathcal{B})$ design.

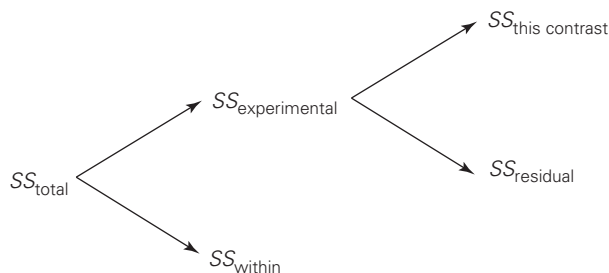


Figure 17.3 A partition of the sum of squares for a contrast which is not orthogonal to the standard decomposition of the experimental sum of squares of an $S(\mathcal{A} \times \mathcal{B})$ design into $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{AB}$.

17.5 What error term should be used?

17.5.1 The easy case: fixed factors

When the design involves only fixed factors, under the null hypothesis, the mean square of any comparison (no matter how many degrees of freedom it has) has the same expected value as the $\mathcal{S}(\mathcal{AB})$ sum of squares. Therefore, this mean square ($MS_{\mathcal{S}(\mathcal{AB})}$) should always be used as the test mean square (i.e. the denominator mean square) to compute the F ratios.

17.5.2 The harder case: one or two random factors

When random factors are involved, then the comparisons must be restricted to comparisons that partition the standard decomposition. In this case the mean square used for testing will be the same mean square that would be used to test the source that the contrast partitions. For example, in an $\mathcal{S}(\mathcal{A} \times \mathcal{B})$ design with $\mathcal{B}$ being a random factor, if a contrast partitions the sum of squares of $\mathcal{A}$, then it should be tested by dividing its mean square by the mean square of interaction ($MS_{\mathcal{AB}}$). If a contrast partitions the sum of squares of $\mathcal{B}$, then it should be tested by dividing its mean square by the ‘within group’ mean square ($MS_{\mathcal{S}(\mathcal{AB})}$).

17.6 Example: partitioning the standard decomposition

Let us go back to Table 17.1 (page 325) describing a set of orthogonal contrasts corresponding to the experiment of Tulving and Pearlstone. Remember that the first two contrasts decompose the effect of Factor $\mathcal{A}$ (list length). The first contrast opposes the first level to the second and third levels (12 word lists vs 24 and 48 word lists), and the second opposes the second level to the third one. The third contrast represents the effect of Factor $\mathcal{B}$ (presence or absence of cue at recall). Contrasts 4 and 5 represent the interaction of Contrast 1 and 2 with $\mathcal{B}$. The question asked by these contrasts is: does the effect of the contrast on the main effect of $\mathcal{A}$ change as a function of $\mathcal{B}$? For example, Contrast 4 explores the interaction of Contrast 1 with $\mathcal{B}$. Because Contrast 1 evaluates the difference between the short list (12 words) and the other lists (24 and 48 words), its interaction with $\mathcal{B}$ corresponds to the following question: does the difference between the first list and the other lists change as a function of the presence or absence of the cue? From Table 17.1, we can find that when the cue is absent at test (Condition b_1) the value of ψ_1 —which is the contrast opposing Condition a_1 to Condition a_2 and Condition a_3 —is equal to

$$(-2 \times 10) + (1 \times 13) + (1 \times 19) = -20 + 13 + 19 = 12.$$

When the cue is present at test (Condition b_2), however, the value of the contrast opposing Condition a_1 to Condition a_2 and Condition a_3 is equal to

$$(-2 \times 10) + (1 \times 15) + (1 \times 29) = -20 + 15 + 29 = 24.$$

So the difference between Condition a_1 and Conditions a_2 and a_3 is *larger* when there is a cue at test. This reveals an interaction between Contrast 1 and $\mathcal{B}$.

Some readers, at this point, may wonder why the coefficients of Contrast ψ_1 for b_1 are the *inverse* of those of b_2 . This comes from the null hypothesis: if there is *no* interaction between

Contrast 1 and $\mathcal{B}$, then the effect of Contrast 1 should be the same for b_1 and b_2 . Therefore the following equality should hold:

$$(-2 \times \mu_{11.}) + (1 \times \mu_{21.}) + (1 \times \mu_{31.}) = (-2 \times \mu_{12.}) + (1 \times \mu_{22.}) + (1 \times \mu_{32.}).$$

Rewriting this equality as a contrast gives:

$$(-2 \times \mu_{11.}) + (+1 \times \mu_{21.}) + (+1 \times \mu_{31.}) + (+2 \times \mu_{12.}) + (-1 \times \mu_{22.}) + (-1 \times \mu_{32.}) = 0.$$

This shows how the sign reverses.

17.6.1 Testing the contrasts

For each contrast, there is a null hypothesis that can be tested. Under the null hypothesis, the value of the contrast is equal to zero, and the value observed is a consequence of the error fluctuations. Therefore, the null hypothesis corresponding to a contrast can be evaluated using an F_ψ criterion computed as:

$$F_\psi = \frac{MS_\psi}{MS_{\text{error}}} = \frac{SS_\psi}{MS_{S(AB)}}. \quad (17.4)$$

For example, remembering that $MS_{S(AB)} = 9.00$, the null hypothesis for Contrast 1 is evaluated as:

$$F_{\psi_1} = \frac{SS_\psi}{MS_{S(AB)}} = \frac{1,080.00}{9.00} = 120.00.$$

Under the null hypothesis, the criterion F_{ψ_1} follows a Fisher distribution with $\nu_1 = 1$ and $\nu_2 = 54$ [why 54? Because $AB(S-1) = 2 \times 3 \times 9 = 54$]. If we want to use the critical value approach, we will find no critical value for these specific numbers of degrees of freedom, but the critical value for $\nu_1 = 1$ and $\nu_2 = 50$ will do. In the F_{critical} table we find for $\alpha = .01$ a value of 7.17. Because $F_{\psi_1} = 120.00 > F_{\text{critical}} = 7.17$, the probability of such a value occurring by chance alone is smaller than $\alpha = .01$. And so, we can reject the null hypothesis. Table 17.4 gives the detail of the contrast analysis.

Experimental group							SS_ψ	F_ψ
	a_1b_1	a_1b_2	a_2b_1	a_2b_2	a_3b_1	a_3b_2		
$M_{ab.}$	9	10	13	14	19	28		
ψ_1	-2	-2	1	1	1	1	1,080.00	120.00**
ψ_2	0	0	1	1	-1	-1	1,000.00	111.11**
ψ_3	1	-1	1	-1	1	-1	240.00	26.67**
ψ_4	-2	2	1	-1	1	-1	120.00	13.33**
ψ_5	0	0	1	-1	-1	1	160.00	17.78**

Table 17.4 A possible family of orthogonal contrasts for analyzing the $S(A \times B)$ of Tulving and Pearlstone (Section 16.7, page 303), along with their sums of squares and F . The sign ** (resp. *) on the right of an F value means that it is significant at the $\alpha = .01$ (resp. $\alpha = .05$) level. The error mean square is equal to 9. The number of subjects per group was 10. Contrasts 1 and 2 partition the sum of squares of $\mathcal{A}$. Contrast 3 expresses the sum of squares of $\mathcal{B}$. Contrasts 4 and 5 partition the sum of squares of $\mathcal{AB}$. The coefficients for Contrasts ψ_4 and ψ_5 are obtained by multiplication of the coefficients of (respectively) Contrasts ψ_1 by ψ_3 , and ψ_2 by ψ_3 .

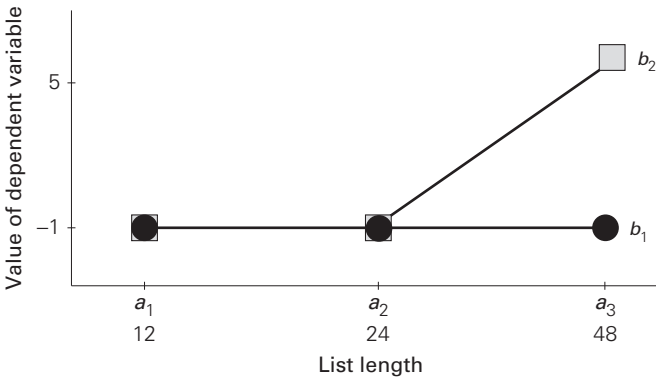


Figure 17.4 A prediction for a 3×2 design (like the Tulving and Pearlstone example). This contrast does *not* constitute a subpartition of the standard decomposition.

17.7 **Example: a contrast non-orthogonal to the standard decomposition**

Consider the prediction drawn in Figure 17.4. It states that the cue should have an effect only when subjects learn 48-word lists. It corresponds to the following contrast:

		Experimental group					
		a_1b_1	a_1b_2	a_2b_1	a_2b_2	a_3b_1	a_3b_2
$M_{a,b}$	ψ	10	10	13	15	19	29
		-1	-1	-1	-1	-1	5

This contrast is not a contrast on the main effect of $\mathcal{A}$ (because it does not give the same coefficients to groups with the same value of the index a), neither is it a contrast on the main effect of $\mathcal{B}$ (same reason but for $\mathcal{B}$). But nor is this a contrast on the interaction $\mathcal{AB}$, as can be checked in the following table:

	a_1	a_2	a_3	$\sum_a C_{ab}$
b_1	$C_{1,1} = -1$	$C_{2,1} = -1$	$C_{3,1} = -1$	$\sum_a C_{a,1} = -3$
b_2	$C_{1,2} = -1$	$C_{2,2} = -1$	$C_{3,2} = 5$	$\sum_a C_{a,2} = -3$
$\sum_b C_{ab}$	$\sum_b C_{1,b} = -2$	$\sum_b C_{2,b} = -2$	$\sum_b C_{3,b} = 4$	0

But we still want to test this contrast in lieu of the standard decomposition. The first step is to compute the estimated value of this contrast:

$$\begin{aligned}\widehat{\psi} &= \sum_a \sum_b C_{ab} M_{ab} \\ &= (-1 \times 10) + (-1 \times 10) + (-1 \times 13) + (-1 \times 15) + (-1 \times 19) + (5 \times 29) \\ &= 78.00\end{aligned}$$

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p(F)</i>	<i>v</i> ₁	<i>v</i> ₂
ψ	1	2,028.00	2,028.00	225.33	<i>Very very small!</i>	1	54
Residual of experimental Error [<i>S</i> (<i>AB</i>)]	4	572.00	143.00	15.89	.000000023	4	54
Total	59	3,086.00					

Table 17.5 ANOVA table for a contrast used in lieu of the standard decomposition.

Using the formula given in Equation 17.2, we find that:

$$SS_{\psi} = \frac{S\hat{\psi}^2}{\sum C_{ab}^2} = \frac{10 \times (78.00)^2}{30} = \frac{60,840.00}{30} = 2,028.00.$$

Because this contrast is not orthogonal to the standard decomposition, it corresponds to a different decomposition of the experimental sum of squares. Specifically, the experimental sum of squares is now partitioned into two sums of squares: this contrast sum of squares and the residual sum of squares which is obtained by subtraction. The experimental sum of squares is equal to

$$SS_{\text{experimental}} = SS_A + SS_B + SS_{AB} = 2,080.00 + 240.00 + 280.00 = 2,600.00.$$

The residual sum of squares is obtained by subtraction:

$$SS_{\text{residual}} = SS_{\text{experimental}} - SS_{\psi} = 2,600.00 - 2,028.00 = 572.00.$$

This residual sum of squares has 4 degrees of freedom ($5 - 1$, where 5 is the number of degrees of freedom of the experimental sum of squares minus one degree of freedom for our contrast). Therefore the residual sum of squares is:

$$MS_{\text{residual}} = \frac{SS_{\text{residual}}}{df_{\text{residual}}} = \frac{572.00}{4} = 143.00.$$

From these values we can build the ANOVA table used in lieu of the standard decomposition ANOVA table (see Table 17.5).

17.8 *A posteriori* comparisons

The *a posteriori* comparisons follow the same pattern that we have seen before and therefore they can be extended in a straightforward manner. We just need to adjust properly the number of observations and the degrees of freedom when computing mean squares and looking for critical values. For example, with an $\mathcal{A} \times \mathcal{B}$ design with $A = 3$ and $B = 4$, a Scheffé test for all possible contrasts will be tested with a critical F with $v_1 = AB - 1$ and $v_2 = AB(S - 1)$. If the comparisons are restricted to the main effect of one factor (say $\mathcal{A}$) the critical value for Scheffé test will be obtained with $v_1 = A - 1$ and $v_2 = AB(S - 1)$.

Chapter summary

17.9 Key notions of the chapter

Below are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

Main effect and interaction contrasts	Sub-design analysis
Interaction contrasts are obtained as the product of main effect contrasts	Standard decomposition

17.10 Key questions of the chapter

Below are some questions about the content of this chapter. All the answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- * How can we check that a contrast is a main effect contrast?
- * How can we check that a contrast is an interaction contrast?
- * When should we use contrasts instead of the standard decomposition?

18

ANOVA, one-factor repeated measures design: $S \times \mathcal{A}$

18.1 Introduction

For the previous experimental designs [$S(\mathcal{A})$ and $S(\mathcal{A} \times \mathcal{B})$], we insisted on the fact that a subject participated in only *one* condition of the experiment (i.e. only one measurement was recorded per subject). There are, however, some experimental designs in which several measures are recorded for each subject. In this chapter, we will present the specific case in which subjects participate in all the conditions of the experiment (one subject provides a score for each level of the independent variable). This design is generally referred to as a *within-subjects* design or a *repeated measures* design. It is symbolized as: $S \times \mathcal{A}$ design (read ' S by $\mathcal{A}$ design').

18.2 Advantages of repeated measurement designs

Repeated measurement designs are frequently used in studies dealing with learning, learning transfer, or effects linked to habituation phenomena. In these types of studies, the independent variable expresses explicitly the notion of time or succession. Note that these learning phenomena can be a problem if they are not taken explicitly into account in the experimental design (i.e. they can act as a *confound*). We have already encountered this point in the chapter dedicated to multi-factorial designs (see Chapter 15, Section 15.5, page 287).

Repeated measures designs are not restricted to experiments involving the notions of time or repetition. They are also used in other types of experiments because of their greater *sensitivity* (i.e. the fact that they increase the probability of correctly rejecting the null hypothesis). This greater sensitivity is due to the fact that they allow for the reduction of the experimental error. Recall that, intuitively, the criterion F can be expressed as

$$F = \frac{\text{Effect of independent variable} + \text{Experimental error}}{\text{Experimental error}}.$$

The estimation of the experimental error appears in the denominator of this ratio, and therefore the smaller the error, the greater the F (everything else being equal), and the easier it will be to reject the null hypothesis.

An additional bonus of the use of repeated measurements is that, from a strictly practical point of view, they have the advantage of requiring smaller numbers of subjects than independent or between-subjects designs. To emphasize and illustrate this point, let us compare a between-subjects design (independent measures) and a within-subjects design (repeated measures). Suppose that the independent variable is made of 3 levels (a_1, a_2, a_3) and that 4 observations are recorded for each level. The subject assignment for the between-subjects design is described in the table below:

Levels of the independent variable		
a_1	a_2	a_3
s_1	s_5	s_9
s_2	s_6	s_{10}
s_3	s_7	s_{11}
s_4	s_8	s_{12}

For the within-subjects design, the following table illustrates the subject assignment.

Levels of the independent variable		
a_1	a_2	a_3
s_1	s_1	s_1
s_2	s_2	s_2
s_3	s_3	s_3
s_4	s_4	s_4

There are 12 subjects in the case of the between-subjects design but only 4 subjects for the within-subjects design.

Note that compared to the $\mathcal{S}(\mathcal{A})$ design, the notation Y_{as} takes on a slightly different meaning here. Y_{as} is the score of individual s for level a of the independent variable, and s refers to the *same* subject across all the levels of $\mathcal{A}$.

An equivalent way of looking at a within-subjects design is to imagine it as a specific case of a two-factor design: $\mathcal{S}$ and $\mathcal{A}$ (where $\mathcal{S}$ is a random factor).¹ However, unlike the $\mathcal{A} \times \mathcal{B}$ design where there are multiple scores for each ab condition, there is only one score for each combination of factors (level of $\mathcal{A}$ and subject $\mathcal{S}$). Since there is only one score for each as condition, the interaction between $\mathcal{S}$ and $\mathcal{A}$ is *confounded* with the experimental error.

¹ By saying that $\mathcal{S}$ is a random factor, we mean that subjects are assumed to be a representative *random* sample of the population of all possible subjects, and that the statistical conclusions that we want to reach should be valid for the whole population and not only for the specific sample used in the experiment.

18.3 Examination of the F ratio

In this section, we show how to estimate the sum of squares of error in a repeated measurement design. Recall that unlike the between-subjects design, the same subjects are assigned to all conditions. Therefore the main idea is to subtract from the within sum of squares the sum of squares representing the variability due to the subjects.

Let us start by a brief refresher: in an $\mathcal{S}(\mathcal{A})$ design, the total sum of squares is partitioned into two parts:

$$SS_{\text{total}} = SS_A + SS_{S(A)}.$$

SS_A expresses the effect of the independent variable, and $SS_{S(A)}$ expresses the effect of the experimental error.

The same components can be isolated in the case of a repeated measures design. However, although SS_A can still be used to express the effect of the independent variable, $SS_{S(A)}$ cannot be used any longer to estimate the error. The statistic $SS_{S(A)}$ *overestimates* the value of the experimental error in the case of repeated measures. To understand this point, it suffices to note that the variability due to the difference between subjects from one condition to the other no longer exists. To convince yourself, imagine the case—indeed a caricature—in which the independent variable does not have any effect at all, and in which—in addition—the experimental error is equal to zero.

For example, take a look at the pattern of results shown in Table 18.1. It is clear from this table that the variability within groups, expressed by $SS_{S(A)}$, depends only upon the individual differences or the variability *between subjects*. Hence, $SS_{S(A)}$ obviously overestimates the experimental error.

The example shown in Table 18.2 illustrates the results of a hypothetical experiment where there is an effect of the independent variable $\mathcal{A}$. We can imagine that a_1 represents the control group, scores in the condition a_2 are 1 point above the scores in the control condition, and scores for a_3 are 2 points above these control scores. Once again it is clear that $SS_{S(A)}$ is far from being a good estimation of the experimental error (which is in both examples equal to zero).

In the next section we show that in a within-subjects design the experimental error is smaller than it is in a between-subjects design.

Subject	Levels of the independent variable			
	a_1	a_2	a_3	M_s
s_1	8.00	8.00	8.00	8.00
s_2	5.00	5.00	5.00	5.00
s_3	6.00	6.00	6.00	6.00
s_4	9.00	9.00	9.00	9.00
$M_{a\cdot}$	7.00	7.00	7.00	7.00

Table 18.1 Results of a hypothetical experiment using an $S \times \mathcal{A}$ design.

Subject	Levels of the independent variable			
	a_1	a_2	a_3	$M_{\cdot s}$
s_1	5.00	6.00	7.00	6.00
s_2	2.00	3.00	4.00	3.00
s_3	3.00	4.00	5.00	4.00
s_4	6.00	7.00	8.00	7.00
$M_{a\cdot}$	4.00	5.00	6.00	5.00

Table 18.2 Results of a second hypothetical experiment using an $S \times \mathcal{A}$ design.

18.4 Partition of the within-group variability: $S(\mathcal{A}) = S + \mathcal{AS}$

Observation of Tables 18.1 and 18.2 indicates that part of the variability within the experimental groups—denoted $S(\mathcal{A})$ —can be attributed to the difference *between subjects* (i.e. the effect of factor S). For example, in both tables, the mean of the first subject is 1 above the grand mean. The mean of the second subject is at 2 below the grand mean. The subjects' idiosyncrasies appear also in the results obtained for each of the experimental conditions.

If we suppose that the pattern of differences between subjects is similar in each experimental condition, it is possible to estimate these differences from the deviations of the subject means from the grand mean:

$$M_{\cdot s} - M_{\cdot\cdot}$$

Note that we just introduced a new notation,

$$M_{\cdot s} = \frac{Y_{\cdot s}}{A}$$

which is read as the 'mean of subject s '.

Hence, two sources of variability can be identified within an experimental group:

- the difference between the subjects,
- the experimental error.

Recall that the effect of an experimental treatment is tested by dividing the *between-group mean square* (i.e. the experimental variance) by the *mean square estimating the experimental error* (remember that 'under the null hypothesis' the mean square of $\mathcal{A}$ reflects the experimental error because there is *no* effect of $\mathcal{A}$). So, in general, an F looks like:

$$F_{\text{something}} = \frac{MS_{\text{something}}}{MS_{\text{error}}}$$

For the one-factor between-subjects designs (or independent measures), the error is estimated by the within-group mean square [$S(\mathcal{A})$]. For a one-factor within-subjects design (or repeated measures), to estimate the experimental error, the 'variability due to the subjects' (i.e. the difference between subjects) should be subtracted first from the 'within-group variability'. Thus the experimental error corresponds to what remains of the within-group variability

once the variability due to the subjects has been taken into account. In other words, the experimental error is estimated by the ‘residual’ (of the within sum of squares after the subject sum of squares has been subtracted).

Thus, the sum of the squares within groups (i.e. $SS_{S(A)}$) can be decomposed into two quantities: one due to the subject factor, the other one estimating the error (i.e. the residual). The general approach is similar to the one previously used to partition the total sum of squares into two parts: A and $S(A)$. We start by noting that:

$$\begin{array}{lll} (Y_{as} - M_{a\cdot}) & = & (M_{\cdot s} - M_{\cdot\cdot}) \quad + \quad (Y_{as} - M_{a\cdot} - M_{\cdot s} + M_{\cdot\cdot}) \\ \text{within-group} & & \text{between-subject} \quad \text{residual} \\ \text{deviation} & & \text{deviation} \quad \text{deviation} \end{array}$$

The ‘residual’ is simply obtained by subtraction:

$$\text{residual} = \text{within-group deviation} - \text{between-subject deviation}$$

(you may have noticed that the residual corresponds to the interaction between A and S).

The sum of squares within groups is then obtained by squaring and summing over a and s the deviations of the scores from their group means. With a formula this is written as:

$$\begin{aligned} SS_{S(A)} &= \sum (Y_{as} - M_{a\cdot})^2 \\ &= \sum [(M_{\cdot s} - M_{\cdot\cdot}) + (Y_{as} - M_{a\cdot} - M_{\cdot s} + M_{\cdot\cdot})]^2 . \end{aligned}$$

By expanding the squared term between the brackets and distributing the $\sum$ sign, this equation becomes:

$$\begin{aligned} SS_{S(A)} &= \sum (M_{\cdot s} - M_{\cdot\cdot})^2 + \sum (Y_{as} - M_{a\cdot} - M_{\cdot s} + M_{\cdot\cdot})^2 \\ &\quad + 2 \sum (Y_{as} - M_{a\cdot} - M_{\cdot s} + M_{\cdot\cdot})(M_{\cdot s} - M_{\cdot\cdot}) . \end{aligned}$$

Because

$$\sum (Y_{as} - M_{a\cdot} - M_{\cdot s} + M_{\cdot\cdot})(M_{\cdot s} - M_{\cdot\cdot}) = 0$$

[try to prove it as an exercise—the technique is the same as the one used previously to show the additivity of SS_A and $SS_{S(A)}$], the sum of squares within groups reduces to:

$$\begin{array}{ll} \sum (Y_{as} - M_{a\cdot})^2 & = A \sum (M_{\cdot s} - M_{\cdot\cdot})^2 \\ \text{sum of the squares} & \text{sum of the squares} \\ \text{within group} & \text{between subjects} \\ & + \sum (Y_{as} - M_{a\cdot} - M_{\cdot s} + M_{\cdot\cdot})^2 \\ & \text{residual} \\ & \text{sum of squares} . \end{array} \quad (18.1)$$

The sum of squares ‘between subjects’ expresses the effect of the subject factor S . It is denoted by SS_S . The ‘residual’ sum of squares corresponds to the interaction between factor A and factor S . It is denoted by SS_{AS} . With these notations, the partition of the within-groups sum of squares can be rewritten as:

$$SS_{S(A)} = SS_S + SS_{AS} .$$

18.5 Computing F in an $S \times A$ design

The criterion F is similar to the one used in a between-subjects design with the exception that MS_{AS} will be used to estimate the error instead of $MS_{S(A)}$. Hence, the effect of A is tested with the following F ratio:

$$F_A = \frac{MS_A}{MS_{AS}}.$$

To compute the mean square of error MS_{AS} we first need to find the number of degrees of freedom associated with the residual sum of squares SS_{AS} (recall that a mean square is obtained by dividing a sum of squares by its degrees of freedom). We already know that (from Chapter 7, Section 7.3.6, page 140)

$$df_A = A - 1$$

using the same principle as before we can deduce that

$$df_S = S - 1,$$

and we have shown that

$$SS_{\text{total}} = SS_S + SS_A + SS_{AS}.$$

Remember that when the sums of squares are added together, the degrees of freedom are also added. Hence,

$$df_{\text{total}} = df_S + df_A + df_{AS}.$$

From that we can derive that

$$df_{AS} = df_{\text{total}} - df_A - df_S,$$

and hence,

$$\begin{aligned} df_{AS} &= (AS - 1) - (A - 1) - (S - 1) \\ &= (A - 1)(S - 1) \end{aligned}$$

and

$$MS_{AS} = \frac{SS_{AS}}{df_{AS}} = \frac{SS_{AS}}{(A - 1)(S - 1)}.$$

An example should help us ‘digest’ this formula.

18.6 Numerical example: $S \times A$ design

Let us suppose that in a hypothetical experiment a within-subjects design was used. The independent variable consists of 4 levels and the size of the experimental group is 5 (i.e. 5 subjects participated in the experiment; how many subjects would we need for a between-subjects design?). The results are presented in Table 18.3.

Subject	Levels of the independent variable				
	a_1	a_2	a_3	a_4	$M_{.s}$
s_1	31	30	23	24	27
s_2	24	25	20	23	23
s_3	24	18	16	18	19
s_4	18	18	16	12	16
s_5	18	14	15	13	15
$M_{a.}$	23	21	18	18	20

Table 18.3 A numerical example of an $S \times A$ design

Before starting to compute the F ratio, don't forget the following preliminary steps:

1. Statistical hypothesis.
2. Choice of a test.
3. Sampling distribution.
4. Significance level.
5. Decision rule and rejection area.

After completing those steps, try to compute the F ratio. You should obtain the following results.

For the sum of squares:

$$\begin{aligned}
 SS_A &= S \sum (M_{a.} - M_{..})^2 \\
 &= 5[(23.00 - 20.00)^2 + (21.00 - 20.00)^2 + (18.00 - 20.00)^2 + (18.00 - 20.00)^2] \\
 &= 5[3.00^2 + 1.00^2 + (-2.00)^2 + (-2.00)^2] \\
 &= 5[9 + 1 + 4 + 4] \\
 &= 90.00
 \end{aligned}$$

$$\begin{aligned}
 SS_S &= A \sum (M_{.s} - M_{..})^2 \\
 &= 4[(27.00 - 20.00)^2 + (23.00 - 20.00)^2 + (19.00 - 20.00)^2 \\
 &\quad + (16.00 - 20.00)^2 + (15.00 - 20.00)^2] \\
 &= 4[7.00^2 + 3.00^2 + (-1.00)^2 + (-4.00)^2 + (-5.00)^2] \\
 &= 4[49.00 + 9.00 + 1.00 + 16.00 + 25.00] \\
 &= 400.00
 \end{aligned}$$

$$\begin{aligned}
 SS_{AS} &= \sum (Y_{as} - M_{a.} - M_{.s} + M_{..})^2 \\
 &= (31.00 - 23.00 - 27.00 + 20.00)^2 + (24.00 - 23.00 - 23.00 + 20.00)^2 + \cdots \\
 &\quad + \cdots + (12.00 - 18.00 - 16.00 + 20.00)^2 + (13.00 - 18.00 - 15.00 + 20.00)^2 \\
 &= 1.00^2 + (-2.00)^2 + \cdots + (-2.00)^2 + 0.00^2 \\
 &= 48.00 .
 \end{aligned}$$

Numbers of degrees of freedom:

$$df_A = A - 1 = 4 - 1 = 3$$

$$df_S = S - 1 = 5 - 1 = 4$$

$$df_{AS} = (S - 1)(A - 1) = (5 - 1)(4 - 1) = 12.$$

Mean squares:

$$MS_A = \frac{SS_A}{df_A} = \frac{90.00}{3} = 30.00$$

$$MS_S = \frac{SS_S}{df_S} = \frac{400.00}{4} = 100.00$$

$$MS_{AS} = \frac{SS_{AS}}{df_{AS}} = \frac{48.00}{12} = 4.00$$

Statistical index F :

$$F_A = \frac{MS_A}{MS_{AS}} = \frac{30.00}{4.00} = 7.50$$

With $\nu_1 = 3$, $\nu_2 = 12$, and $\alpha = .01$, we find a value of $F_{\text{critical}} = 5.95$. Therefore, we could conclude that, in this experiment, there is a significant effect of the independent variable on the dependent variable, $F(3, 12) = 7.50$, $MS_e = 4.00$, $p < .01$.

Note that, although only two quantities are necessary to compute the criterion, MS_A and MS_{AS} , we have actually partitioned the total sum of squares into three additive (or orthogonal) parts: $\mathcal{A}$, S , and $\mathcal{AS}$. Therefore it seems natural to present the details of this partitioning in an ANOVA table such as Table 18.4. This table makes it clear that an $S \times \mathcal{A}$ design is actually a specific case of a two-factor design.

18.6.1 An alternate way of partitioning the total sum of squares

The total variability may also be approached from the subject perspective instead of the group perspective, separating the *between-subjects variability* from the *within-subjects variability*. In this case, the variability between subjects is expressed by SS_S . The variability within subjects can be considered as being composed of two quantities: the effect of the independent variable and the experimental error (i.e. the interaction between the subject and the independent variable). This is equivalent to saying that the difference between a subject's score across experimental conditions can be attributed to two causes: the effect of the

Source	df	SS	MS	F	$p(F)$
$\mathcal{A}$	3	90.00	30.00	7.50	.0045
S	4	400.00	100.00		
$\mathcal{AS}$	12	48.00	4.00		
Total	19	538.00			

Table 18.4 ANOVA table for an $S \times \mathcal{A}$ design.

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Between subjects	4	400.00	100.00	
Within subjects:				
Treatment	3	90.00	30.00	7.50**
Residual	12	48.00	4.00	
Total	19	538.00		

** $p < .01$

Table 18.5 An alternative ANOVA table for an $S \times A$ design.

independent variable and the interaction between the subject and the independent variable. With a formula we have:

$$SS_{\text{within-subjects}} = SS_A + SS_{AS}.$$

This partition is often mentioned in old texts. Following this perspective, the results of our hypothetical experiment would be presented as illustrated in Table 18.5.

Look carefully at Tables 18.4 and 18.5 to understand the meaning of the notation and the equivalence of the two types of presentation.

18.7 Score model: Models I and II for repeated measures designs

The distinction between Model I and Model II previously described for independent measures designs applies here also. Recall that for Model II (both factors random) the estimation of the ‘error variance’ is no longer the within-group mean square $MS_{S(A)}$, but the mean square of the interaction, MS_{AS} . To help you review this material why don’t you try to prove this to yourself?

18.8 Effect size: R , R , and R

In order to make the explanation in terms of the coefficients of correlation smoother, we can introduce some notations parallel to the ones used in Chapter 16 on $S(A \times B)$.

In an $S \times A$ design, we decompose the deviation from any score to the grand mean as:

$$(Y_{as} - M..) = (M_{a.} - M..) + (M_{.s} - M..) + (Y_{as} - M_{a.} - M_{.s} + M..),$$

which can be rewritten as

$$Y_{as} = M.. + (M_{a.} - M..) + (M_{.s} - M..) + (Y_{as} - M_{a.} - M_{.s} + M..).$$

We can now denote each component of a score as

- $M..$ the grand mean.
- $A_a = (M_{a.} - M..)$ the effect of the a th level of A .
- $S_s = (M_{.s} - M..)$ the effect of the s th level of S (i.e. of the s th subject).

- $\mathcal{AS}_{as} = (Y_{as} - M_{a\cdot} - M_{\cdot s} + M_{\cdot\cdot})$ the effect of the interaction between the a th level of $\mathcal{A}$ and the s th level of $\mathcal{S}$ (i.e. of the s th subject).

With these notations, a score Y_{as} is expressed as

$$Y_{as} = M_{\cdot\cdot} + \mathcal{A}_a + \mathcal{S}_s + \mathcal{AS}_{as}.$$

The coefficient $R^2_{Y\cdot A}$, which was previously defined for the $\mathcal{S}(\mathcal{A})$ design, can still be used here. Because we have three sources of variation in an $\mathcal{S} \times \mathcal{A}$ design, we have also three possible coefficients. The general interpretation of the coefficient remains the same as previously seen. Specifically, $R^2_{Y\cdot \text{source of variation}}$ is the correlation between the scores actually observed and the prediction derived from the knowledge of the source of variation in question. Hence:

- The coefficient

$$R^2_{Y\cdot A} = \frac{SS_A}{SS_{\text{total}}}$$

is the correlation between Y_{as} and $\mathcal{A}_a = (M_{a\cdot} - M_{\cdot\cdot})$.

- The coefficient

$$R^2_{Y\cdot S} = \frac{SS_S}{SS_{\text{total}}}$$

is the correlation between Y_{as} and $\mathcal{S}_s = (M_{\cdot s} - M_{\cdot\cdot})$.

- The coefficient

$$R^2_{Y\cdot AS} = \frac{SS_{AS}}{SS_{\text{total}}}$$

is the correlation between Y_{as} and $\mathcal{AS}_{as} = (Y_{as} - M_{a\cdot} - M_{\cdot s} + M_{\cdot\cdot})$.

The procedure detailed in Chapter 16 linking the regression technique and the analysis of variance via the coefficients of correlation is indeed still valid. (Try to do it as an additional exercise.) The formula for the F ratio can also be rewritten in terms of the coefficient of correlation as

$$F = \frac{R^2_{Y\cdot A}}{R^2_{Y\cdot AS}} \times \frac{df_{AS}}{df_A}.$$

(The derivation is left as an exercise, because it is almost the same as the one seen in Chapter 16.)

18.9 Problems with repeated measures

Repeated measurement designs have a lot of advantages, but the simple fact of using the same subjects creates potential problems which may threaten the validity of these designs. In this section, we describe some of these potential problems.

18.9.1 Carry-over effects

As already mentioned, repeated measurement designs present a number of advantages: a smaller number of subjects, better sensitivity (i.e. greater power) and so on. However, the application of such a design implies that no carry-over should occur from one condition to the other. This would be the case, for example, if we were using a repeated measurement design to study the quality of several reading methods: clearly, a child who already knows how to read cannot be compared with a novice child (i.e. a child who has not yet learned how to read).

When such problems occur, experimental validity becomes questionable, since it is impossible to separate the effect of the independent variable from the effect of the carry-over (i.e. the effects of carry-over and of the independent variable are confounded). Randomizing the order in which the stimuli are presented to the subject is, unfortunately, not sufficient to prevent the risk of carry-over.

18.9.2 Pre-test and post-test

Imagine that we wish to reduce racial prejudice (which is a worthy goal). To do this, we have developed a procedure (which is here left to your imagination, but which is based on the evidence reviewed in Lindsey and Aronson, 1968). We wish to evaluate the effectiveness of our manipulation. The following experimental design comes quite naturally to mind. Quite fortuitously we have at hand a test of racism (for example, the measurement scales developed in the work of Adorno *et al.*, 1950). To begin with, we measure the amount of prejudice of our subjects. Then we subject them to our experimental manipulation designed to reduce prejudice. Finally we measure their amount of prejudice again. This seems like a straightforward way to proceed.

Alas...

Suppose that, in effect, we observe a reduction (or an increase) in prejudice for our subjects. Have we demonstrated the effectiveness (or the 'counter-effectiveness') of our manipulation? Well maybe not! The result may be due to some confounded independent variables. Some of these (confounded) variables are listed below (see also Cozby, 1977; Cook and Campbell, 1979):

History: A whole series of events separates the pre-test from the post-test, which could cause the observed effect (or prevent its appearance, or interact with the dependent variable). For example, the showing of a sensational film accompanied with discussions in the media, or sensational news events, or an economic crisis, etc.

Maturation of the subjects: As time passes, people change (not a great discovery). This type of problem, always present, is felt particularly strongly in developmental studies in which it is often difficult to separate experimental effects from the effects of maturation of the subjects.

Sensitization by the test: Simply having been tested affects the subjects. In particular, it could make the subjects more aware of their attitudes, and so lead some of them to change those attitudes (for example, to be more in accord with what they believe the experimenter wants, to correspond more closely to their self-image, or for other reasons). This type of problem is most likely to arise in studies making use of a panel of subjects (that is, the same group of subjects tested or interviewed several times over a rather long time period). The subjects, expecting to be tested, tend to adopt behavior appropriate to 'experts': they seek out information, try to learn about the topic, etc.

18.9.3 Statistical regression, or regression toward the mean

Imagine that in order to show clearly the effect of your treatment, you decide to select your subjects so as to concentrate on the ones most in need of a change. You use the pre-test to select a group of definitely racist subjects. You observe a reduction in prejudice on the post-test. Can this be attributed to the effects of your treatment? No! This reduction can be attributed to the phenomenon of ‘regression toward the mean’. This problem occurs because any measure includes some error. If we select subjects whose pre-test scores are extreme (in relation to the mean), we thereby select subjects with scores whose error components are maximal (why?), and in fact carrying their scores away from the mean. On a re-test, these errors will tend to diminish, and the subjects’ scores will appear to move toward the mean, giving the false impression of an effect of the treatment. To understand this phenomenon, imagine that a score on a test (or some such measure) can be seen as composed of two parts: the quantity itself that we were trying to measure (and that we never observe directly), and a random error component. In choosing subjects with extreme measures we are choosing those for whom the error component is large. To be specific, we are choosing subjects for whom both the original measurement *and* the error of measurement are large. Because of this, the measured quantity appears larger than it really is. On re-test, the random error component will be on the average smaller (because large errors are rarer than small errors). Consequently the score observed on the post-test (which is composed of the quantity measured plus an error that is now on the average smaller) will lie closer to the mean (on the average).

18.10 Score model (Model I) $S \times A$ design: A fixed

The score of subject s in experimental condition a can be decomposed as:

$$Y_{as} = \mu_{..} + \alpha_a + s_s + \alpha s_{as} + e_{as}$$

with α_a : the effect of the a th level of A

s_s : the effect of the subject s
(sth level of the random subject Factor S)

αs_{as} : the effect of the a th level of A and the subject s

e_{as} : the experimental error associated with the condition as
(confounded with the interaction AS)

with the following conditions:

$$\sum \alpha_a = 0, \quad \sum \alpha_a^2 = (A-1)\vartheta_a^2 \quad \sum_a \alpha s_{as} = 0$$

$$\mathbf{E}\{s_s\} = 0, \quad \mathbf{E}\{s_s^2\} = \sigma_s^2$$

$$\mathbf{E}\{\alpha s_{as}\} = 0, \quad \mathbf{E}\{\alpha s_{as}^2\} = \sigma_{as}^2$$

$$\mathbf{E}\{e_{as}\} = 0, \quad \mathbf{E}\{e_{as}^2\} = \sigma_e^2.$$

The previously described conditions of normality and homogeneity of variance also apply, along with a new condition called ‘*sphericity*’ (also called ‘*circularity*’ or ‘*homogeneity of covariance*’). This last condition means, intuitively, that the correlation (calculated for each subject) between two experimental conditions is constant for all pairs of conditions chosen. This condition is described in more detail in Section 18.12 (page 348).

Under the usual rules, the following mathematical expectations can be found for the different mean squares:

$$\mathbf{E}\{MS_A\} = \sigma_e^2 + \sigma_{as}^2 + S\vartheta_a^2$$

$$\mathbf{E}\{MS_S\} = \sigma_e^2 + A\sigma_s^2$$

$$\mathbf{E}\{MS_{AS}\} = \sigma_e^2 + \sigma_{as}^2.$$

Under the null hypothesis of the absence of an effect of $\mathcal{A}$, the mean square of $\mathcal{A}$ and the mean square of $\mathcal{AS}$ are estimated by the same quantity $\sigma_e^2 + \sigma_{as}^2$. When the conditions specified above are satisfied, the ratio of these quantities follows a Fisher distribution with

$$\nu_1 = (A - 1) \quad \text{and} \quad \nu_2 = (A - 1)(S - 1) \quad \text{degrees of freedom.}$$

This justifies the formula for the calculation of F given previously.

As in all other designs, the sums of squares can be represented by estimations of the score model components.

$$SS_A = S \sum (M_{a.} - M_{..})^2 = S \sum \text{est} \{\alpha_a^2\}$$

$$SS_S = A \sum (M_{.s} - M_{..})^2 = A \sum \text{est} \{s_s^2\}$$

$$SS_{AS} = \sum (Y_{as} - M_{a.} - M_{.s} + M_{..})^2 = \sum \text{est} \{\alpha s_{as}^2\}.$$

Note: If you study the score model you will notice that (as usual) the subject factor is random. You may also notice a new interaction term αs . This interaction term means that each subject may react to each experimental condition in a specific (idiosyncratic) way. This interaction cannot be separated from the experimental error, because there is only one score per subject in each condition. The fact that the indices of the interaction term (αs) and the error term (e) are the same (as), points this out: these two quantities vary simultaneously, they are *confounded*. And this is why they are represented as two different quantities with the same indices. As a consequence of this confound the experimental error cannot be estimated separately from the interaction term. A conservative estimation of the error can be obtained, however, by assuming that the term αs is null.

18.11 Score model (Model II) $\mathcal{S} \times \mathcal{A}$ design: $\mathcal{A}$ random

The score of subject s in experimental condition a can be decomposed as

$$Y_{as} = \mu_{..} + a_a + s_s + as_{as} + e_{as}$$

with the following conditions:

$$\mathbf{E}\{a_a\} = 0, \quad \mathbf{E}\{a_a^2\} = \sigma_a^2$$

$$\mathbf{E}\{s_s\} = 0, \quad \mathbf{E}\{s_s^2\} = \sigma_s^2$$

$$\mathbf{E}\{as_{as}\} = 0, \quad \mathbf{E}\{as_{as}^2\} = \sigma_{as}^2$$

$$\mathbf{E}\{e_{as}\} = 0, \quad \mathbf{E}\{e_{as}^2\} = \sigma_e^2.$$

Following the usual rules, the mathematical expectations can be found for the different mean squares as

$$\mathbf{E}\{MS_A\} = \sigma_e^2 + \sigma_{as}^2 + S\sigma_a^2$$

$$\mathbf{E}\{MS_S\} = \sigma_e^2 + \sigma_{as}^2 + A\sigma_s^2$$

$$\mathbf{E}\{MS_{AS}\} = \sigma_e^2 + \sigma_{as}^2.$$

Note: When $\mathcal{A}$ is random, examination of the mathematical expectations shows that the null hypothesis of no effect of $\mathcal{S}$ (like that for $\mathcal{A}$) can be tested by dividing the mean square of $\mathcal{S}$ by the mean square of $\mathcal{AS}$, because the mathematical expectations are identical. However, this particular test for an effect of $\mathcal{S}$ is rarely used; simply because, in general, the null hypothesis for $\mathcal{S}$ (there is no difference between subjects) is known or assumed to be false.

18.12 A new assumption: sphericity (circularity)

In addition to the usual assumptions of normality of the error and homogeneity of variance, the F test for repeated measurement designs assumes a new condition called ‘sphericity’ (a variation of sphericity is called ‘circularity’). In this section, we describe this condition first intuitively and then more formally. Then we suggest some possible palliative actions when we suspect that the sphericity condition is not valid.

18.12.1 Sphericity: intuitive approach

Under the sphericity assumption, we postulate that the subjects’ behavior is fairly constant through the different levels of the variable (e.g. a ‘good’ subject will be good in every condition, and a ‘bad’ subject will be bad in every condition). This implies that we should find the same value for the coefficient of correlation between any pair of conditions. In other words, we assume that there is no interaction between the subjects and the independent variable (an example of interaction would be that one subject performs below the average in one condition and above the average in another condition). Recall that MS_{AS} which represents this interaction is used as the error term in the computation of F . If the interaction effect is large it will mask the effects we are trying to uncover. We describe below five ways of detecting such perverse effects.

- First. Examine the raw data carefully, and compare the subjects’ scores in each level of the independent variable. A more thorough examination can also be done by looking at the ‘residuals’ (i.e. the interaction between subjects and conditions) instead of the raw data:

$$Y_{as} - M_{a\cdot} - M_{\cdot s} + M_{\cdot\cdot}$$

- Second. Plot the interaction as was explained in the previous chapter.
- Third. Compare the different sums of squares (using the coefficients of correlation described earlier on in this chapter, for example).
- Fourth (by the way, this is the best.) Compute the covariance (or correlation) between all the possible pairs of experimental conditions. A wide difference between these covariances (or the correlations) indicates that the assumptions of validity are probably not met.
- Fifth. Compute an index of sphericity (this is described in Section 18.12.2).

Sphericity, which is an important issue for repeated measurement designs, has been explored in detail by several authors (cf. Box, 1954; Greenhouse and Geisser, 1959; Huynh and Feldt, 1970; Rouanet and Lépine, 1970; Huynh and Mandeville, 1979; Keselman, Rogan, Mendoza, and Breen, 1980; Kirk, 1982). A discussion of this problem along with some other possible statistical techniques not relying on sphericity such as the T^2 can also be found in: Winer (1971), McCall and Appelbaum (1973), Myers (1979), Rogan *et al.* (1979), Kirk (1982), Appelbaum and McCall (1983), Hertzog and Rovine (1985), Crowder and Hand (1990), Kirk (1995), and Algina and Keselman (1997).

18.12.2 Box's index of sphericity: ϵ

We will illustrate Box's index of sphericity with a fictitious example given in Table 18.6 where we collected the data from $S = 5$ subjects whose responses were measured for $A = 4$ different treatments. The standard analysis of variance will give an $F_A = \frac{600}{112} = 5.36$. With $\nu_1 = 3$ and $\nu_2 = 12$, this gives a p value of .018.

In order to evaluate the degree of non-sphericity, the first step is to collect into a table called a *covariance matrix* the variance and covariance between all treatments. This matrix is given in Table 18.7.

Box (1954) defined an index of non-sphericity which quantifies the degree of non-sphericity of a *population covariance* table. This table gives the values of the population covariances between A experimental conditions, and the values in the diagonal are the population variances for these A conditions. If we call $\zeta_{a,a'}$ (read 'zeta a a prime') the entries of this $A \times A$ table,

	a_1	a_2	a_3	a_4	M_s
S_1	76	64	34	26	50
S_2	60	48	46	30	46
S_3	58	34	32	28	38
S_4	46	46	32	28	38
S_5	30	18	36	28	28
$M_{a.}$	54	42	36	28	$M_{..} = 40$

Table 18.6 A data set to illustrate the sphericity problem for a repeated measurement design.

	a_1	a_2	a_3	a_4	
a_1	294	258	8	−8	
a_2	258	294	8	−8	
a_3	8	8	34	6	
a_4	−8	−8	6	2	
$\bar{t}_{a.}$	138	138	14	−2	$\bar{t}_{..} = 72$
$\bar{s}_{a.} - \bar{s}_{..}$	66	66	−58	−74	

Table 18.7 The covariance matrix for the data set of Table 18.6.

Box defined ε (read ‘epsilon’), the index of non-sphericity, as

$$\varepsilon = \frac{\left(\sum_a \zeta_{a,a} \right)^2}{(A-1) \sum_{a,a'} \zeta_{a,a'}^2}, \quad (18.2)$$

(where a and a' correspond to all possible pairs of indices going from 1 to A). The value of ε varies between 0 and 1; the smaller the value of ε , the less spherical the data.

Box also showed that when sphericity fails, the F_A ratio still follows a Fisher distribution but with different degrees of freedom, namely: $\nu_1 = \varepsilon(A-1)$ and $\nu_2 = \varepsilon(A-1)(S-1)$.

18.12.3 Greenhouse–Geisser correction

Box’s approach works when we use the *population* variance covariance matrix, but unfortunately this matrix is not known in general. In order to estimate ε we need to transform the sample covariance matrix into an *estimate* of the population covariance matrix. If we denote by $t_{a,a'}$ the sample estimate of the covariance between groups a and a' (these values are given in Table 18.7), by $\bar{t}_{a\cdot}$ the mean of the covariances for group a and by $\bar{t}_{\cdot\cdot}$ the grand mean of the covariance table, we can compute a new matrix which is an estimate of the population covariance matrix and whose general term will be denoted by $s_{a,a'}$ and computed as

$$s_{a,a'} = (t_{a,a'} - \bar{t}_{\cdot\cdot}) - (\bar{t}_{a\cdot} - \bar{t}_{\cdot\cdot}) - (\bar{t}_{a'\cdot} - \bar{t}_{\cdot\cdot}) = t_{a,a'} - \bar{t}_{a\cdot} - \bar{t}_{a'\cdot} + \bar{t}_{\cdot\cdot}. \quad (18.3)$$

(this procedure is called ‘double-centering’). For example the estimated value of the covariance between Conditions 2 and 3 is:

$$s_{1,3} = t_{1,3} - \bar{t}_{1\cdot} - \bar{t}_{3\cdot} + \bar{t}_{\cdot\cdot} = 8 - 138 - 14 + 72 = -72. \quad (18.4)$$

Table 18.8 gives the double-centered covariance matrix. From this matrix, we can compute the estimate of ε which is denoted $\hat{\varepsilon}$ (compare with Equation 18.2):

$$\hat{\varepsilon} = \frac{\left(\sum_a s_{a,a} \right)^2}{(A-1) \sum_{a,a'} s_{a,a'}^2}. \quad (18.5)$$

	a_1	a_2	a_3	a_4
a_1	90	54	−72	−72
a_2	54	90	−72	−72
a_3	−72	−72	78	66
a_4	−72	−72	66	78

Table 18.8 The double-centered covariance matrix used to estimate the population covariance matrix.

In our example, this formula gives:

$$\begin{aligned}\widehat{\varepsilon} &= \frac{(90 + 90 + 78 + 78)^2}{(4 - 1)(90^2 + 54^2 + \dots + 66^2 + 78^2)} = \frac{336^2}{3 \times 84,384} \\ &= \frac{112,896}{253,152} = .4460 .\end{aligned}\quad (18.6)$$

We use the value of $\widehat{\varepsilon} = .4460$ to correct the number of degrees of freedom of F_A as $\nu_1 = \widehat{\varepsilon}(A - 1) = 1.34$ and $\nu_2 = \widehat{\varepsilon}(A - 1)(S - 1) = 5.35$. These corrected values of ν_1 and ν_2 give for $F_A = 5.36$ a probability of $p = .059$. If you want to use the critical value approach, you need to round the values of these corrected degrees of freedom to the nearest integer (which will give here the values of $\nu_1 = 1$ and $\nu_2 = 5$).

18.12.4 Extreme Greenhouse–Geisser correction

A *conservative* (i.e. increasing the risk of Type II error: the probability of not rejecting the null hypothesis when it is false) correction for sphericity has been suggested by Greenhouse and Geisser (1959). Their idea is to choose the largest possible value of $\widehat{\varepsilon}$, which is equal to $A - 1$. This leads to consider that F_A follows a Fisher distribution with $\nu_1 = 1$ and $\nu_2 = S - 1$ degrees of freedom. In this case, these corrected values of $\nu_1 = 1$ and $\nu_2 = 4$ give for $F_A = 5.36$ a probability of $p = .081$.

18.12.5 Huynh–Feldt correction

Huynh and Feldt (1976) suggested a better (more powerful) approximation for ε . It is denoted $\widetilde{\varepsilon}$ and it is computed as

$$\widetilde{\varepsilon} = \frac{S(A - 1)\widehat{\varepsilon} - 2}{(A - 1)[S - 1 - (A - 1)\widehat{\varepsilon}]} .\quad (18.7)$$

In our example, this formula gives:

$$\widetilde{\varepsilon} = \frac{5(4 - 1).4460 - 2}{(4 - 1)[5 - 1 - (4 - 1).4460]} = .5872 .$$

We use the value of $\widetilde{\varepsilon} = .5872$ to correct the number of degrees of freedom of F_A as $\nu_1 = \widetilde{\varepsilon}(A - 1) = 1.76$ and $\nu_2 = \widetilde{\varepsilon}(A - 1)(S - 1) = 7.04$. These corrected values give for $F_A = 5.36$ a probability of $p = .041$. If you want to use the critical value approach, you need to round these corrected values for the number of degrees of freedom to the nearest integer (which will give here the values of $\nu_1 = 2$ and $\nu_2 = 7$). In general, the correction of Huynh and Feldt is to be preferred because it is more powerful (and Greenhouse–Geisser is too conservative).

18.12.6 Stepwise strategy for sphericity

Greenhouse and Geisser (1959) suggest using a stepwise strategy for the implementation of the correction for lack of sphericity. If F_A is not significant with the standard degrees of freedom, there is no need to implement a correction (because it will make it even less significant). If F_A is significant with the extreme correction [i.e. with $\nu_1 = 1$ and $\nu_2 = (S - 1)$], then there is

no need to correct either (because the correction will make it more significant). If F_A is not significant with the extreme correction but is significant with the standard number of degrees of freedom, then use the ε correction (Greenhouse and Geisser recommend using $\widehat{\varepsilon}$, but the subsequent $\widetilde{\varepsilon}$ is a better estimate and should be preferred).

18.13 An example with computational formulas

In this section we present an additional example along with a routine in nine steps and four ‘numbers in a square’

A

S

AS

1

to compute the analysis of variance.

In a psychopharmacological experiment, we want to test the effect of two types of amphetamine-like drugs on latency performing a motor task. In order to control for any potential sources of variation due to individual reactions to amphetamines, the same six subjects were used in the three conditions of the experiment: Drug A, Drug B, and Placebo. The dependent variable is the reaction time measured in milliseconds. The results of the experiment are presented in Table 18.9.

Before looking at the computational routines, try to use the comprehension formula given in the text (the ideal would be that you would discover the following nine steps by yourself).

$$Q1 = Y_{..} = 124 + 105 + \cdots + 84 = 1,800.00$$
$$Q2 = \text{AS} = \sum Y_{as}^2 = 124^2 + 105^2 + \cdots + 84^2 = 182,850.00$$
$$Q3 = \text{A} = \sum \frac{Y_{a.}^2}{S} = \frac{660^2 + 570^2 + 570^2}{6} = 180,900.00$$
$$Q4 = \text{S} = \sum \frac{Y_{.s}^2}{A} = \frac{336^2 + \cdots + 276^2}{3} = 180,750.00$$

Experimental conditions				
Subject	Drug A	Placebo	Drug B	Total
s ₁	124.00	108.00	104.00	336.00
s ₂	105.00	107.00	100.00	312.00
s ₃	107.00	90.00	100.00	297.00
s ₄	109.00	89.00	93.00	291.00
s ₅	94.00	105.00	89.00	288.00
s ₆	121.00	71.00	84.00	276.00
Total	660.00	570.00	570.00	1,800.00

Table 18.9 Results of a fictitious hypothetical experiment illustrating the computational routine for an $S \times A$ design.

Source	df	SS	MS	F	p(F)
$\mathcal{A}$	2	900.00	450.00	3.75	.060
$\mathcal{S}$	5	750.00	150.00		
$\mathcal{AS}$	10	1,200.00	120.00		
Total	17	2,850.00			

Table 18.10 ANOVA table for the data of Table 18.9.

Source	df	Expanded df	Comprehension formula	Computational formula
$\mathcal{A}$	$A - 1$	$A - 1$	$S \sum (M_{a\cdot} - M_{\cdot\cdot})^2$	$\boxed{\mathbf{A}} - \boxed{\mathbf{1}}$
$\mathcal{S}$	$S - 1$	$S - 1$	$A \sum (M_{\cdot s} - M_{\cdot\cdot})^2$	$\boxed{\mathbf{S}} - \boxed{\mathbf{1}}$
$\mathcal{AS}$	$(A - 1)(S - 1)$	$AS - A$ $-S + 1$	$\sum (Y_{as} - M_{\cdot s} - M_{a\cdot} + M_{\cdot\cdot})^2$	$\boxed{\mathbf{AS}} - \boxed{\mathbf{S}}$ $- \boxed{\mathbf{A}} + \boxed{\mathbf{1}}$

Table 18.11 Relationships between the name of the sources of variability, the degree of freedom and the sums of squares for an $\mathcal{S} \times \mathcal{A}$ design.

$$Q5 = \boxed{\mathbf{1}} = \frac{Y_{\cdot\cdot}^2}{AS} = \frac{1,800^2}{18} = 180,000.00$$

$$Q6 = SS_{\text{total}} = \boxed{\mathbf{AS}} - \boxed{\mathbf{1}} = 2,850.00$$

$$Q7 = SS_{\mathcal{A}} = \boxed{\mathbf{A}} - \boxed{\mathbf{1}} = 900.00$$

$$Q8 = SS_{\mathcal{S}} = \boxed{\mathbf{S}} - \boxed{\mathbf{1}} = 750.00$$

$$Q9 = SS_{\mathcal{AS}} = \boxed{\mathbf{AS}} - \boxed{\mathbf{S}} - \boxed{\mathbf{A}} + \boxed{\mathbf{1}} = 1,200.00$$

(18.8)

These quantities are summarized in Table 18.10. The relationship between the names of the sources of variation, the degrees of freedom, and the comprehension and computational formulas for the sums of squares that were suggested in Chapter 8 can be generalized to this specific case as illustrated in Table 18.11.

18.14 Another example: proactive interference

In an experiment on proactive interference, subjects were asked to learn a list of ten pairs of words. Two days later they were asked to recall these words. Once they finished recalling this first list, they were asked to learn a second list of ten pairs of words which they would be asked to recall after a new delay of two days. Recall of the second list was followed by a third list and so on until they learned and recalled six lists. The independent variable is the rank of

Subject	Rank of the list						Total
	1	2	3	4	5	6	
s_1	30	18	21	15	18	12	114
s_2	21	23	16	17	13	12	102
s_3	19	21	13	15	13	9	90
s_4	19	19	16	9	11	10	84
s_5	21	16	12	15	9	11	84
s_6	22	17	14	12	10	9	84
s_7	19	20	17	10	13	5	84
s_8	17	18	11	11	9	12	78
Total	168	152	120	104	96	80	720

Table 18.12 Results of an experiment about proactive interference effects in memory

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p</i> (<i>F</i>)
$\mathcal{A}$	5	720.00	144.00	24.00**	> .00001
$\mathcal{S}$	7	168.00	24.00		
$\mathcal{AS}$	35	210.00	6.00		
Total	47	1,098.00			

Table 18.13 ANOVA table for the experiment on proactive interference effects

the list in the learning sequence (first list, second list, ... , sixth list). The dependent variable is the number of pairs correctly recalled. The authors of this experiment predict that recall performance will decrease as a function of the rank of the lists (this effect is called ‘proactive interference’). The results are presented in Table 18.12.

As usual, do the computations on your own and then check your results. The correct answer is given in Table 18.13. To interpret the results of this experiment draw a graph like the one presented in the previous chapters. This should help you to reach an additional (or more subtle) conclusion—which one? Re-do the two previous exercises bearing in mind that the results were obtained using a between-subjects experimental design. What do you notice?

Chapter summary

18.15 Key notions of the chapter

Opposite are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they

are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

Greater sensitivity of repeated measurement designs $\mathcal{S} \times \mathcal{A}$ compared to the completely randomized designs $\mathcal{S}(\mathcal{A})$.

In the $\mathcal{S} \times \mathcal{A}$, the interaction $\mathcal{AS}$ is used to estimate the error. The actual error and the interaction are *confounded*.

Sum of squares within subjects and sum squares between subjects.

What is a *carry-over* effect, and how to detect it. If you find or suspect one, what do you need to do?

Problems linked to pre-test and post-test: history, maturation of subjects, and sensitization by the test.

Regression toward the mean.

Experimental mortality.

Sphericity (a.k.a. circularity).

Box sphericity index.

Greenhouse–Geisser correction (extreme or not).

18.16 New notations

Below are the new notations introduced in this chapter. Test yourself on their meaning.

$$SS_{AS}, \quad MS_{AS}, \quad df_{AS}$$

$$SS_S, \quad MS_S, \quad df_S$$

$$M_{\cdot s} = \frac{Y_{\cdot s}}{A}$$

$$R^2_{Y \cdot A}, \quad R^2_{Y \cdot AS}, \quad R^2_{Y \cdot S}$$

$$\sum \frac{Y_{\cdot s}^2}{A} = \boxed{\mathbf{S}}$$

$$\sum Y_{as}^2 = \boxed{\mathbf{AS}}$$

$$\varepsilon, \hat{\varepsilon}, \tilde{\varepsilon}, \zeta, s_{a,a}, t_{a,a},$$

18.17 Key formulas of the chapter

Below are the main formulas introduced in this chapter: try to go through them and understand what they mean.

$$SS_{S(A)} = SS_S + SS_{AS}$$

$$F_A = \frac{MS_A}{MS_{AS}}$$

$$SS_{\text{total}} = SS_S + SS_A + SS_{AS}$$

$$df_{\text{total}} = df_S + df_A + df_{AS}$$

$$df_{AS} = (A - 1)(S - 1)$$

$$MS_{AS} = \frac{SS_{AS}}{df_{AS}} = \frac{SS_{AS}}{(A - 1)(S - 1)}$$

$$SS_A = S \sum (M_{a\cdot} - M_{..})^2 = S \sum \mathcal{A}_a$$

$$SS_S = A \sum (M_{\cdot s} - M_{..})^2 = A \sum \mathcal{S}_s$$

$$SS_{AS} = \sum (Y_{as} - M_{a\cdot} - M_{\cdot s} + M_{..})^2 = \sum \mathcal{AS}_{as}$$

$$Y_{as} = M_{..} + (M_{a\cdot} - M_{..}) + (M_{\cdot s} - M_{..}) + (Y_{as} - M_{a\cdot} - M_{\cdot s} + M_{..})$$

$$\mathcal{A}_a = (M_{a\cdot} - M_{..})$$

$$\mathcal{S}_s = (M_{\cdot s} - M_{..})$$

$$\mathcal{AS}_{as} = (Y_{as} - M_{a\cdot} - M_{\cdot s} + M_{..})$$

$$Y_{as} = M_{..} + \mathcal{A}_a + \mathcal{S}_s + \mathcal{AS}_{as} .$$

$$R_{Y \cdot A}^2 = \frac{SS_A}{SS_{\text{total}}}$$

$$R_{Y \cdot S}^2 = \frac{SS_S}{SS_{\text{total}}}$$

$$R_{Y \cdot AS}^2 = \frac{SS_{AS}}{SS_{\text{total}}}$$

$$F_A = \frac{R_{Y \cdot A}^2}{R_{Y \cdot AS}^2} \times \frac{df_{AS}}{df_A} .$$

$$\varepsilon = \frac{\left(\sum_a \zeta_{a,a} \right)^2}{(A-1) \sum_{a,a'} \zeta_{a,a'}^2} ,$$

$$\nu_1 = \varepsilon(A-1) \text{ and } \nu_2 = \varepsilon(A-1)(S-1).$$

$$s_{a,a'} = (t_{a,a'} - \bar{t}_{..}) - (\bar{t}_{a\cdot} - \bar{t}_{..}) - (\bar{t}_{\cdot a'} - \bar{t}_{..}) = t_{a,a'} - \bar{t}_{a\cdot} - \bar{t}_{\cdot a'} + \bar{t}_{..} .$$

$$\widehat{\varepsilon} = \frac{\left(\sum_a s_{a,a} \right)^2}{(A-1) \sum_{a,a'} s_{a,a'}^2} .$$

$$\widetilde{\varepsilon} = \frac{S(A-1)\widehat{\varepsilon} - 2}{(A-1)[S-1-(A-1)\widehat{\varepsilon}]} .$$

18.18 Key questions of the chapter

Below are some questions about the content of this chapter. All the answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- * What are the reasons to prefer an $\mathcal{S} \times \mathcal{A}$ design over a $\mathcal{S}(\mathcal{A})$ design?
- * When is it better to avoid using an $\mathcal{S} \times \mathcal{A}$?
- * What is sphericity? Why is it a problem?
- * When should we correct for lack of sphericity?

19

ANOVA, two-factor completely repeated measures: $\mathcal{S} \times \mathcal{A} \times \mathcal{B}$

19.1 Introduction

We have already seen repeated measures designs. In the present chapter we will simply generalize what we have seen previously with one-factor, completely repeated measure designs, to two-factor, completely repeated measure designs. For convenience, we will call the experimental factors $\mathcal{A}$ and $\mathcal{B}$.

19.2 Example: plungin'!

Suppose we try to replicate Godden and Baddeley's (1975) experiment. The goal of this experiment was to show the effects of context on memory. Godden and Baddeley's hypothesis was that memory should be better when the conditions at test are more similar to the conditions experienced during learning. To operationalize this idea, Godden and Baddeley decided to use a very particular population: deepsea divers. The divers were asked to learn a list of 50 unrelated words either on the beach or under about 10 feet of water. The divers were then tested either on the beach or under sea. The divers needed to be tested in both environments in order to make sure that any effect observed could not be attributed to a global effect of one of the environments. The rationale behind using divers was two-fold. The first reason was practical: is it worth designing training programs on dry land for divers if they are not able to recall under water what they have learned? There is strong evidence, incidentally, that the problem is real. The second reason was more akin to good principles of experimental design. The difference between contexts (under sea and on the beach) seems quite important, hence a context effect should be easier to demonstrate in this experiment.

Because it is not very easy to find deepsea divers (willing in addition to participate in a memory experiment) it was decided to use a small number of divers in all possible conditions of the design. The list of words was randomly created and assigned to each subject. The order of testing was randomized in order to eliminate any possible carry-over effects by confounding them with the experimental error.

The first independent variable is the place of learning. It has 2 levels (on the beach and under sea), and it is denoted $\mathcal{A}$. The second independent variable is the place of testing. It has 2 levels (on the beach and under sea, like $\mathcal{A}$), and it is denoted $\mathcal{B}$. Crossing these 2 independent variables gives 4 experimental conditions:

1. Learning on the beach and recalling on the beach.
2. Learning on the beach and recalling under sea.
3. Learning under sea and recalling on the beach.
4. Learning under sea and recalling under sea.

Because each subject in this experiment participates in all four experimental conditions, the factor S is crossed with the 2 experimental factors. Hence, the design can be symbolized as an $S \times \mathcal{A} \times \mathcal{B}$ design. For this (fictitious) replication of Godden and Baddeley's (1980) experiment we have been able to convince $S = 5$ (fictitious) subjects to take part in this experiment (the original experiment had 16 subjects). Before enjoying the delicate pleasure of contemplating the results, let us enjoy a small theoretical *détour*.

19.3 Sum of squares, mean squares and F ratios

The deviation of each score from its mean is decomposed into components corresponding to the different sources of variability. When squared and summed, each of these components gives a sum of squares. These different sums of squares are orthogonal to each other as can be proven using the technique detailed for the $S(\mathcal{A})$ design (cf. Chapter 7, Section 7.3.5, 138). Their comprehension formula follows:

$$\begin{aligned}
 SS_A &= BS \sum (M_{a..} - M_{...})^2 \\
 SS_B &= AS \sum (M_{.b.} - M_{...})^2 \\
 SS_S &= AB \sum (M_{..s} - M_{...})^2 \\
 SS_{AB} &= S \sum (M_{ab.} - M_{a..} - M_{.b.} + M_{...})^2 \\
 SS_{AS} &= B \sum (M_{a.s} - M_{a..} - M_{..s} + M_{...})^2 \\
 SS_{BS} &= A \sum (M_{.bs} - M_{.b.} - M_{..s} + M_{...})^2 \\
 SS_{ABS} &= \sum (Y_{abs} - M_{ab.} - M_{a.s} - M_{.bs} + M_{a..} + M_{.b.} + M_{..s} - M_{...})^2
 \end{aligned}$$

Each sum of squares has its associated degrees of freedom:

$$\begin{aligned}
 df_A &= A - 1 \\
 df_B &= B - 1 \\
 df_S &= S - 1 \\
 df_{AB} &= (A - 1)(B - 1) \\
 df_{AS} &= (A - 1)(S - 1)
 \end{aligned}$$

$$df_{BS} = (B - 1)(S - 1)$$

$$df_{ABS} = (A - 1)(B - 1)(S - 1)$$

$$df_{\text{total}} = ABS - 1.$$

As usual the F ratios are obtained by dividing the mean squares of the sources by the error mean square. Specifically,

- For the main effect of $\mathcal{A}$, the F is

$$F_A = \frac{MS_A}{MS_{AS}}.$$

Under the null hypothesis this F ratio will follow a Fisher distribution with $\nu_1 = (A - 1)$ and $\nu_2 = (A - 1)(S - 1)$ degrees of freedom.

- For the main effect of $\mathcal{B}$ the F is

$$F_B = \frac{MS_B}{MS_{BS}}.$$

Under the null hypothesis this F ratio will follow a Fisher distribution with $\nu_1 = (B - 1)$ and $\nu_2 = (B - 1)(S - 1)$ degrees of freedom.

- For the interaction effect between $\mathcal{A}$ and $\mathcal{B}$ the F is

$$F_{AB} = \frac{MS_{AB}}{MS_{ABS}}.$$

Under the null hypothesis this F ratio will follow a Fisher distribution with $\nu_1 = (A - 1)(B - 1)$ and $\nu_2 = (A - 1)(B - 1)(S - 1)$ degrees of freedom.

The similarity between these tests and those used for an $\mathcal{S} \times \mathcal{A}$ design is worth noting. In an $\mathcal{S} \times \mathcal{A}$ design, the interaction mean square (i.e. MS_{AS}) is used as the denominator of the F ratio for the source $\mathcal{A}$. In an $\mathcal{S} \times \mathcal{A} \times \mathcal{B}$ design the effect of an experimental source is tested against the interaction of that source and $\mathcal{S}$. Hence the $\mathcal{S} \times \mathcal{A} \times \mathcal{B}$ designs generalize relatively straightforwardly the $\mathcal{S} \times \mathcal{A}$ designs.

The procedure for the different tests can be condensed in a single table showing the mean square used to test each source of variation. The general label for the mean squares used to test a source is ‘mean square test’ and it is abbreviated as MS_{test} .

Source	MS_{test}
$\mathcal{A}$	MS_{AS}
$\mathcal{B}$	MS_{BS}
$\mathcal{S}$	-----
$\mathcal{AB}$	MS_{ABS}
$\mathcal{AS}$	-----
$\mathcal{BS}$	-----
$\mathcal{ABS}$	-----

The dashes (-----) are placed in the table to indicate that some sources of variation cannot be tested because there is no mean square that will match them under the null hypothesis.

It is worth noting, however, that the experimental sources (i.e. $\mathcal{A}$, $\mathcal{B}$, $\mathcal{AB}$) can be tested. Hence, when $\mathcal{A}$ and $\mathcal{B}$ are fixed factors, the same statistical hypotheses can be tested with both an $\mathcal{S} \times \mathcal{A} \times \mathcal{B}$ design and an $\mathcal{S}(\mathcal{A} \times \mathcal{B})$ design.

As shown for the designs previously presented, the formulas for calculating the F indices can be justified by examining the score model. The score model also makes clear why certain sources of variation cannot be tested.

19.4 Score model (Model I), $\mathcal{S} \times \mathcal{A} \times \mathcal{B}$ design: $\mathcal{A}$ and $\mathcal{B}$ fixed

By using notations similar to those in the preceding chapters, we can decompose the score of subject s in experimental group ab as follows:

$$Y_{abs} = \mu \dots + \alpha_a + \beta_b + s_s + \alpha\beta_{ab} + \alpha s_{as} + \beta s_{bs} + \alpha\beta s_{abs} + e_{abs}.$$

Note that we have introduced two new interaction terms: one term is the interaction between the subject factor and $\mathcal{B}$ (i.e. the $\mathcal{BS}$ interaction), and the other term is the interaction between the subject factor and both independent variables (i.e. the $\mathcal{ABS}$ interaction). The latter interaction, which involves *three* variables, is termed a *second-order interaction* because there are *two* ‘crosses’ in $\mathcal{A} \times \mathcal{B} \times \mathcal{S}$. Note also that $\mathcal{S}$ is a random factor (it is written as a roman letter in the formula), and that (as in the $\mathcal{S} \times \mathcal{A}$ design) the error term is confounded with the highest level of interaction. We can see that the two terms, $\alpha\beta s$ and e , have the same indices. They vary in a concomitant manner and so it is not possible to separate their effects.

As usual, each part of the score model (except e) corresponds to a source of effect and thus is represented by a source of variance. So, we can decompose the difference of a score from the grand mean into ‘slices’ with one slice for each effect. Squaring these slices then gives the different sums of squares, which were listed in the previous section. As an exercise, try to express the different sums of squares in terms of the score model (as done for the $\mathcal{A} \times \mathcal{B}$ design). Don’t forget that $\mathcal{A}$ and $\mathcal{B}$ are fixed factors, but $\mathcal{S}$ is always random. Then, using these expressions, you should be able to construct the following formulas for the mathematical expectation of the different mean squares:

$$\mathbf{E}\{MS_A\} = \sigma_e^2 + B\sigma_{as}^2 + BS\vartheta_a^2$$

$$\mathbf{E}\{MS_B\} = \sigma_e^2 + A\sigma_{bs}^2 + AS\vartheta_b^2$$

$$\mathbf{E}\{MS_S\} = \sigma_e^2 + AB\sigma_s^2$$

$$\mathbf{E}\{MS_{AB}\} = \sigma_e^2 + \sigma_{abs}^2 + S\vartheta_{ab}^2$$

$$\mathbf{E}\{MS_{AS}\} = \sigma_e^2 + B\sigma_{as}^2$$

$$\mathbf{E}\{MS_{BS}\} = \sigma_e^2 + A\sigma_{bs}^2$$

$$\mathbf{E}\{MS_{ABS}\} = \sigma_e^2 + \sigma_{abs}^2.$$

These expected values for the mean squares give the rationale behind the formulas given in the previous section. Note that there is no appropriate MS_{test} for the sources of variability: $\mathcal{S}$, $\mathcal{AS}$, $\mathcal{BS}$ or $\mathcal{ABS}$. Only the sources of variation $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{AB}$ can be tested.

After that beautiful theoretical *détour*, we are now ready to plunge into the results of the replication of the experiment of Godden and Baddeley (1980).

19.5 Results of the experiment: plungin’

In this experiment, we asked subjects to learn lists made of 50 short words each. Each list has been made by drawing randomly words from a dictionary. Each list is used just once (hence, because we have $S = 5$ subjects and $A \times B = 2 \times 2 = 4$ experimental conditions, we have $5 \times 4 = 20$ lists). The dependent variable is the number of words recalled 10 minutes after learning (in order to have enough time to plunge or to come back to the beach).

The results of the (fictitious) replication are given in Table 19.1. Please take a careful look at it and make sure you understand the way the results are laid out. The usual implicit summation conventions are followed to denote the sums and the means. Recall that the first independent variable, $\mathcal{A}$, is the place of learning. It has 2 levels: on the beach (a_1) and under sea (a_2). The second independent variable, $\mathcal{B}$, is the place of testing. It has 2 levels: on the beach (b_1) and under sea (b_2).

Before starting to compute, let us recall that the prediction of the authors was that memory should be better when the contexts of encoding and testing are the same than when the contexts of encoding and testing are different. This means that the authors had a very specific shape of effects (i.e. an X-shaped interaction) in mind. As a consequence,

		$\mathcal{A}$ Learning place		Σ	Means
		a_1 On land	a_2 Under sea		
b_1 Testing place on land	s_1	34	14	48	24
	s_2	37	21	58	29
	s_3	27	31	58	29
	s_4	43	27	70	35
	s_5	44	32	76	38
		$Y_{11\cdot} = 185$ $M_{11\cdot} = 37$	$Y_{21\cdot} = 125$ $M_{21\cdot} = 25$	$Y_{\cdot 1\cdot} = 310$ $M_{\cdot 1\cdot} = 31$	
				$Y_{\cdot 2s}$	$M_{\cdot 2s}$
b_2 Testing place under sea	s_1	18	22	40	20
	s_2	21	25	46	23
	s_3	25	33	58	29
	s_4	37	33	70	35
	s_5	34	42	76	38
		$Y_{12\cdot} = 135$ $M_{12\cdot} = 27$	$Y_{22\cdot} = 155$ $M_{22\cdot} = 31$	$Y_{\cdot 2\cdot} = 290$ $M_{\cdot 2\cdot} = 29$	

Table 19.1 Result of a (fictitious) replication of Godden and Baddeley’s (1980) experiment with deepsea divers (see text for explanation).

$\mathcal{A}$						
Learning place (sums and means)						
	a_1 On land		a_2 Under sea		Σ	Means
	$Y_{1..s}$	$M_{1..s}$	$Y_{2..s}$	$M_{2..s}$	$Y_{..s}$	$M_{..s}$
s_1	52	26	36	18	88	22
s_2	58	29	46	23	104	26
s_3	52	26	64	32	116	29
s_4	80	40	60	30	140	35
s_5	78	39	74	37	152	38
	$Y_{1..} = 320$		$Y_{2..} = 280$		$Y_{...} = 600$	
	$M_{1..} = 32$		$M_{2..} = 28$		$M_{...} = 30$	

Table 19.2 Some sums that will help computing the sums of squares for Godden and Baddeley's (1980) experiment.

they predict that all of the experimental sums of squares should correspond to the sum of squares of interaction. Table 19.2 gives the different sums that will be needed to compute the sums of squares.

We show that you can obtain the sums of squares two ways; using either the comprehension formulas or the computational formulas. By now you should be quite confident in your use of the comprehension formulas. But as you probably have noted, the computation of the various sums of squares grows more difficult when the experimental design becomes more complex. It is therefore quite advisable to become familiar with the computational formulas because these are computationally less challenging. Since you now have both ways presented, make sure that you can get the same answer either way you go!

We can now start the detailed computations leading to the ANOVA table. First, let us arrive by the way of the comprehension formulas.

First, find the various degrees of freedom:

$$df_A = (A - 1) = 2 - 1 = 1 \quad (19.1)$$

$$df_B = (B - 1) = 2 - 1 = 1$$

$$df_S = (S - 1) = 5 - 1 = 4$$

$$df_{AB} = (A - 1)(B - 1) = (2 - 1)(2 - 1) = 1$$

$$df_{AS} = (A - 1)(S - 1) = (2 - 1)(5 - 1) = 4$$

$$df_{BS} = (B - 1)(S - 1) = (2 - 1)(5 - 1) = 4$$

$$df_{ABS} = (A - 1)(B - 1)(S - 1) = (2 - 1)(2 - 1)(5 - 1) = 4$$

$$df_{total} = A \times B \times S - 1 = 2 \times 2 \times 5 - 1 = 19$$

Now, compute SS_A , SS_B , SS_S , SS_{AB} , SS_{AS} , SS_{BS} , and SS_{ABS} :

$$\begin{aligned}
 SS_A &= BS \sum (M_{a..} - M_{...})^2 \\
 &= 2 \times 5 \times [(32 - 30)^2 + (28 - 30)^2]
 \end{aligned} \quad (19.2)$$

$$\begin{aligned}
&= 10 \times [2^2 + (-2)^2] \\
&= 10 \times [4 + 4] \\
&= 10 \times 8 \\
&= 80.00
\end{aligned}$$

$$\begin{aligned}
SS_B &= AS \sum (M_{.b.} - M_{..})^2 & (19.3) \\
&= 2 \times 5 \times [(31 - 30)^2 + (29 - 30)^2] \\
&= 10 \times [1^2 + (-1)^2] \\
&= 10 \times [1 + 1] \\
&= 10 \times 2 \\
&= 20.00
\end{aligned}$$

$$\begin{aligned}
SS_S &= AB \sum (M_{..s} - M_{...})^2 & (19.4) \\
&= 2 \times 2 [(22 - 30)^2 + \dots + (38 - 30)^2] \\
&= 4 \times [(-8)^2 + \dots + 8^2] \\
&= 4 \times [64 + \dots + 64] \\
&= 680.00
\end{aligned}$$

$$\begin{aligned}
SS_{AB} &= S \sum (M_{ab.} - M_{a..} - M_{.b.} + M_{...})^2 & (19.5) \\
&= 5 \times [(37 - 32 - 31 + 30)^2 + (27 - 32 - 29 + 30)^2 \\
&\quad + (25 - 31 - 28 + 30)^2 + (31 - 29 - 28 + 30)^2] \\
&= 5 \times [4^2 + (-4)^2 + (-4)^2 + 4^2] \\
&= 5 \times [16 + 16 + 16 + 16] \\
&= 5 \times 64.00 \\
&= 320.00
\end{aligned}$$

$$\begin{aligned}
SS_{AS} &= B \sum (M_{a.s} - M_{a..} - M_{..s} + M_{...})^2 & (19.6) \\
&= 2 \times [(26 - 32 - 22 + 30)^2 + (29 - 32 - 26 + 30)^2 + \dots \\
&\quad \dots + (30 - 28 - 35 + 30)^2 + (37 - 28 - 38 + 30)^2] \\
&= 2 \times [2^2 + 1^2 + \dots + (-3)^2 + 1^2] \\
&= 2 \times [4 + 1 + \dots + 9 + 1] \\
&= 160.00
\end{aligned}$$

$$\begin{aligned}
SS_{BS} &= A \sum (M_{.bs} - M_{.b.} - M_{..s} + M_{...})^2 & (19.7) \\
&= 2 \times [(24 - 31 - 22 + 30)^2 + (29 - 31 - 26 + 30)^2 + \dots
\end{aligned}$$

$$\begin{aligned}
& + \dots + (35 - 29 - 35 + 30)^2 + (38 - 29 - 38 + 30)^2] \\
& = 2 \times [1^2 + 2^2 + \dots + 1^2 + 1^2] \\
& = 2 \times [1 + 4 + \dots + 1 + 1] \\
& = 32.00
\end{aligned}$$

$$\begin{aligned}
SS_{ABS} &= (Y_{abs} - M_{ab.} - M_{a.s} - M_{.bs} + M_{a..} + M_{.b.} + M_{..s} - M_{...})^2 \quad (19.8) \\
&= [(34 - 37 - 26 - 24 + 32 + 31 + 22 - 30)^2 + \\
&\quad + (37 - 37 - 29 - 29 + 32 + 31 + 26 - 30)^2 + \dots \\
&\quad + \dots + (33 - 31 - 30 - 35 + 28 + 29 + 35 - 30)^2 + \\
&\quad + (42 - 31 - 37 - 38 + 28 + 29 + 38 - 30)^2] \\
&= [2^2 + 1^2 + \dots + (-1)^2 + 1^2] \\
&= [4 + 1 + \dots + 1 + 1] \\
&= 64.00
\end{aligned}$$

Here follow the routines for computing the sums of squares using the computational formulas.

$$Q1: Y_{...} = 34 + 37 + \dots + 33 + 42 = 600.00$$

$$Q2: \boxed{\mathbf{ABS}} = \sum Y_{abs}^2 = 34^2 + 37^2 + \dots + 33^2 + 42^2 = 19,356.00$$

$$Q3: \boxed{\mathbf{A}} = \sum \frac{Y_{a..}^2}{BS} = \frac{320^2 + 280^2}{10} = 18,080.00$$

$$Q4: \boxed{\mathbf{B}} = \sum \frac{Y_{.b.}^2}{AS} = \frac{310^2 + 290^2}{10} = 18,020.00$$

$$Q5: \boxed{\mathbf{S}} = \sum \frac{Y_{..s}^2}{AB} = \frac{88^2 + \dots + 152^2}{4} = 18,680.00$$

$$Q6: \boxed{\mathbf{AB}} = \sum \frac{Y_{ab.}^2}{S} = \frac{185^2 + 135^2 + 125^2 + 155^2}{5} = 18,420.00$$

$$Q7: \boxed{\mathbf{AS}} = \sum \frac{Y_{a.s}^2}{B} = \frac{52^2 + 58^2 \dots + 60^2 + 74^2}{2} = 18,920.00$$

$$Q8: \boxed{\mathbf{BS}} = \sum \frac{Y_{.bs}^2}{A} = \frac{48^2 + 58^2 \dots + 70^2 + 76^2}{2} = 18,732.00$$

$$Q9: \boxed{\mathbf{1}} = \frac{Y_{...}^2}{ABS} = \frac{600^2}{20} = 18,000.00$$

$$Q10: SS_A = \boxed{\mathbf{A}} - \boxed{\mathbf{1}} = 18,080.00 - 18,000.00 = 80.00$$

$$Q11: SS_B = \boxed{\mathbf{B}} - \boxed{\mathbf{1}} = 18,020.00 - 18,000.00 = 20.00$$

$$Q12: SS_S = \boxed{\mathbf{S}} - \boxed{\mathbf{1}} = 18,680.00 - 18,000.00 = 680.00$$

$$Q13: SS_{AB} = \boxed{AB} - \boxed{A} - \boxed{B} + \boxed{1}$$

$$= 18,420.00 - 18,080.00 - 18,020.00 + 18,000.00 = 320.00$$

$$Q14: SS_{AS} = \boxed{AS} - \boxed{A} - \boxed{S} + \boxed{1}$$

$$= 18,920.00 - 18,080.00 - 18,680.00 + 18,000.00 = 160.00$$

$$Q15: SS_{BS} = \boxed{BS} - \boxed{B} - \boxed{S} + \boxed{1}$$

$$= 18,732.00 - 18,020.00 - 18,680.00 + 18,000.00 = 32.00$$

$$Q16: SS_{ABS} = \boxed{ABS} + \boxed{A} + \boxed{B} + \boxed{S} - \boxed{AB} - \boxed{AS} - \boxed{BS} - \boxed{1} =$$

$$= 19,356.00 - 18,420.00 - 18,920.00 - 18,732.00$$

$$+ 18,080.00 + 18,020.00 + 18,680.00 - 18,000.00 = 64.00$$

$$Q17: SS_{\text{total}} = \boxed{ABS} - \boxed{1} = 19,356.00 - 18,000.00 = 1,356.00$$

Now that we have found the sums of squares, we can proceed to finding the mean squares:

$$MS_A = \frac{SS_A}{df_A} = \frac{80.00}{1} = 80.00 \quad (19.9)$$

$$MS_B = \frac{SS_B}{df_B} = \frac{20.00}{1} = 20.00$$

$$MS_S = \frac{SS_S}{df_S} = \frac{680.00}{4} = 170.00$$

$$MS_{AB} = \frac{SS_{AB}}{df_{AB}} = \frac{320.00}{1} = 320.00$$

$$MS_{AS} = \frac{SS_{AS}}{df_{AS}} = \frac{160.00}{4} = 40.00$$

$$MS_{BS} = \frac{SS_{BS}}{df_{BS}} = \frac{32.00}{4} = 8.00$$

$$MS_{ABS} = \frac{SS_{ABS}}{df_{ABS}} = \frac{64.00}{4} = 16.00.$$

And, finally, we compute the F values:

$$F_A = \frac{MS_A}{MS_{AS}} = \frac{80.00}{40.00} = 2.00 \quad (19.10)$$

$$F_B = \frac{MS_B}{MS_{BS}} = \frac{20.00}{8.00} = 2.50$$

$$F_{AB} = \frac{MS_{AB}}{MS_{ABS}} = \frac{320.00}{16.00} = 20.00$$

Armed with all the necessary numbers we can now fill in the ANOVA table.

Source	R^2	df	SS	MS	F	$p(F)$	ν_1	ν_2
$\mathcal{A}$	0.05900	1	80.00	80.00	2.00	.22973	1	4
$\mathcal{B}$	0.01475	1	20.00	20.00	2.50	.18815	1	4
S	0.50147	4	680.00	170.00	—			
$\mathcal{AB}$	0.23599	1	320.00	320.00	20.00	.01231	1	4
$\mathcal{AS}$	0.11799	4	160.00	40.00	—			
$\mathcal{BS}$	0.02360	4	32.00	8.00	—			
$\mathcal{ABS}$	0.04720	4	64.00	16.00	—			
Total	1.00	19	1,356.00					

Could you find the different critical values for each F ratio? If not, you may want to re-read this chapter or some previous chapters.

A picture is always better than a lot of words, and the results of this experiment are displayed in Figure 19.1. As you can see in this figure, an $S \times \mathcal{A} \times \mathcal{B}$ design is illustrated the same way as an $S(\mathcal{A} \times \mathcal{B})$, because only the experimental sources are displayed.

To conclude, we can say that there is a very clear interaction between the place of learning and the place of testing, $F_{AB}(1, 4) = 20.00$, $MS_e = 16.00$, $p < .05$ (almost $p < .01$ but not quite!). But there are no main effects of the place of learning, $F_A(1, 4) = 2.00$, $MS_e = 40.00$, $p > .10$, nor of the place of testing, $F_B(1, 4) = 2.50$, $MS_e = 8.00$, $p > .10$.

19.6 Score model (Model II): $S \times \mathcal{A} \times \mathcal{B}$ design, $\mathcal{A}$ and $\mathcal{B}$ random

As we have seen for the experimental design $S(\mathcal{A} \times \mathcal{B})$, we are going to see that things change when the two independent variables are random in an $S \times \mathcal{A} \times \mathcal{B}$ design. For $S(\mathcal{A} \times \mathcal{B})$

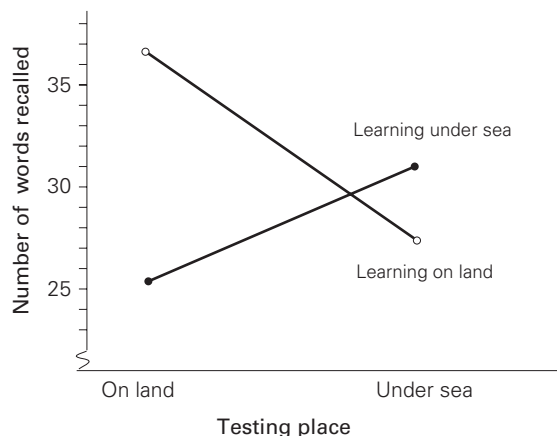


Figure 19.1 Results of a (fictitious) replication of Godden and Baddeley's (1980) experiment. Five deepsea divers learned lists of 50 words on the beach or under sea. They were tested on the beach or under sea. The dependent variable is the number of words recalled.

designs the main effects are tested by comparing them to the interaction $\mathcal{AB}$. Here we are not able to do this same test of the main effects. Let us see why.

If we derive the expected values of the mean squares, taking into account that $\mathcal{A}$ and $\mathcal{B}$ are random, we obtain

$$\begin{aligned}\mathbf{E}\{MS_A\} &= \sigma_e^2 + \sigma_{abs}^2 + B\sigma_{as}^2 + S\sigma_{ab}^2 + BS\sigma_a^2 \\ \mathbf{E}\{MS_B\} &= \sigma_e^2 + \sigma_{abs}^2 + A\sigma_{bs}^2 + S\sigma_{ab}^2 + AS\sigma_b^2 \\ \mathbf{E}\{MS_S\} &= \sigma_e^2 + \sigma_{abs}^2 + A\sigma_{bs}^2 + B\sigma_{as}^2 + AB\sigma_s^2 \\ \mathbf{E}\{MS_{AB}\} &= \sigma_e^2 + \sigma_{abs}^2 + S\sigma_{ab}^2 \\ \mathbf{E}\{MS_{AS}\} &= \sigma_e^2 + \sigma_{abs}^2 + B\sigma_{as}^2 \\ \mathbf{E}\{MS_{BS}\} &= \sigma_e^2 + \sigma_{abs}^2 + A\sigma_{bs}^2 \\ \mathbf{E}\{MS_{ABS}\} &= \sigma_e^2 + \sigma_{abs}^2.\end{aligned}$$

From these different expected values we can construct the following table:

Source	MS_{test}
$\mathcal{A}$	----
$\mathcal{B}$	----
$\mathcal{S}$	----
$\mathcal{AB}$	MS_{ABS}
$\mathcal{AS}$	MS_{ABS}
$\mathcal{BS}$	MS_{ABS}
$\mathcal{ABS}$	----

Note the importance of carefully choosing a statistical model, since in this case the main effects cannot be tested. However, there are statistical techniques which allow us to test these main effects by calculating a modified criterion called a *Quasi-F*. This modified criterion is then compared to a particularly determined Fisher distribution. Although we only touch lightly on this approach for now, we will return to it in more detail at the end of this chapter. For this Model II version of the $\mathcal{S} \times \mathcal{A} \times \mathcal{B}$ design (and also for the Model III version presented in the next section), the degrees of freedom and the mean squares are calculated in the same manner as when both $\mathcal{A}$ and $\mathcal{B}$ are fixed factors.

19.7 Score model (Model III): $\mathcal{S} \times \mathcal{A} \times \mathcal{B}$ design, $\mathcal{A}$ fixed, $\mathcal{B}$ random

When $\mathcal{A}$ is fixed and $\mathcal{B}$ is random, the following expected values are found for the mean squares:

$$\begin{aligned}\mathbf{E}\{MS_A\} &= \sigma_e^2 + \sigma_{abs}^2 + B\sigma_{as}^2 + S\sigma_{ab}^2 + BS\vartheta_a^2 \\ \mathbf{E}\{MS_B\} &= \sigma_e^2 + A\sigma_{bs}^2 + AS\sigma_b^2 \\ \mathbf{E}\{MS_S\} &= \sigma_e^2 + A\sigma_{bs}^2 + AB\sigma_s^2 \\ \mathbf{E}\{MS_{AB}\} &= \sigma_e^2 + \sigma_{abs}^2 + S\sigma_{ab}^2\end{aligned}$$

$$\mathbf{E}\{MS_{AS}\} = \sigma_e^2 + \sigma_{abs}^2 + B\sigma_{as}^2$$

$$\mathbf{E}\{MS_{BS}\} = \sigma_e^2 + A\sigma_{bs}^2$$

$$\mathbf{E}\{MS_{ABS}\} = \sigma_e^2 + \sigma_{abs}^2.$$

From these different expected values the following table can be constructed:

Source	MS_{test}
$\mathcal{A}$	-----
$\mathcal{B}$	MS_{BS}
$\mathcal{S}$	MS_{BS}
$\mathcal{AB}$	MS_{ABS}
$\mathcal{AS}$	MS_{ABS}
$\mathcal{BS}$	-----
$\mathcal{ABS}$	-----

It can be seen that, for $\mathcal{S} \times \mathcal{A} \times \mathcal{B}$ designs, the decision to consider a factor as fixed or random has greater consequences than for an $\mathcal{S}(\mathcal{A} \times \mathcal{B})$ design. This is a good time to review the reasons for, and the problems of, making this decision, as first described at the end of the initial chapter on the score model (Chapter 10, Section 10.6.3, page 206).

19.8 Quasi- $F : F'$

As we have just seen for some factorial designs, some sources of variation cannot be tested due to the lack of a mean square having—under the null hypothesis—the same expected value as the source to be tested. This is the case, for example, for both $\mathcal{A}$ and $\mathcal{B}$ in an $\mathcal{S} \times \mathcal{A} \times \mathcal{B}$ design when $\mathcal{A}$ and $\mathcal{B}$ are random factors.

In order to evaluate the source of an effect, despite the lack of a specific test mean square, a criterion called a *Quasi- F* and denoted F' (read as F prime) can be used. To understand the logic of this approach recall that the expected values of the mean square of $\mathcal{A}$ when $\mathcal{A}$ and $\mathcal{B}$ are random are:

$$\mathbf{E}\{MS_A\} = \sigma_e^2 + \sigma_{abs}^2 + B\sigma_{as}^2 + S\sigma_{ab}^2 + BS\sigma_a^2.$$

Under the null hypothesis, σ_a^2 is zero; so, when the null hypothesis is true, the expected value of the mean square of $\mathcal{A}$ is

$$\mathbf{E}\{MS_A\} = \sigma_e^2 + \sigma_{abs}^2 + B\sigma_{as}^2 + S\sigma_{ab}^2.$$

This value can be obtained by combining the expected values of several mean squares. Precisely, the combination needed is

$$MS_{AB} + MS_{AS} - MS_{ABS}.$$

We can calculate, therefore, a quantity denoted $MS'_{test,A}$ using the formula

$$MS'_{test,A} = MS_{AB} + MS_{AS} - MS_{ABS}$$

and test the null hypothesis ‘ $\mathcal{A}$ has no effect’ by calculating the criterion

$$F'_A = \frac{MS_A}{MS'_{test,A}}.$$

It can be shown (Satterthwaite, 1946), that the criterion F' follows, approximately, a Fisher distribution with $\nu_1 = A - 1$ degrees of freedom and ν_2 being given by the closest integer to the result of this (quite frightening!) formula:

$$\frac{(MS_{AB} + MS_{AS} - MS_{ABS})^2}{(MS_{AB}^2/df_{AB}) + (MS_{AS}^2/df_{AS}) + (MS_{ABS}^2/df_{ABS})}.$$

Caution: The mean squares in the numerator are added or subtracted, but the terms in the denominator are *always* added. Incidentally, if you use a computer program to obtain the probability associated with F' , you can plug in the actual value of ν_2 without rounding, this will make the result more precise.

As a numerical example, imagine that the results of the Godden and Baddeley experiment came from a design where $\mathcal{A}$ and $\mathcal{B}$ were random factors. We have:

- $MS_A = 80.00$, with $df_A = 1$
- $MS_{AS} = 40.00$, with $df_{AS} = 4$
- $MS_{AB} = 320.00$, with $df_{AB} = 1$
- $MS_{ABS} = 16.00$ with $df_{ABS} = 4$

and

$$MS'_{\text{test},A} = 320.00 + 40.00 - 16.00 = 344.00$$

and

$$F'_A = \frac{MS_A}{MS'_{\text{test},A}} = \frac{80}{344} = .232.$$

The F'_A criterion is approximately distributed as a Fisher distribution, with $\nu_1 = 1$ and ν_2 being the closest integer to

$$\begin{aligned} \nu_2 &\approx \frac{(MS_{AB} + MS_{AS} - MS_{ABS})^2}{(MS_{AB}^2/df_{AB}) + (MS_{AS}^2/df_{AS}) + (MS_{ABS}^2/df_{ABS})} \\ &= \frac{(320.00 + 40.00 - 16.00)^2}{[(320.00^2/1) + (40.00^2/4) + (16.00^2/4)]} \\ &= 1.15. \end{aligned}$$

Therefore, $\nu_2 = 1$ (1 is the closest integer to 1.15). The same procedure can be used for all factorial designs, whatever the complexity.

19.9 A cousin F''

By studying the formula for calculating the criterion F' we see that the denominator (and therefore F') can take on negative values. In this event, another type of *quasi-F* denoted F'' (read as '*F* second') may be used. F'' can be obtained by the following formula (verify by using the score model that this formula can test the null hypothesis for $\mathcal{A}$):

$$F''_A = \frac{MS_{ABS} + MS_A}{MS_{AB} + MS_{AS}}.$$

This criterion F'' approximately follows a Fisher distribution with ν_1 being the closest integer to

$$\frac{(MS_{ABS} + MS_A)^2}{(MS_{ABS}^2/df_{ABS}) + (MS_A^2/df_A)},$$

and ν_2 the closest integer to

$$\frac{(MS_{AB} + MS_{AS})^2}{(MS_{AB}^2/df_{AB}) + (MS_{AS}^2/df_{AS})}.$$

In general, given a linear combination of mean squares obtained by

$$MS_1 \pm MS_2 \pm \dots \pm MS_N,$$

the number of degrees of freedom is given by the closest integer to

$$\frac{(MS_1 \pm MS_2 \pm \dots \pm MS_N)^2}{(MS_1^2/df_1) + (MS_2^2/df_2) + \dots + (MS_N^2/df_N)}.$$

19.9.1 Digression: what to choose?

The answer to this question is not completely clear. However, the literature in this area (Davenport and Webster, 1973; Foster and Dickinson, 1976; Gaylor and Hopper, 1969; Hudson and Krutchkoff, 1968; Myers, 1979; Santa *et al.*, 1979), indicates that F' provides, in most cases, a satisfactory approximation for testing the null hypothesis. The criterion F'' is preferred if F' is negative, or if the number of subjects per group is small (4 or less). In most cases, however, the two criteria agree, and the choice is simply a matter of taste. To conclude, it is necessary to point out that both of the Quasi- F criteria are very conservative (i.e. sometimes they do not lead to the rejection of the null hypothesis when it is actually false). Nevertheless, their use is to be preferred over other procedures.

19.10 Validity assumptions, measures of intensity, key notions, etc.

The $\mathcal{S} \times \mathcal{A} \times \mathcal{B}$ designs are simply an extension of the $\mathcal{S} \times \mathcal{A}$ designs, and so all the same considerations and problems previously presented for a one-factor repeated measures design apply. The transposition is left to the reader as an exercise. The only new notion introduced in this chapter is the Quasi- F : F' and F'' .

Chapter summary

19.11 New notations

Over the page are the new notations introduced in this chapter. Test yourself on their meaning.

$$SS_{AS}$$

$$MS'_{\text{test},A}$$

$$SS_{BS}$$

$$F'$$

$$SS_{ABS}$$

$$F''$$

$$\text{Quasi-}F$$

19.12 Key formulas of the chapter

Below are the main formulas introduced in this chapter: try to go through them and understand what they mean.

$$SS_A = BS \sum (M_{a..} - M_{...})^2$$

$$SS_B = AS \sum (M_{.b.} - M_{...})^2$$

$$SS_S = AB \sum (M_{..s} - M_{...})^2$$

$$SS_{AB} = S \sum (M_{ab.} - M_{a..} - M_{.b.} + M_{...})^2$$

$$SS_{AS} = B \sum (M_{a.s} - M_{a..} - M_{..s} + M_{...})^2$$

$$SS_{BS} = A \sum (M_{.bs} - M_{.b.} - M_{..s} + M_{...})^2$$

$$SS_{ABS} = \sum (Y_{abs} - M_{ab.} - M_{a.s} - M_{.bs} + M_{a..} + M_{.b.} + M_{..s} - M_{...})^2$$

$$F_A = \frac{MS_A}{MS_{AS}}$$

$$F_B = \frac{MS_B}{MS_{BS}}$$

$$F_{AB} = \frac{MS_{AB}}{MS_{ABS}}$$

$$MS'_{\text{test},A} = MS_{AB} + MS_{AS} - MS_{ABS}$$

$$F'_A = \frac{MS_A}{MS'_{\text{test},A}}$$

$$v_2 = \frac{(MS_{AB} + MS_{AS} - MS_{ABS})^2}{(MS_{AB}^2/df_{AB}) + (MS_{AS}^2/df_{AS}) + (MS_{ABS}^2/df_{ABS})}$$

$$F''_A = \frac{MS_{ABS} + MS_A}{MS_{AB} + MS_{AS}}$$

$$v_1 = \frac{(MS_{ABS} + MS_A)^2}{(MS_{ABS}^2/df_{ABS}) + (MS_A^2/df_A)}$$

$$v_2 = \frac{(MS_{AB} + MS_{AS})^2}{(MS_{AB}^2/df_{AB}) + (MS_{AS}^2/df_{AS})}$$

20

ANOVA, two-factor partially repeated measures: $S(\mathcal{A}) \times \mathcal{B}$

20.1 Introduction

We have already seen that, in general, repeated measures designs are more sensitive (i.e. make it easier to reject the null hypothesis) than independent measures designs. In addition, repeated measures designs need a smaller number of subjects than the equivalent independent measures designs (and subjects are often rare and hard to find, first-year psychology students excepted). For these reasons, psychologists tend to prefer designs with repeated measures. In some cases, however, it is difficult or even impossible to use the same subjects in all the experimental conditions. For example, suppose you want to study the effect of *gender* and *alcohol* on risk taking when driving. Suppose that you have decided to operationalize the independent variable ‘alcohol level’ by having four concentration levels. Taking into account the important between-subject variability in reactivity to alcohol, you decide to have each subject observed in each alcohol condition. The order of administration of each condition will be randomized for each subject. The experimenters themselves will be ignorant of the alcohol condition the subjects are in (this is a double-blind study—see Chapter 1). It is, however, much more difficult to have repeated measures for the independent variable ‘sex of the subject’ (which would require measures for each subject in both the ‘male’ and ‘female’ condition).

The design that we described above is called a *partially* repeated measures design or mixed design. The measures are independent (i.e. non-repeated) for the factor *gender* that we will denote $\mathcal{A}$, and repeated for the factor *alcohol* level that we will denote $\mathcal{B}$. This design is symbolized by the notation $S(\mathcal{A}) \times \mathcal{B}$. The nesting factor $\mathcal{A}$ is the factor for which measures are independent. An equivalent way of saying this is that $\mathcal{A}$ is a *between-subject* independent variable or factor. Recall that a factor (say S) is nested in another factor (say $\mathcal{A}$) when each level of the first factor (S) appears in conjunction with one and only one level of the second factor ($\mathcal{A}$). By contrast, a given subject will be present in all conditions of factor $\mathcal{B}$ (the alcohol level). Hence the S factor is *crossed* with the $\mathcal{B}$ factor. The independent variable $\mathcal{B}$ is often called the *within-subject* factor, because its effect will be seen within each subject (because each subject

$S(\mathcal{A}) \times \mathcal{B}$ design Factor $\mathcal{B}$				
$\mathcal{A}$	b_1	b_2	b_3	b_4
a_1	s_1	s_1	s_1	s_1
	s_2	s_2	s_2	s_2
	s_3	s_3	s_3	s_3
	s_4	s_4	s_4	s_4
	s_5	s_5	s_5	s_5
a_2	s_6	s_6	s_6	s_6
	s_7	s_7	s_7	s_7
	s_8	s_8	s_8	s_8
	s_9	s_9	s_9	s_9
	s_{10}	s_{10}	s_{10}	s_{10}

Table 20.1 Layout of an $S(\mathcal{A}) \times \mathcal{B}$ design.

$S(\mathcal{A} \times \mathcal{B})$ design Factor $\mathcal{B}$				
$\mathcal{A}$	b_1	b_2	b_3	b_4
a_1	s_1	s_2	s_3	s_4
	s_5	s_6	s_7	s_8
	s_9	s_{10}	s_{11}	s_{12}
	s_{13}	s_{14}	s_{15}	s_{16}
	s_{17}	s_{18}	s_{19}	s_{20}
a_2	s_{21}	s_{22}	s_{23}	s_{24}
	s_{25}	s_{26}	s_{27}	s_{28}
	s_{29}	s_{30}	s_{31}	s_{32}
	s_{33}	s_{34}	s_{35}	s_{36}
	s_{37}	s_{38}	s_{39}	s_{40}

Table 20.2 Layout of an $S(\mathcal{A} \times \mathcal{B})$ design.

is measured in all the levels of $\mathcal{B}$). The experimental conditions are defined by crossing the levels of $\mathcal{A}$ and $\mathcal{B}$.

Suppose that you have at your disposal 10 volunteers (5 women and 5 men). The experimental layout is described in Table 20.1.

In order to make clearer the similarity and the differences between the 2-factor experimental designs, compare Table 20.1 with Tables 20.2 and 20.3 that describe respectively the experimental layout for an $S(\mathcal{A} \times \mathcal{B})$ and an $S \times \mathcal{A} \times \mathcal{B}$ design.

Hence in order to obtain the number of measures for the 8 ($\mathcal{A} \times \mathcal{B} = 2 \times 4 = 8$) experimental conditions, we need: 5 subjects with an $S \times \mathcal{A} \times \mathcal{B}$ design, 10 subjects with an $S(\mathcal{A}) \times \mathcal{B}$, and 40 subjects with an $S(\mathcal{A} \times \mathcal{B})$ design.

$S \times A \times B$ design Factor B				
A	b_1	b_2	b_3	b_4
a_1	s_1	s_1	s_1	s_1
	s_2	s_2	s_2	s_2
	s_3	s_3	s_3	s_3
	s_4	s_4	s_4	s_4
	s_5	s_5	s_5	s_5
a_2	s_1	s_1	s_1	s_1
	s_2	s_2	s_2	s_2
	s_3	s_3	s_3	s_3
	s_4	s_4	s_4	s_4
	s_5	s_5	s_5	s_5

Table 20.3 Layout of an $S \times A \times B$ design.

20.2 Example: bat and hat

The example will be a (fictitious) replication of an experiment by Conrad (1972). The general idea was to explore the hypothesis that young children do not use phonological coding in short-term memory. In order to do this, we selected 10 children: 5 five year olds and 5 twelve year olds. This constitutes the first independent variable (A or *age* with 2 levels), which happens also to be what we have called a ‘tag’ or ‘classificatory’ variable in Chapter 1. Because a subject is either five years old or twelve years old, the Subject factor (S) is nested in the (A) Age factor.

The second independent variable deals with phonological similarity, and we will use the letter B to symbolize it. But before describing it, we need to explain in more detail the experiment. Each child was shown 100 pairs of pictures of objects. A pilot study had made sure that children always used the same name for these pictures (i.e. the cat picture was always called ‘a cat’, never ‘a pet’ or ‘an animal’).

After the children had looked at the pictures, the pictures were turned over so that the children could only see the backs of the pictures. Then the experimenter gave an identical pair of pictures to the children and asked them to position each new picture on top of the old ones (that were hidden by now) such that the new pictures matched the hidden ones. For half of the pairs of pictures, the sound of the name of the objects was similar (i.e. hat and cat), whereas for the other half of the pairs, the sound of the names of the objects in a pair was dissimilar (i.e. horse and chair). This manipulation constitutes the second experimental factor B or ‘*phonological similarity*’. It has two levels: b_1 phonologically dissimilar and b_2 phonologically similar. The dependent variable will be the number of pairs of pictures correctly positioned by the child.

Conrad reasoned that if the older children use a phonological code to rehearse information, then it would be more difficult for them to remember the phonologically similar pairs than the phonologically dissimilar pairs. This should happen because of an interference effect. If the young children do not use a phonological code to rehearse the material they want to learn, then their performance should be unaffected by phonological similarity, and they should perform at the same level for both conditions of phonological similarity. In addition, because of the usual age effect, one can expect the older children to perform on the whole better than the

		<i>B</i> Phonological similarity		Σ	Means
		<i>b</i> ₁ similar	<i>b</i> ₂ dissimilar		
<i>a</i> ₁ Age: 5 years	<i>s</i> ₁	15	13	<i>Y</i> _{1..} 28	<i>M</i> _{1..} 14
	<i>s</i> ₂	23	19	42	21
	<i>s</i> ₃	12	10	22	11
	<i>s</i> ₄	16	16	32	16
	<i>s</i> ₅	14	12	26	13
		<u><i>Y</i>_{11.} = 80</u>	<u><i>Y</i>_{12.} = 70</u>	<u><i>Y</i>_{1..} = 150</u>	
		<i>M</i> _{11.} = 16	<i>M</i> _{12.} = 14	<i>M</i> _{1..} = 15	
<i>a</i> ₂ Age: 12 years				<i>Y</i> _{2..}	<i>M</i> _{2..}
	<i>s</i> ₆	39	29	68	34
	<i>s</i> ₇	31	15	46	23
	<i>s</i> ₈	40	30	70	35
	<i>s</i> ₉	32	26	58	29
	<i>s</i> ₁₀	38	30	68	34
		<u><i>Y</i>_{21.} = 180</u>	<u><i>Y</i>_{22.} = 130</u>	<u><i>Y</i>_{2..} = 310</u>	
		<i>M</i> _{21.} = 36	<i>M</i> _{22.} = 26	<i>M</i> _{2..} = 31	

Table 20.4 Results of a replication of Conrad’s (1972) experiment.

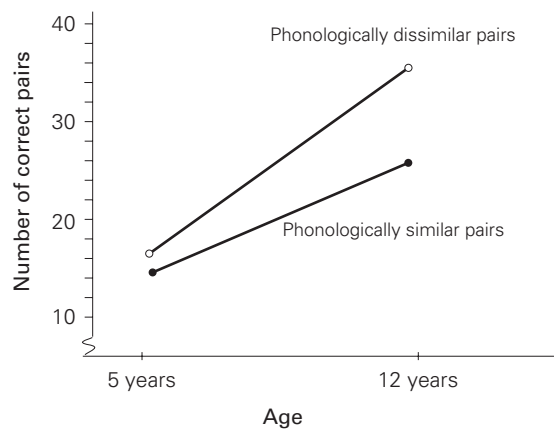


Figure 20.1 Results of a replication of Conrad’s (1971) experiment.

younger ones. Could you draw the graph corresponding to the expected pattern of results? Could you express these predictions in terms of the ANOVA model?

We expect a main effect of age (which is rather trivial), and also (and this is the *crucial* point) we expect an interaction effect. This interaction will be the really important test of Conrad’s theoretical prediction (cf. the discussion about the importance of the interaction as test for theories in Chapter 15).

The results of this replication are given in Table 20.4, along with the different sums needed for computing the ANOVA table. Make sure you understand its layout and try to determine whether the experimental predictions are supported by the results.

20.3 Sums of squares, mean squares, and F ratios

Most of the sources of variation are already familiar. We have, however, two new sources: $S(\mathcal{A})$ and $BS(\mathcal{A})$. These notations serve as reminders that because S is nested in $\mathcal{A}$, the effect of a subject cannot be dissociated from the interaction with the a th condition in which this subject uniquely appears. Also the effect of the subject cannot be dissociated from the interaction with $\mathcal{B}$. This happens because a given subject s appears only in association with a given level a of $\mathcal{A}$, as he or she also appears only in association with a given level ab of $\mathcal{AB}$.

The different sums of squares for each of the sources of variation for an $S(\mathcal{A}) \times \mathcal{B}$ design are given below:

$$SS_A = BS \sum (M_{a..} - M_{...})^2$$

$$SS_B = AS \sum (M_{.b.} - M_{...})^2$$

$$SS_{S(\mathcal{A})} = B \sum (M_{a.s} - M_{a..})^2$$

$$SS_{AB} = S \sum (M_{ab.} - M_{a..} - M_{.b.} + M_{...})^2$$

$$SS_{BS(\mathcal{A})} = \sum (Y_{abs} - M_{ab.} - M_{a.s} + M_{a..})^2$$

With the usual reasoning, we find the following values for the degrees of freedom associated with the sums of squares:

$$df_A = A - 1$$

$$df_B = B - 1$$

$$df_{S(\mathcal{A})} = A(S - 1)$$

$$df_{AB} = (A - 1)(B - 1)$$

$$df_{BS(\mathcal{A})} = A(B - 1)(S - 1).$$

If you compare the formulas for the sums of squares and degrees of freedom for the $S(\mathcal{A} \times \mathcal{B})$, $S \times \mathcal{A} \times \mathcal{B}$, and $S(\mathcal{A}) \times \mathcal{B}$ designs, you will find that the experimental sources per se (i.e. $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{AB}$) are computed in the same way for all three designs. Only the ‘residual’ terms vary from one design to the other. An even closer look may reveal that the name of a source gives the formula for computing its sum of squares, number of degrees of freedom, and mean square.

The formulas to compute the different F ratios are given in the ‘test-table’ in which we follow a long tradition of separating the *between*-subjects sources (in here $\mathcal{A}$) from the *within*-subjects sources (in here $\mathcal{B}$ and $\mathcal{AB}$).

We see from Table 20.5 that the *between* source ($\mathcal{A}$) is tested with the subjects nested in $\mathcal{A}$: namely the $S(\mathcal{A})$ source. Also the *within* sources ($\mathcal{B}$ and $\mathcal{AB}$) are tested with the interaction term $BS(\mathcal{A})$ corresponding to the interaction between the subject factor (nested in $\mathcal{A}$) and factor $\mathcal{B}$.

Source	MS_{test}
<i>Between subjects</i>	
$\mathcal{A}$	$MS_{S(\mathcal{A})}$
$S(\mathcal{A})$	----
<i>Within subjects</i>	
$\mathcal{B}$	$MS_{BS(\mathcal{A})}$
$\mathcal{AB}$	$MS_{BS(\mathcal{A})}$
$BS(\mathcal{A})$	----

Table 20.5 Mean squares for test for an $S(\mathcal{A}) \times \mathcal{B}$ design with $\mathcal{A}$ and $\mathcal{B}$ fixed

20.4 The comprehension formula routine

By now you should be thoroughly familiar with the steps for finding the F values. Once again, make sure you can do the computations.

First, find the degrees of freedom:

$$\begin{aligned}df_A &= (A - 1) = 2 - 1 = 1 \\df_B &= (B - 1) = 2 - 1 = 1 \\df_{S(\mathcal{A})} &= A(S - 1) = 2(5 - 1) = 8 \\df_{AB} &= (A - 1)(B - 1) = (2 - 1)(2 - 1) = 1 \\df_{BS(\mathcal{A})} &= A(B - 1)(S - 1) = 2(2 - 1)(5 - 1) = 2 \times 1 \times 4 = 8 \\df_{\text{total}} &= A \times B \times S - 1 = 2 \times 2 \times 5 - 1 = 19\end{aligned}$$

Next, compute SS_A , SS_B , $SS_{S(\mathcal{A})}$, SS_{AB} , and $SS_{BS(\mathcal{A})}$.

$$\begin{aligned}SS_A &= BS \sum (M_{a..} - M_{...})^2 \\&= 2 \times 5 \times [(15 - 23)^2 + (31 - 23)^2] \\&= 10 \times [(-8)^2 + 8^2] \\&= 10 \times [64 + 64] \\&= 10 \times 128 \\&= 1,280.00\end{aligned} \tag{20.1}$$

$$\begin{aligned}SS_B &= AS \sum (M_{.b.} - M_{...})^2 \\&= 2 \times 5 \times [(26 - 23)^2 + (20 - 23)^2] \\&= 10 \times [3^2 + (-3)^2] \\&= 10 \times [9 + 9] \\&= 10 \times 18 \\&= 180.00\end{aligned} \tag{20.2}$$

$$\begin{aligned}
SS_{S(A)} &= B \sum (M_{a \cdot s} - M_{a \cdot \cdot})^2 & (20.3) \\
&= 2 [(14 - 15)^2 + (21 - 15)^2 + \cdots + (29 - 31)^2 + (34 - 31)^2] \\
&= 2 \times [1^2 + 6^2 + \cdots + (-2)^2 + 3^2] \\
&= 2 \times [1 + 36 + \cdots + 4 + 9] \\
&= 2 \times 160 \\
&= 320.00.
\end{aligned}$$

$$\begin{aligned}
SS_{AB} &= S \sum (M_{ab \cdot} - M_{a \cdot \cdot} - M_{\cdot b \cdot} + M_{\cdot \cdot \cdot})^2 & (20.4) \\
&= 5 \times [(16 - 15 - 26 + 23)^2 + (14 - 15 - 20 + 23)^2 \\
&\quad + (36 - 31 - 26 + 23)^2 + (26 - 31 - 20 + 23)^2] \\
&= 5 \times [(-2)^2 + 2^2 + 2^2 + (-2)^2] \\
&= 5 \times [4 + 4 + 4 + 4] \\
&= 5 \times 16.00 \\
&= 80.00
\end{aligned}$$

$$\begin{aligned}
SS_{BS(A)} &= \sum (Y_{abs} - M_{ab \cdot} - M_{a \cdot s} + M_{a \cdot \cdot})^2 & (20.5) \\
&= [(15 - 16 - 14 + 15)^2 + (23 - 16 - 21 + 15)^2 + \cdots \\
&\quad \cdots + (26 - 26 - 29 + 31)^2 + (30 - 26 - 34 + 31)^2] \\
&= [0^2 + 1^2 + \cdots + 2^2 + 1^2] \\
&= [0 + 1 + \cdots + 4 + 1] \\
&= 32.00.
\end{aligned}$$

We now proceed to finding the mean squares:

$$\begin{aligned}
MS_A &= \frac{SS_A}{df_A} = \frac{1,280.00}{1} = 1,280.00 & (20.6) \\
MS_B &= \frac{SS_B}{df_B} = \frac{180.00}{1} = 180.00 \\
MS_{AB} &= \frac{SS_{AB}}{df_{AB}} = \frac{80.00}{1} = 80.00 \\
MS_{S(A)} &= \frac{SS_{S(A)}}{df_{S(A)}} = \frac{320.00}{8} = 40.00 \\
MS_{BS(A)} &= \frac{SS_{BS}}{df_{BS}} = \frac{32.00}{8} = 4.00.
\end{aligned}$$

Finally, we can compute the F values:

$$F_A = \frac{MS_A}{MS_{S(A)}} = \frac{1,280.00}{40.00} = 32.00 \quad (20.7)$$

$$F_B = \frac{MS_B}{MS_{BS(A)}} = \frac{180.00}{4.00} = 45.00$$

$$F_{AB} = \frac{MS_{AB}}{MS_{BS(A)}} = \frac{80.00}{4.00} = 20.00.$$

20.5 The 13-point computational routine

For computational ease, try using the computational formulas.

$$Q1: Y_{...} = 15 + 23 + \cdots + 26 + 30 = 460.00$$

$$Q2: \boxed{\mathbf{ABS}} = \sum Y_{abs}^2 = 15^2 + 23^2 + \cdots + 26^2 + 30^2 = 12,472.00$$

$$Q3: \boxed{\mathbf{A}} = \sum \frac{Y_{a..}^2}{BS} = \frac{150^2 + 310^2}{10} = 11,860.00$$

$$Q4: \boxed{\mathbf{B}} = \sum \frac{Y_{.b.}^2}{AS} = \frac{260^2 + 200^2}{10} = 10,760.00$$

$$Q5: \boxed{\mathbf{AB}} = \sum \frac{Y_{ab.}^2}{S} = \frac{80^2 + 70^2 + 180^2 + 130^2}{5} = 12,120.00$$

$$Q6: \boxed{\mathbf{AS}} = \sum \frac{Y_{a.s}^2}{B} = \frac{28^2 + 42^2 + \cdots + 58^2 + 68^2}{2} = 12,180.00$$

$$Q7: \boxed{\mathbf{1}} = \frac{Y_{...}^2}{ABS} = \frac{211,600}{20} = 10,580.00$$

$$Q8: SS_A = \boxed{\mathbf{A}} - \boxed{\mathbf{1}} = 11,860.00 - 10,580.00 = 1,280.00$$

$$Q9: SS_B = \boxed{\mathbf{B}} - \boxed{\mathbf{1}} = 10,760.00 - 10,580.00 = 180.00$$

$$Q10: SS_{S(A)} = \boxed{\mathbf{AS}} - \boxed{\mathbf{A}} = 12,180.00 - 11,860.00 = 320.00$$

$$Q11: SS_{AB} = \boxed{\mathbf{AB}} - \boxed{\mathbf{A}} - \boxed{\mathbf{B}} + \boxed{\mathbf{1}}$$

$$= 12,120.00 - 11,860.00 - 10,760.00 + 10,580.00 = 80.00$$

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p(F)</i>
<i>Between subjects</i>					
$\mathcal{A}$	1	1,280.00	1,280.00	32.00	.00056
$S(\mathcal{A})$	8	320.00	40.00	-----	
<i>Within subjects</i>					
$\mathcal{B}$	1	180.00	180.00	45.00	.00020
$\mathcal{AB}$	1	80.00	80.00	20.00	.00220
$\mathcal{BS}(\mathcal{A})$	8	32.00	4.00	-----	
Total	19	1,892.00			

Table 20.6 ANOVA table for a replication of Conrad's (1971) experiment (data from Table 20.4).

$$\begin{aligned} \text{Q12: } SS_{\mathcal{BS}(\mathcal{A})} &= \boxed{\text{ABS}} - \boxed{\text{AB}} - \boxed{\text{AS}} + \boxed{\text{A}} \\ &= 12,472.00 - 12,120.00 - 12,180.00 + 11,860.00 = 32.00 \end{aligned}$$

$$\text{Q13: } SS_{\text{total}} = \boxed{\text{ABS}} - \boxed{1} = 12,472.000 - 10,580.00 = 1,892.00 .$$

With all these Qs we can now complete the ANOVA table (see Table 20.6).

As you can see from Figure 20.1 as well as the results of the analysis of variance, the experimental predictions are supported by the experimental results. The results section would indicate the following information:

The results were treated as an age $\times$ phonological similarity analysis of variance design with age (5 year olds vs 12 year olds) being a between-subject factor and phonological similarity (similar vs dissimilar) being a within-subject factor. There was a very clear effect of age, $F(1, 8) = 32.00$, $MS_e = 40.00$, $p < .01$. The expected interaction of age by phonological similarity was also very reliable $F(1, 8) = 20.00$, $MS_e = 4.00$, $p < .01$. A main effect of phonological similarity was also detected $F(1, 8) = 45.00$, $MS_e = 4.00$, $p < .01$, but as Figure 20.1 shows, its interpretation as a main effect is delicate because of the strong interaction between phonological similarity and age.

20.6 Score model (Model I), $S(\mathcal{A}) \times \mathcal{B}$ design: $\mathcal{A}$ and $\mathcal{B}$ fixed

The formulas given previously for calculating F can, as usual, be justified by looking at the score model. In this chapter we will look only at the score model for the case when both factors are fixed or random. The extension to other cases is left to the reader: just replace the Greek letters by Roman letters when the corresponding factor is random.

Some components of this score model have been seen before, but two terms, namely $s_{s(a)}$ and $\beta_{s_{bs(a)}}$, are new. These terms remind us that a subject is assigned to only one level of $\mathcal{A}$ and so appears in only one level of $\mathcal{A}$ crossed with all levels of $\mathcal{B}$.

As for the repeated measures designs $\mathcal{S} \times \mathcal{A} \times \mathcal{B}$, the error term $e_{bs(a)}$ carries the same index as the highest level of interaction $\beta s_{bs(a)}$. Therefore, these two terms are confounded.

$$Y_{abs} = \mu \dots + \alpha_a + \beta_b + s_{s(a)} + \alpha\beta_{ab} + \beta s_{bs(a)} + e_{bs(a)}.$$

Notice the absence of interaction terms between $\mathcal{A}$ and $\mathcal{S}$. Because a subject is assigned to only one level of $\mathcal{A}$, it is impossible to estimate separately the interaction between $\mathcal{S}$ and $\mathcal{A}$ and the error.

The quantity $s_{s(a)}$ can be thought of as the sum of the $\mathcal{S} \times \mathcal{A} \times \mathcal{B}$ design terms s_s and αs_{as} . In the same manner $\beta s_{bs(a)}$ may be interpreted as the sum of the terms βs_{bs} and $\alpha\beta_{abs}$ of $\mathcal{S} \times \mathcal{A} \times \mathcal{B}$ designs.

This relation also applies to the sums of squares. Thus,

$$SS_{\mathcal{S}(\mathcal{A})} \text{ of } \mathcal{S}(\mathcal{A}) \times \mathcal{B} = SS_{\mathcal{S}} + SS_{AS} \text{ of } \mathcal{S} \times \mathcal{A} \times \mathcal{B}$$

$$SS_{BS(\mathcal{A})} \text{ of } \mathcal{S}(\mathcal{A}) \times \mathcal{B} = SS_{BS} + SS_{ABS} \text{ of } \mathcal{S} \times \mathcal{A} \times \mathcal{B}.$$

The degrees of freedom are also related this way. This general relationship makes it possible to calculate the analysis of variance for partially repeated measures from the results of fully repeated measures. This can come in handy if the statistical computer package that you are using does not provide support for partially repeated designs.

When the sums of squares for $\mathcal{S}(\mathcal{A}) \times \mathcal{B}$ designs are expressed in the terms of the score model, the following expectations for the mean squares are obtained.

$$\mathbf{E}\{MS_A\} = \sigma_e^2 + B\sigma_{s(a)}^2 + BS\vartheta_a^2$$

$$\mathbf{E}\{MS_B\} = \sigma_e^2 + \sigma_{bs(a)}^2 + AS\vartheta_b^2$$

$$\mathbf{E}\{MS_{\mathcal{S}(\mathcal{A})}\} = \sigma_e^2 + B\sigma_{s(a)}^2$$

$$\mathbf{E}\{MS_{AB}\} = \sigma_e^2 + \sigma_{bs(a)}^2 + S\vartheta_{ab}^2$$

$$\mathbf{E}\{MS_{BS(\mathcal{A})}\} = \sigma_e^2 + \sigma_{bs(a)}^2.$$

These expected values justify the test mean squares, and lack of test values, previously presented in Table 20.5.

20.7 Score model (Model II), $\mathcal{S}(\mathcal{A}) \times \mathcal{B}$ design: $\mathcal{A}$ and $\mathcal{B}$ random

The following expected values can be derived for the mean squares when $\mathcal{A}$ and $\mathcal{B}$ are random factors.

$$\mathbf{E}\{MS_A\} = \sigma_e^2 + \sigma_{bs(a)}^2 + S\sigma_{ab}^2 + B\sigma_{s(a)}^2 + BS\sigma_a^2$$

$$\mathbf{E}\{MS_B\} = \sigma_e^2 + \sigma_{bs(a)}^2 + S\sigma_{ab}^2 + AS\sigma_b^2$$

$$\mathbf{E}\{MS_{\mathcal{S}(\mathcal{A})}\} = \sigma_e^2 + \sigma_{bs(a)}^2 + B\sigma_{s(a)}^2$$

$$\mathbf{E}\{MS_{AB}\} = \sigma_e^2 + \sigma_{bs(a)}^2 + S\sigma_{ab}^2$$

$$\mathbf{E}\{MS_{BS(\mathcal{A})}\} = \sigma_e^2 + \sigma_{bs(a)}^2$$

From these expectations the following table can be constructed. Note that the sources are separated into between- and within-subjects effects.

Source	MS_{test}
<i>Between subjects</i>	
$\mathcal{A}$	-----
$\mathcal{S}(\mathcal{A})$	$MS_{\mathcal{S}(\mathcal{A})}$
<i>Within subjects</i>	
$\mathcal{B}$	$MS_{BS(\mathcal{A})}$
$\mathcal{AB}$	$MS_{BS(\mathcal{A})}$
$BS(\mathcal{A})$	-----

As before, the dashes signify the sources of effect that cannot be tested directly. But we can use a Quasi- F again (see Chapter 19, Section 19.8, page 369). Specifically, it is possible to test the effect of $\mathcal{A}$ with

$$MS'_{\text{test},\mathcal{A}} = MS_{AB} + MS_{\mathcal{S}(\mathcal{A})} - MS_{BS(\mathcal{A})}.$$

20.8 Score for Model III, $\mathcal{S}(\mathcal{A}) \times \mathcal{B}$ design: $\mathcal{A}$ fixed and $\mathcal{B}$ random

Because the independent variables $\mathcal{A}$ and $\mathcal{B}$ do not have symmetric roles [unlike designs $\mathcal{S}(\mathcal{A} \times \mathcal{B})$ and $\mathcal{S} \times \mathcal{A} \times \mathcal{B}$], two cases must be specified for $\mathcal{S}(\mathcal{A}) \times \mathcal{B}$ Model III. In one case $\mathcal{A}$ is random and $\mathcal{B}$ is fixed, while in the other case $\mathcal{A}$ is fixed and $\mathcal{B}$ is random.

In the first case ($\mathcal{A}$ random, $\mathcal{B}$ fixed) the expected values of the mean squares¹ give rise to the following test table:

Source	MS_{test}
<i>Between subjects</i>	
$\mathcal{A}$	$MS_{\mathcal{S}(\mathcal{A})}$
$\mathcal{S}(\mathcal{A})$	-----
<i>Within subjects</i>	
$\mathcal{B}$	MS_{AB}
$\mathcal{AB}$	$MS_{BS(\mathcal{A})}$
$BS(\mathcal{A})$	-----

¹ If you want to know what are these expected values, the rules to compute them are given in Chapter 22.

When $\mathcal{A}$ is fixed and $\mathcal{B}$ is random, the test table obtained is

Source	MS_{test}
<i>Between subjects</i>	
$\mathcal{A}$	-----
$\mathcal{S}(\mathcal{A})$	$MS_{\mathcal{BS}(\mathcal{A})}$
<i>Within subjects</i>	
$\mathcal{B}$	$MS_{\mathcal{BS}(\mathcal{A})}$
$\mathcal{AB}$	$MS_{\mathcal{BS}(\mathcal{A})}$
$\mathcal{BS}(\mathcal{A})$	-----

The considerable difference between these two tables shows the importance of distinguishing between fixed and random factors. The correct choice of a statistical model clearly depends upon this decision.

20.9 Coefficients of intensity

The coefficient $R^2_{Y \cdot \mathcal{A}}$ can be used here as usual. For each of the identifiable sources of variation we divide the sum of squares of effect by the total sum of squares. For example, the importance of the interaction $\mathcal{AB}$ is evaluated by computing the ratio

$$R^2_{Y \cdot \mathcal{AB}} = \frac{SS_{\mathcal{AB}}}{SS_{\text{total}}}.$$

The importance of the interaction $\mathcal{BS}(\mathcal{A})$ is evaluated by computing the ratio

$$R^2_{Y \cdot \mathcal{BS}(\mathcal{A})} = \frac{SS_{\mathcal{BS}(\mathcal{A})}}{SS_{\text{total}}}.$$

The interpretation of $R^2_{Y \cdot \mathcal{A}}$ as a coefficient of correlation is the same as seen previously. The derivation is left as an exercise.

It is also possible to define partial correlation coefficients. For example, this will be the case if we want to evaluate the effect of $\mathcal{A}$ relative to the *experimental* sources of variation (i.e. the sum of $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{AB}$). In this case we will divide the sum of squares for $\mathcal{A}$ by the sum of the sums of squares of $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{AB}$. Any other combination could be valid as long as it makes sense for the experimenter and casts some light on the research question at stake.

20.10 Validity of $\mathcal{S}(\mathcal{A}) \times \mathcal{B}$ designs

The $\mathcal{S}(\mathcal{A}) \times \mathcal{B}$ design can be seen as a compromise between the $\mathcal{S}(\mathcal{A} \times \mathcal{B})$ and the $\mathcal{S} \times \mathcal{A} \times \mathcal{B}$ designs. The set of technical assumptions and validity conditions required for these two cases are still supposed to hold here. An additional condition is required, however. It is called in general ‘homogeneity of the covariance matrices’. It means essentially that the covariance matrices computed from the $\mathcal{B}$ factor should be independent of the level of $\mathcal{A}$. This is equivalent to stating that there is no interaction between the covariance configuration for $\mathcal{B}$ and the factor $\mathcal{A}$. This problem is detailed in the references given in Chapter 18 (Section 18.12, pages 348ff.).

20.11 Prescription

An $S(\mathcal{A}) \times \mathcal{B}$ design can be interpreted as a juxtaposition of several $S \times \mathcal{B}$ designs (one per level of $\mathcal{A}$), or a repeated measurement of an $S(\mathcal{A})$ design.

When we introduced the $S \times \mathcal{A}$ designs, we mentioned that an important advantage of those designs was their greater sensitivity or power [as compared with an equivalent $S(\mathcal{A})$ design]. This advantage carries over to the more complex designs, such as the ones studied in this chapter.

The effect of the *within*-subjects factor $\mathcal{B}$, is, in general, easier to detect than the effect on the *between*-subjects factor $\mathcal{A}$. Consequently, if a choice is possible, it is a good idea to use repeated measures for the experimental factor that we judge of greater interest, or for which we think that we need the most powerful test. However, it is not always possible to choose. In particular when a factor is a tag or subject factor, it will be impossible to obtain repeated measures.

Chapter summary

20.12 Key notions of the chapter

Below are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

$$SS_{BS(A)}$$

$$R^2_{Y \cdot BS(A)}$$

20.13 Key formulas of the chapter

Below are the main formulas introduced in this chapter: try to go through them and understand what they mean.

$$SS_A = BS \sum (M_{a..} - M_{...})^2$$

$$SS_B = AS \sum (M_{.b.} - M_{...})^2$$

$$SS_{S(A)} = B \sum (M_{a.s} - M_{a..})^2$$

$$SS_{AB} = S \sum (M_{ab.} - M_{a..} - M_{.b.} + M_{...})^2$$

$$SS_{BS(A)} = \sum (Y_{abs} - M_{ab.} - M_{a.s} + M_{a..})^2$$

$$R^2_{Y \cdot BS(A)} = \frac{SS_{BS(A)}}{SS_{\text{total}}}$$

20.14 Key questions of the chapter

Below are two questions about the content of this chapter. Both the answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- * What are the advantages provided by an $\mathcal{S}(\mathcal{A}) \times \mathcal{B}$ design?
- * Why is extra thought required when deciding if factors will be fixed or random?

21

ANOVA, nested factorial design: $\mathcal{S} \times \mathcal{A}(\mathcal{B})$

21.1 Introduction

For all the two-factor designs that we have seen so far, the experimental sum of squares was always partitioned the same way in three experimental sources of variation: $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{AB}$. In this chapter, we will see a new partition of the experimental sum of squares. This new partition occurs because, for the first time, the two experimental factors are not completely crossed. Specifically, one of the factors (called $\mathcal{A}$ for convenience) is nested in the other one (called $\mathcal{B}$). This happens typically when $\mathcal{A}$ is an *item* factor such as words, faces, melodies, etc.

For example, suppose we study how long it takes for Caucasian subjects to identify as faces, 10 faces of their own race, 10 Japanese faces, and 10 African faces. There are two experimental factors here. The race of the faces (with three levels: Caucasian, Japanese, African) is one factor. Let's call this factor: $\mathcal{B}$. (Why $\mathcal{B}$ rather than $\mathcal{A}$ will soon be made clear.) It is, presumably, a fixed factor because any replication of the experiment will use the same levels of this variable. The faces make another factor. Let's call this factor: $\mathcal{A}$. But, a given face can belong to only one race (i.e. a face is either Caucasian, or Japanese, or African). Therefore, the faces are *nested* in the race factor. This is symbolized by the notation $\mathcal{A}(\mathcal{B})$ (read $\mathcal{A}$ nested in $\mathcal{B}$, or in words: 'Faces nested in Race', or even briefer: 'Face in Race').

In this design, $\mathcal{A}$ is almost always an item factor, and, consequently, is almost always a *random factor*, whereas $\mathcal{B}$ is a fixed factor. So we will look in detail only at this case and mention briefly the other cases created by the various choices of fixed and random as the experimental factors. An additional word of caution is needed here. Even though $\mathcal{B}$ is generally a fixed factor, it is almost always a *tag* variable. This implies that it cannot be manipulated (if it could, then we could have $\mathcal{A}$ crossed with $\mathcal{B}$). Because of this, it is often difficult to separate the effect of the variable itself from all the potential confounding factors. For example, if atypical faces are easier to recognize than typical faces, is it due

to typicality *per se* or is it due to all the variables that are confounded with typicality? For example, typical faces are judged more average, and also are often judged more beautiful than atypical faces. What variable(s) is (are) really responsible of the memory effect? Is it typicality, or beauty, or being average, or something else?

21.2 Example: faces in space

Some faces give the impression of being original or bizarre. Some other faces, by contrast, give the impression of being average or common. We say that original faces are *atypical*; and that common faces are *typical*. In terms of design factors, we say that: faces vary on the typicality factor (which has two levels: typical vs atypical).¹

In this example, we are interested in the effect of typicality on reaction time. Presumably, typical faces should be easier to process as faces than atypical faces. In this example,² we measured the reaction time of four subjects in a face identification task. Ten faces, mixed with ten ‘distractor faces’, were presented on the screen of a computer. The distractor faces were jumbled faces (e.g. with the nose at the place of the mouth). Five of the ten faces were typical faces, and the other five faces were atypical faces. Subjects were asked to respond as quickly as they could. Only the data recorded from the normal faces (i.e. not jumbled) were kept for further analysis. All the subjects identified correctly the faces as faces. The data (which are made of ‘nice numbers’) are given in Table 21.1. As usual, make sure that you understand the layout, and try to determine whether there is some effect of typicality.

Here, as in most $\mathcal{S} \times \mathcal{A}(\mathcal{B})$ designs, we are mainly interested in the nesting factor (i.e. $\mathcal{B}$). The nested factor [i.e. $\mathcal{A}(\mathcal{B})$] is not, however, without interest. If it is statistically significant, this may indicate that the pattern of effects, which we see in the results, depends upon the *specific* sample of items used in this experiment.

21.2.1 A word of caution: it is very hard to be random

As we mentioned previously, the nested factor [i.e. $\mathcal{A}(\mathcal{B})$], is, in general, an item factor. We take it into account because we need it to show the effect of $\mathcal{B}$ on this type of item. For example we need faces to show the effect of typicality (*on* faces). The statistical theory used to perform the analysis *assumes* that the items (e.g. the faces) are a *random* sample which is representative of the set of possible items. Unfortunately, it is often difficult to draw a random sample of items. For example, how can I be sure that the set of faces obtained is random? Think about the set of biases that can be at work: these faces are probably faces of students, friends of students, or friends of the experimenters (we need to have volunteers agreeing to give images of their faces to science!). This is hardly random! Problems like these are frequent in these designs. And often, a thorny question is to evaluate if the results of an experiment may have been influenced by a biased sample of items.

¹ Note that we can think of (or ‘operationalize’) the typicality factor as a quantitative factor, and then use a multiple regression approach instead of an ANOVA.

² Somewhat fictitious, but close to some standard experiments in face recognition.

Factor $\mathcal{B}$													
(Typicality: Typical vs Atypical)													
b_1 : (Typical)							b_2 : (Atypical)						
$\mathcal{A}_{a(b_1)}$: (Typical Faces)							$\mathcal{A}_{a(b_2)}$: (Atypical Faces)						
	a_1	a_2	a_3	a_4	a_5	M_{1s}	a_1	a_2	a_3	a_4	a_5	M_{2s}	$M_{..s}$
s_1	20	22	25	24	19	22	37	37	43	48	45	42	32
s_2	9	8	21	21	21	16	34	35	35	37	39	36	26
s_3	18	20	18	21	33	22	35	39	39	37	40	38	30
s_4	5	14	16	22	23	16	38	49	51	50	52	48	32
	$M_{11\cdot}$	$M_{21\cdot}$	$M_{31\cdot}$	$M_{41\cdot}$	$M_{51\cdot}$		$M_{12\cdot}$	$M_{22\cdot}$	$M_{32\cdot}$	$M_{42\cdot}$	$M_{52\cdot}$		
	13	16	20	22	24		36	40	42	43	44	$M_{\dots} = 30$	
	$M_{1\cdot} = 19$						$M_{2\cdot} = 41$						

Table 21.1 Data from a fictitious experiment with an $S \times \mathcal{A}(\mathcal{B})$ design. Factor $\mathcal{B}$ is Typicality. Factor $\mathcal{A}(\mathcal{B})$ is Faces (nested in Typicality). There are 4 subjects in this experiment. The dependent variable is measured in centiseconds (1 centisecond equals 10 milliseconds); and it is the time taken by a subject to respond that a given face was a face.

21.3 How to analyze an $S \times \mathcal{A}(\mathcal{B})$ design

Most of the sources of variation are already familiar. We have, however, two new sources: $\mathcal{A}(\mathcal{B})$ and $\mathcal{AS}(\mathcal{B})$. These notations serve as reminders that since the item factor $\mathcal{A}(\mathcal{B})$ is nested in $\mathcal{B}$, the effect of $\mathcal{A}(\mathcal{B})$ cannot be dissociated from its interaction with the b condition of $\mathcal{B}$ in which the item uniquely appears. Also, because a given combination of the item a and the subject s appears only in association with a given level b of $\mathcal{B}$, the interaction of $\mathcal{A}(\mathcal{B})$ and S is itself nested in $\mathcal{B}$. Therefore, the effect of the combination of the a th item and the s th subject cannot be dissociated from their (second-order) interaction with $\mathcal{B}$.³

21.3.1 Sums of squares

The different sums of squares for each of the sources of variation for an $S \times \mathcal{A}(\mathcal{B})$ design are given below:

$$SS_{\mathcal{A}(\mathcal{B})} = S \sum (M_{ab.} - M_{.b.})^2 \quad (21.1)$$

$$SS_{\mathcal{B}} = AS \sum (M_{.b.} - M_{...})^2 \quad (21.2)$$

$$SS_S = AB \sum (M_{..s} - M_{...})^2 \quad (21.3)$$

³ If you have the impression of being in the ‘no context’ condition of Romeo and Juliet, don’t panic. Keep on reading. It will get better!

$$SS_{BS} = A \sum (M_{.bs} - M_{.b.} - M_{..s} + M_{...})^2 \quad (21.4)$$

$$SS_{\mathcal{A}S(\mathcal{B})} = \sum (Y_{abs} - M_{ab.} - M_{.bs} + M_{.b.})^2 \quad (21.5)$$

The following sums of squares and their formulas are already familiar: $\mathcal{B}$, $\mathcal{S}$, and $\mathcal{BS}$. The two new sums of squares comes from factor $\mathcal{A}(\mathcal{B})$ being nested in $\mathcal{B}$. The sum of squares of $\mathcal{A}$ is obtained by looking at the deviation of each item from the mean of the b condition in which it appears (e.g. we look at the deviation of a typical face to the mean of all the typical faces). That should be straightforward.

A bit more difficult is the formula of the sum of squares of the interaction $\mathcal{A} \times \mathcal{S}(\mathcal{B})$. Look at Equation 21.5 which shows the general form of an interaction between two factors ($\mathcal{A}$ and $\mathcal{S}$). In the case that we are familiar with, this interaction is indicated by the pattern of the indices: ' $a \cdot s - a \cdot - - \cdot s + \dots$ '. Instead, we have this pattern of indices: ' $abs - ab \cdot - \cdot bs + \cdot b \cdot$ '. The difference between these two patterns comes from the index b which is always present in the second one but absent in the first one. This is because any given combination of a and s occurs in a specific level of $\mathcal{B}$ (e.g. when a subject looks at a specific face, this face is either a typical face *or* an atypical face). Another way to look at this formula is to interpret it as the residual of the total deviation after subtraction of all the other deviations corresponding to the sources of variation that we have already identified. With a formula, we will say that the $\mathcal{A} \times \mathcal{S}(\mathcal{B})$ deviation is obtained as:

$$\begin{aligned} (Y_{abs} - M_{ab.} - M_{.bs} + M_{.b.}) &= (Y_{abs} - M_{...}) - (M_{ab.} - M_{.b.}) - (M_{.b.} - M_{...}) \\ &\quad - (M_{..s} - M_{...}) - (M_{.bs} - M_{.b.} - M_{..s} + M_{...}). \end{aligned} \quad (21.6)$$

Or in words:

$$\begin{aligned} \text{Deviation } \mathcal{A} \times \mathcal{S}(\mathcal{B}) &= \text{Total deviation} - \text{Deviation } \mathcal{A}(\mathcal{B}) \\ &\quad - \text{Deviation } \mathcal{B} - \text{Deviation } \mathcal{S} - \text{Deviation } \mathcal{BS}. \end{aligned} \quad (21.7)$$

The sum of squares of $\mathcal{A} \times \mathcal{S}(\mathcal{B})$ is obtained by squaring and adding all the deviations for $\mathcal{A} \times \mathcal{S}(\mathcal{B})$.

It can be shown, as an exercise, that the different sums of squares (as shown in Equations 21.1 to 21.5) in an $S \times \mathcal{A}(\mathcal{B})$ are pairwise orthogonal (i.e. the coefficient of correlation between any pair of deviations will always be zero). As a consequence, the sum of all these sums of squares will be equal to the total sum of squares.

21.3.2 Degrees of freedom and mean squares

With each sum of squares is associated its number of degrees of freedom. With the usual reasoning, we find the following values for the degrees of freedom associated with the sums of squares:

$$df_{\mathcal{A}(\mathcal{B})} = B(A - 1)$$

$$df_{\mathcal{B}} = (B - 1)$$

$$df_{\mathcal{S}} = (S - 1)$$

$$df_{BS} = (B - 1)(S - 1)$$

$$df_{A(S)} = B(A - 1)(S - 1) .$$

Having the degrees of freedom and the sums of squares, we can compute the mean squares, as usual, with a division:

$$MS_{A(B)} = \frac{SS_{A(B)}}{df_{A(B)}}$$

$$MS_B = \frac{SS_B}{df_B}$$

$$MS_S = \frac{SS_S}{df_S}$$

$$MS_{BS} = \frac{SS_{BS}}{df_{BS}}$$

$$MS_{AS(B)} = \frac{SS_{AS(B)}}{df_{AS(B)}} .$$

21.3.3 F and quasi- F ratios

Remember the standard procedure? In order to evaluate the reliability of a source of variation, we need to find the expected value of its mean square. Then we assume that the null hypothesis is true, and we try to find another source whose mean square has the same expected value. This mean square is the test mean square. Dividing the first mean square (the effect mean square) by the test mean square gives an F ratio. When there is no test mean square, there is no way to compute an F ratio. However, combining several mean squares may give a test mean square called a test ‘quasi-mean square’ or a ‘test mean square prime’ (see Chapter 19, 19.8, page 369). The ratio of the effect mean square by its ‘quasi-mean square’ gives a ‘quasi- F ratio’ (or F'). The expected values of the mean squares for an $\mathcal{S} \times \mathcal{A}(\mathcal{B})$ design with $\mathcal{A}(\mathcal{B})$ random and $\mathcal{B}$ fixed are given in Table 21.2.

From Table 21.2, we find that most sources of variation may be evaluated by computing F ratios using the $\mathcal{A} \times \mathcal{S}(\mathcal{B})$ mean square. Unfortunately, the experimental factor of prime interest (i.e. $\mathcal{B}$), cannot be tested this way, but requires the use of a quasi- F' . The test mean square for the main effect of $\mathcal{B}$ is obtained as

$$MS'_{\text{test}, \mathcal{B}} = MS_{A(B)} + MS_{BS} - MS_{AS(B)} . \quad (21.8)$$

Source	Expected mean squares	MS_{test}
$\mathcal{B}$	$\sigma_e^2 + \sigma_{as(b)}^2 + A\sigma_{bs}^2 + S\sigma_{a(b)}^2 + AS\vartheta_b^2$	$MS_{A(B)} + MS_{BS} - MS_{AS(B)}$
$\mathcal{S}$	$\sigma_e^2 + \sigma_{as(b)}^2 + AB\sigma_s^2$	$MS_{AS(B)}$
$\mathcal{A}(\mathcal{B})$	$\sigma_e^2 + \sigma_{as(b)}^2 + S\sigma_{a(b)}^2$	$MS_{AS(B)}$
$\mathcal{BS}$	$\sigma_e^2 + \sigma_{as(b)}^2 + A\sigma_{bs}^2$	$MS_{AS(B)}$
$\mathcal{AS}(\mathcal{B})$	$\sigma_e^2 + \sigma_{as(b)}^2$	

Table 21.2 The expected mean squares when $\mathcal{A}$ is random and $\mathcal{B}$ is fixed for an $\mathcal{S} \times \mathcal{A}(\mathcal{B})$ design.

The number of degrees of freedom of the mean square of test is approximated by the following formula:

$$v'_2 = \frac{(MS_{A(B)} + MS_{BS} - MS_{AS(B)})^2}{\frac{MS_{A(B)}^2}{df_{A(B)}} + \frac{MS_{BS}^2}{df_{BS}} + \frac{MS_{AS(B)}^2}{df_{AS(B)}}}. \quad (21.9)$$

21.4 Back to the example: faces in space

From Table 21.1, we find the following values for the sums of squares.

For the source $\mathcal{A}(\mathcal{B})$:

$$\begin{aligned} SS_{A(B)} &= S \sum (M_{ab.} - M_{.b.})^2 \\ &= 4 \times [(13 - 19)^2 + \dots + (24 - 19)^2 + (36 - 41)^2 + \dots + (44 - 41)^2] \\ &= 480.00 . \end{aligned}$$

For the source $\mathcal{B}$:

$$\begin{aligned} SS_B &= AS \sum (M_{.b.} - M_{...})^2 \\ &= 5 \times 4 \times [(19 - 30)^2 + (41 - 30)^2] \\ &= 4,840.00 . \end{aligned} \quad (21.10)$$

For the source $\mathcal{S}$:

$$\begin{aligned} SS_S &= AB \sum (M_{..s} - M_{...})^2 \\ &= 5 \times 2 \times [(32 - 30)^2 + \dots + (32 - 30)^2] \\ &= 240.00 . \end{aligned} \quad (21.11)$$

For the source $\mathcal{BS}$:

$$\begin{aligned} SS_{BS} &= A \sum (M_{.bs} - M_{.b.} - M_{..s} + M_{...})^2 \\ &= 5 \times [(22 - 19 - 32 + 30)^2 + \dots + (48 - 41 - 32 + 30)^2] \\ &= 360.00 . \end{aligned} \quad (21.12)$$

For the source $\mathcal{A} \times \mathcal{S}(\mathcal{B})$:

$$\begin{aligned} SS_{AS(B)} &= \sum (Y_{abs} - M_{ab.} - M_{.bs} + M_{.b.})^2 \\ &= (20 - 13 - 22 + 19)^2 + \dots + (52 - 44 - 48 + 41)^2 \\ &= 360.00 . \end{aligned} \quad (21.13)$$

We can arrange all these values into the standard ANOVA table (see Table 21.3).

A quasi- F' can be computed for the main effect of $\mathcal{B}$. The first step is to find the test mean square. It is obtained as

$$MS'_{\text{test}, \mathcal{B}} = MS_{A(B)} + MS_{BS} - MS_{AS(B)} = 60.00 + 120.00 - 15.00 = 165.00 . \quad (21.14)$$

Source	R^2	df	SS	MS	F	$p(F)$	ν_1	ν_2
Face $\mathcal{A}(\mathcal{B})$	0.08	8	480.00	60.00	4.00	.0040	8	24
Typicality $\mathcal{B}$	0.77	1	4,840.00	4,840.00	29.33 [†]	.0031 [†]	1	5 [†]
Subject $\mathcal{S}$	0.04	3	240.00	80.00	5.33	.0060	3	24
Subject by Face $\mathcal{AS}(\mathcal{B})$	0.06	24	360.00	15.00				
Subject by Typicality $\mathcal{BS}$	0.06	3	360.00	120.00	8.00	.0008	3	24
Total	1.00	39	13,254.00					

[†] This value has been obtained using a quasi- F' approach. See text for explanation.

Table 21.3 ANOVA table for the data from Table 21.1.

Then we need to find the number of degrees of freedom of the mean square of test. It is equal to

$$\nu'_2 = \frac{(MS_{\mathcal{A}(\mathcal{B})} + MS_{\mathcal{BS}} - MS_{\mathcal{AS}(\mathcal{B})})^2}{\frac{MS_{\mathcal{A}(\mathcal{B})}^2}{df_{\mathcal{A}(\mathcal{B})}} + \frac{MS_{\mathcal{BS}}^2}{df_{\mathcal{BS}}} + \frac{MS_{\mathcal{AS}(\mathcal{B})}^2}{df_{\mathcal{AS}(\mathcal{B})}}} = \frac{165.00^2}{\frac{60.00^2}{8} + \frac{120.00^2}{3} + \frac{15.00^2}{24}} \approx 5.18. \quad (21.15)$$

The quasi- F ratio is equal to

$$F'_B = \frac{MS_B}{MS'_{\text{test}, \mathcal{B}}} = \frac{4,840.00}{165.00} = 29.33. \quad (21.16)$$

Under the null hypothesis, this F'_B is distributed as a Fisher distribution with $\nu_1 = B - 1 = 1$ and $\nu_2 = \nu'_2 \approx 5$ degrees of freedom. The critical value for $\alpha = .01$, with these degrees of freedom, is equal to 16.26. Because $F'_B = 29.33$ is larger than the critical value for $\alpha = .01$, we can reject the null hypothesis and conclude that typicality does affect reaction time.

In American Psychological Association style we would indicate that

Typicality affects subjects' reaction time. They were faster at detecting that a typical face was a face (tested using a quasi- F), $F'(1, 5) = 29.22$, $MS_e' = 165.00$, $p < .01$. The analysis of variance detected, in addition, an effect of faces on reaction time, $F(8, 24) = 4.00$, $MS_e = 15.00$, $p < .01$. This effect was small compared to the effect of the typicality factor ($R^2 = .08$ vs $R^2 = .77$).

21.5 What to do with $\mathcal{A}$ fixed and $\mathcal{B}$ fixed

See Table 21.4.

Source	Expected mean squares	MS_{test}
$\mathcal{B}$	$\sigma_e^2 + A\sigma_{bs}^2 + AS\vartheta_b^2$	$MS_{\mathcal{BS}}$
$\mathcal{S}$	$\sigma_e^2 + AB\sigma_s^2$	—
$\mathcal{A}(\mathcal{B})$	$\sigma_e^2 + \sigma_{as(b)}^2 + S\vartheta_{a(b)}^2$	$MS_{\mathcal{AS}(\mathcal{B})}$
$\mathcal{BS}$	$\sigma_e^2 + A\sigma_{bs}^2$	—
$\mathcal{AS}(\mathcal{B})$	$\sigma_e^2 + \sigma_{as(b)}^2$	

Table 21.4 The expected mean squares when $\mathcal{A}$ is fixed and $\mathcal{B}$ is fixed for an $\mathcal{S} \times \mathcal{A}(\mathcal{B})$ design.

21.6 When $\mathcal{A}$ and $\mathcal{B}$ are random factors

See Table 21.5.

Source	Expected mean squares	MS_{test}
$\mathcal{B}$	$\sigma_e^2 + \sigma_{as(b)}^2 + A\sigma_{bs}^2 + S\sigma_{a(b)}^2 + AS\vartheta_b^2$	$MS_{A(B)} + MS_{BS} - MS_{AS(B)}$
$\mathcal{S}$	$\sigma_e^2 + \sigma_{as(b)}^2 + A\sigma_{bs}^2 + AB\sigma_s^2$	MS_{BS}
$\mathcal{A}(B)$	$\sigma_e^2 + \sigma_{as(b)}^2 + S\sigma_{a(b)}^2$	$MS_{AS(B)}$
BS	$\sigma_e^2 + \sigma_{as(b)}^2 + A\sigma_{bs}^2$	$MS_{AS(B)}$
$AS(B)$	$\sigma_e^2 + \sigma_{as(b)}^2$	

Table 21.5 The expected mean squares when $\mathcal{A}$ is random and $\mathcal{B}$ is random for an $S \times \mathcal{A}(B)$ design.

21.7 When $\mathcal{A}$ is fixed and $\mathcal{B}$ is random

Even though this case is pretty hard to imagine (how can an item be fixed when its category is random?), we have decided to mention it for the sake of completeness. See Table 21.6.

Source	Expected mean squares	MS_{test}
$\mathcal{B}$	$\sigma_e^2 + A\sigma_{bs}^2 + AS\sigma_b^2$	MS_{BS}
$\mathcal{S}$	$\sigma_e^2 + A\sigma_{bs}^2 + AB\sigma_s^2$	MS_{BS}
$\mathcal{A}(B)$	$\sigma_e^2 + \sigma_{as(b)}^2 + S\vartheta_{a(b)}^2$	$MS_{AS(B)}$
BS	$\sigma_e^2 + A\sigma_{bs}^2$	—
$AS(B)$	$\sigma_e^2 + \sigma_{as(b)}^2$	

Table 21.6 The expected mean squares when $\mathcal{A}$ is fixed and $\mathcal{B}$ is random for an $S \times \mathcal{A}(B)$ design.

Chapter summary

21.8 New notations

We have only two new notations in this chapter. Test yourself on their meaning.

$\mathcal{A}(B)$ and $AS(B)$.

21.9 Key formulas of the chapter

Below are the main formulas introduced in this chapter: try to go through them and understand what they mean.

$$SS_{A(B)} = S \sum (M_{ab\cdot} - M_{\cdot b})^2$$

(21.17)

$$SS_B = AS \sum (M_{.b.} - M_{...})^2 \quad (21.18)$$

$$SS_S = AB \sum (M_{..s} - M_{...})^2 \quad (21.19)$$

$$SS_{BS} = A \sum (M_{.bs} - M_{.b.} - M_{..s} + M_{...})^2 \quad (21.20)$$

$$SS_{AS(B)} = \sum (Y_{abs} - M_{ab.} - M_{.bs} + M_{.b.})^2 \quad (21.21)$$

$$MS_{A(B)} = \frac{SS_{A(B)}}{df_{A(B)}}$$

$$MS_B = \frac{SS_B}{df_B}$$

$$MS_S = \frac{SS_S}{df_S}$$

$$MS_{BS} = \frac{SS_{BS}}{df_{BS}}$$

$$MS_{AS(B)} = \frac{SS_{AS(B)}}{df_{AS(B)}}$$

$$MS'_{\text{test},B} = MS_{A(B)} + MS_{BS} - MS_{AS(B)} \quad (21.22)$$

$$v'_2 = \frac{(MS_{A(B)} + MS_{BS} - MS_{AS(B)})^2}{\frac{MS_{A(B)}^2}{df_{A(B)}} + \frac{MS_{BS}^2}{df_{BS}} + \frac{MS_{AS(B)}^2}{df_{AS(B)}}} \quad (21.23)$$

21.10 Key questions of the chapter

Below are two questions about the content of this chapter. Both the answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- ★ In an $\mathcal{S} \times \mathcal{A}(\mathcal{B})$ why is it rare to have $\mathcal{A}$ fixed?
- ★ How can we analyze an $\mathcal{S} \times \mathcal{A}(\mathcal{B})$ design as an $\mathcal{S} \times \mathcal{B}$ design? What are the pros and cons of such a strategy?

22

How to derive expected values for any design

22.1 Introduction

So far we have looked at the basic experimental designs (in fact, most of the two-factor designs). However, even a rapid perusing of any scientific journal shows that most published research articles nowadays involve designs with more than two factors. Essentially, the analysis of complex designs works the same way as for the basic designs that we have seen so far, *as long as these designs can be written with the crossing and nesting relations*. In other words, if we can write down the formula of a design, we are guaranteed to be able to analyze it. Note that some design formulas cannot be written with these two relations but can still be analyzed easily.¹ (For these designs, see Winer *et al.* 1991; Kirk, 1995; Keppel and Wickens, 2004; Tabachnick and Fidell, 2007.) However, a large number of designs whose formulas are too complicated to be written down with these two relations are likely to be difficult or impossible to analyze. The goal of this chapter is to give a set of simple rules that can be used to derive, from the formula of a design, the following information:

1. What are the sources of variance for this design?
2. What is the score model of the design?
3. What are the number of degrees of freedom associated with each source? From these, what are the comprehension and computational formulas of the sums of squares corresponding to each source of variance?
4. What are the expected values of the mean squares of the sources of variance?

From that information, we can compute an F or a quasi- F for the sources of variation of interest for a particular design (or in some cases we may find that there is no way of computing such an F ratio). Even though most of the computations in real life are likely to be performed by a computer program, computing the expected value of the mean squares is almost always a necessity

¹ The most notorious ones are the latin and greco-latin square designs.

because most computer programs (including the most popular ones) can easily perform faulty analyses. Therefore, it is important to be able to check that the program at hand performs the correct statistical test.

22.2 Crossing and nesting refresher

22.2.1 Crossing

Two factors are *crossed* when each level of one factor is present in conjunction with all levels of the other factor. The crossing relation is denoted $\times$ (i.e. the multiplication sign). For example if $\mathcal{A}$ is a two level factor and if $\mathcal{B}$ is a three-level factor, $\mathcal{A}$ is crossed with $\mathcal{B}$ if all the $2 \times 3 = 6$ combinations of levels of $\mathcal{A}$ and $\mathcal{B}$ are present in the design. In other words $\mathcal{A} \times \mathcal{B}$ gives the following conditions:

$$\begin{array}{ccc} a_1b_1 & a_1b_2 & a_1b_3 \\ a_2b_1 & a_2b_2 & a_2b_3 \end{array} \quad (22.1)$$

22.2.2 Nesting

Factor $\mathcal{A}$ is nested in Factor $\mathcal{B}$ if a given level of $\mathcal{A}$ occurs in conjunction with *one and only one* level of Factor $\mathcal{B}$. For example, suppose we test three experimental groups (call them Groups Blue, Green, and Violet) in a first school, and three other experimental groups in another school (call them Groups Red, Orange, and Yellow). We have two factors: Groups (with three levels) and Schools (with two levels). A given group can be found in one and only one School. Therefore, the Factor Group is nested in the Factor School. We denote the nesting relation with parentheses. If we denote the group factor $\mathcal{A}$ and the school factor $\mathcal{B}$, we denote that $\mathcal{A}$ is nested in $\mathcal{B}$ by

$$\mathcal{A}(\mathcal{B}) \quad \text{or with real names: Group(School)}, \quad (22.2)$$

(read ‘Group is nested in School, or in brief ‘Group in School’). In the previous example, we say that $\mathcal{A}$ is the *nested* factor and that $\mathcal{B}$ is the *nesting* factor.

22.2.2.1 A notational digression

There are several ways of denoting the nesting relation. Some authors prefer to use square brackets or chevrons. So these authors will denote that $\mathcal{A}$ is nested in $\mathcal{B}$ by

$$\mathcal{A}[\mathcal{B}] \quad \text{or } \mathcal{A}\langle\mathcal{B}\rangle. \quad (22.3)$$

This should not be too much of a problem. There is another notation, which is potentially confusing. This notation uses the ‘slash’ (/) sign to indicate the nesting relation. So $\mathcal{A}$ nested in $\mathcal{B}$ will be denoted $\mathcal{A}/\mathcal{B}$. With only two factors, this notation is fine, but with more than two factors it can be ambiguous because it does not indicate clearly what is the nesting set of factors.

22.2.2.2 Back to nesting

A factor can be nested in a factor which is itself nested in a third factor. For example, suppose we survey 10 subjects per school, in 12 different schools; with 6 schools belonging to one city and 6 other schools coming from another city. There are three factors involved in

this experiment: the subjects (call it $\mathcal{S}$), the school (call it $\mathcal{A}$), and the cities (call it $\mathcal{B}$). The subjects are nested in the schools which are nested in the cities. Therefore the formula for this design is $\mathcal{S}(\mathcal{A}(\mathcal{B}))$.

Note, incidentally, that the nesting relation is *transitive*. This means that when a factor is nested in a second factor which is itself nested in a third factor, then the first factor is also nested in the third. For example, the subject factor is nested in the school (a subject is in one and only one school). As the schools are nested in the cities (a school is in one and only one city), the subjects are automatically nested in the cities (because a subject is in one and only one city).

22.3 Finding the sources of variation

For each source of variation, we have a name which is called a *variation term*. These terms come in two types: the *elementary factors* also called elementary or simple terms (they correspond to main effects), and the *interaction terms*, also called composed terms. These terms are defined later on in Sections 22.3.2 and 22.3.3.

There are several algorithms for deriving the sources of variation from the formula of an experimental design (Crump, 1946; Lee, 1966, 1975; Millman and Glass, 1967; and Honeck *et al.* 1983). The present algorithm combines features of previous ones and is (hopefully) rather straightforward.

Each source of variation is represented by a name which is obtained from a three-step algorithm described below.

22.3.1 Sources of variation. Step 1: write down the formula

The first thing is to write down the formula of the design using letters to represent factors, and using the symbols $\times$ and $()$ to represent the crossing and nesting relations. The letter $\mathcal{E}$ should not be used (in order to avoid confusion with ‘error’). The letter $\mathcal{S}$ is reserved for the subject factor.

In this section, we will use the following design as an example:²

$$\mathcal{S}(\mathcal{A} \times \mathcal{B}(\mathcal{C})) \times \mathcal{D}(\mathcal{F}). \quad (22.4)$$

22.3.2 Sources of variation. Step 2: elementary factors

Start by finding the *elementary factors*. As the name indicates, the elementary factors are the building blocks of the design. They are the factors that can be named with one word. For example, the ‘Subject factor’, the ‘School factor’, and the ‘Typicality factor’. When a factor is not nested in any other factors, it is written using one letter only (e.g. factor $\mathcal{S}$, factor $\mathcal{A}$). When a factor is nested into one or more factors, it is always written accompanied by its nesting factor(s) in parentheses. So, for example, if $\mathcal{S}$ is nested in $\mathcal{A}$ the elementary factor will be written $\mathcal{S}(\mathcal{A})$. Only one set of parentheses is used to write down the nesting factors.

² It is not as scary as it looks. Most of the published papers nowadays use designs of that level of complexity.

So for example, in an $\mathcal{S}(\mathcal{A} \times \mathcal{B})$ as well as in $\mathcal{S}(\mathcal{A}(\mathcal{B}))$ design, the elementary subject factor is written as $\mathcal{S}(\mathcal{AB})$.

For our example of an $\mathcal{S}(\mathcal{A} \times \mathcal{B}(\mathcal{C})) \times \mathcal{D}(\mathcal{F})$ design, the elementary factors are (listed in alphabetical order):

$$\mathcal{A}, \mathcal{B}(\mathcal{C}), \mathcal{C}, \mathcal{D}(\mathcal{F}), \mathcal{F}, \mathcal{S}(\mathcal{ABC}). \quad (22.5)$$

22.3.3 Sources of variation. Step 3: interaction terms

Having listed the elementary factors, the next step is to combine them in order to generate the *interaction terms*. An *interaction term* (also called *interaction factor*, or *complex factor*) is obtained by crossing two or more elementary factors together. The easier way to obtain the interaction terms is to use a stepwise approach. There are as many steps as there are $\times$ signs in the design formula. For convenience, we denoted this number by K which gives also the *order* of the design.

Step 1. *Derive the first-order interaction terms.* Procedure: Create all the pairwise combinations of elementary factors. This will create all the first-order interaction terms. Note, incidentally, that a *first-order* interaction term is made of *two* elementary factors. In general, a k th-order interaction involves $k + 1$ elementary factors.

The nesting factors should always stay in parentheses, and we use only one set of parentheses. The crossing relation is obtained by juxtaposition of the letters from the elementary factors. When the elementary factors are simple letters (i.e. they are not nested), the first-order interaction term is made by the juxtaposition of the letters denoting these two factors. So, for example, the first interaction term corresponding to the interaction of the elementary factors $\mathcal{A}$ and $\mathcal{B}$ is $\mathcal{AB}$.

When an interaction term is made of one or two nested elementary factors, the nested factors (i.e. the factors *not* in parentheses) and the nesting factors (i.e. the factors *in* parentheses) are crossed separately. So, for example, the first-order interaction term is obtained by crossing the elementary factors $\mathcal{B}(\mathcal{C})$ and $\mathcal{D}(\mathcal{F})$ is $\mathcal{BD}(\mathcal{CF})$. For convenience, we order the factors alphabetically. Obviously, we order the nested factors and the nesting factors separately.

Having created all the first-order interaction terms, we need to edit them or more precisely to eliminate all the ‘non-grammatical’ terms. This is done by using the following *crossing condition*.

Crossing condition: Eliminate an interaction term when a factor (i.e. a letter) is repeated in this term. This condition applies for nested and nesting factors. So, any term with a repeated letter is eliminated. For example, the following terms should be eliminated: $\mathcal{AS}(\mathcal{ABS})$, $\mathcal{BS}(\mathcal{ABCC})$, $\mathcal{AAB}$.

Step 2. *Derive the second-order interaction terms.* Procedure: Cross the first-order interaction terms (from Step 1) with the elementary factors. Eliminate the ‘non-grammatical’ terms using the *crossing condition* (see Step 1). In addition, an interaction term cannot be present twice (e.g. $\mathcal{ABC}$ and $\mathcal{ABC}$ are the same interaction term, and so we need to keep only one of them).

⋮

And so on, up to step k :

Step k . Derive the k th-order interaction terms. Procedure: Cross the $(k - 1)$ th-order interaction terms (from Step $k - 1$) with the *first*-order interaction terms. Eliminate the ‘non-grammatical’ terms using the *crossing condition* (see Step 1). Keep only one interaction term when a term has been created several times.

⋮

And so on, up to step k (remember k is the number of $\times$ signs in the design formula).

For our example of an $\mathcal{S}(\mathcal{A} \times \mathcal{B}(\mathcal{C})) \times \mathcal{D}(\mathcal{F})$ design, we went through the following steps:

Step 1. Step 1 gives the first-order interaction terms:

$$\begin{aligned} &\mathcal{AB}(\mathcal{C}), \mathcal{AC}, \mathcal{AD}(\mathcal{F}), \mathcal{AF}, \mathcal{BD}(\mathcal{CF}), \\ &\mathcal{BF}(\mathcal{C}), \mathcal{CD}(\mathcal{F}), \mathcal{CF}, \mathcal{DS}(\mathcal{ABCF}), \mathcal{FS}(\mathcal{ABC}). \end{aligned} \quad (22.6)$$

Step 2. (And last step, because there are 2 $\times$ signs in the formula of the design.) We obtain the following second-order interaction terms:

$$\mathcal{ABD}(\mathcal{CF}), \mathcal{ABF}(\mathcal{C}), \mathcal{ACD}(\mathcal{F}), \mathcal{ACF}. \quad (22.7)$$

22.4 Writing the score model

The score model is written from the sources of variation. The main idea is to have one component of the score model for each source of variation plus one term for the grand mean and, in some cases, an additional term for the error. We start by writing down the indices and then the terms corresponding to the indices. The notation for a source will vary whether the elementary factors are fixed or random.

Let us start with the indices. Using the example of an $\mathcal{S}(\mathcal{A} \times \mathcal{B}(\mathcal{C})) \times \mathcal{D}(\mathcal{F})$ design, we have the following sources (starting from the elementary factors) ordered by level of complexity from the elementary factors to the second level interaction term (and alphabetically within each level of complexity):

$\mathcal{A}$	$\mathcal{B}(\mathcal{C})$	$\mathcal{C}$	$\mathcal{D}(\mathcal{F})$	$\mathcal{F}$
$\mathcal{S}(\mathcal{ABC})$	$\mathcal{AB}(\mathcal{C})$	$\mathcal{AC}$	$\mathcal{AD}(\mathcal{F})$	$\mathcal{AF}$
$\mathcal{BD}(\mathcal{CF})$	$\mathcal{BF}(\mathcal{C})$	$\mathcal{CD}(\mathcal{F})$	$\mathcal{CF}$	$\mathcal{DS}(\mathcal{ABCF})$
$\mathcal{FS}(\mathcal{ABC})$	$\mathcal{ABD}(\mathcal{CF})$	$\mathcal{ABF}(\mathcal{C})$	$\mathcal{ACD}(\mathcal{F})$	$\mathcal{ACF}$

(22.8)

The indices are obtained from the names of the sources of variation by using the same letters but in lower case roman instead of the script upper case letters used for the names of the sources. For our example, we obtain the following pattern of indices.

a	$b(c)$	c	$d(f)$	f
$s(abc)$	$ab(c)$	ac	$ad(f)$	af
$bd(cf)$	$bf(c)$	$cd(f)$	cf	$ds(abcf)$
$fs(abc)$	$abd(cf)$	$abf(c)$	$acd(f)$	acf

(22.9)

In order to name the sources, we use the letters that are not in parentheses. When a source is fixed, we replace this term by the corresponding Greek letter. When a term is random, we use a Roman letter.

For our example, suppose that $\mathcal{A}$ and $\mathcal{B}$ are fixed and that all other sources are random. We obtain the following terms printed with their indices:

$$\begin{array}{ccccc}
 \alpha_a & \beta_{b(c)} & c_c & d_{d(f)} & f_f \\
 s_{s(abc)} & \alpha\beta_{ab(c)} & \alpha c_{ac} & \alpha d_{ad(f)} & \alpha f_{af} \\
 \beta d_{bd(cf)} & \beta f_{bf(c)} & c d_{cd(f)} & c f_{cf} & d s_{ds(abcf)} \\
 f s_{fs(abc)} & \alpha\beta d_{abd(cf)} & \alpha\beta f_{abf(c)} & \alpha c d_{acd(f)} & \alpha c f_{acf}
 \end{array} \quad (22.10)$$

The score model should always start with the population mean (i.e. μ). When the design is not a completely independent measurements design (i.e. if at least one factor is crossed with Factor S), we need to add a term corresponding to the error. This term is denoted by the letter e . Its index is the same as the index of the term with all the letters of the design. In our example, this all-letter-term is $\mathcal{DS}(\mathcal{ABC}\mathcal{F})$. Therefore, the error term will be $e_{ds(abcf)}$. In order to obtain the score model, we need to add all the terms together and put them into an equation showing that these terms are the components of a basic score in the design. For our example of an $\mathcal{S}(\mathcal{A} \times \mathcal{B}(\mathcal{C})) \times \mathcal{D}(\mathcal{F})$ design, we obtain the following score model:

$$\begin{array}{l}
 Y_{abcdfs} = \mu \quad +\alpha_a \quad +\beta_{b(c)} \quad +c_c \quad +d_{d(f)} \quad +f_f \\
 \quad +s_{s(abc)} \quad +\alpha\beta_{ab(c)} \quad +\alpha c_{ac} \quad +\alpha d_{ad(f)} \quad +\alpha f_{af} \\
 \quad +\beta d_{bd(cf)} \quad +\beta f_{bf(c)} \quad +c d_{cd(f)} \quad +c f_{cf} \quad +d s_{ds(abcf)} \\
 \quad +f s_{fs(abc)} \quad +\alpha\beta d_{abd(cf)} \quad +\alpha\beta f_{abf(c)} \quad +\alpha c d_{acd(f)} \quad +\alpha c f_{acf} \\
 \quad +e_{ds(abcf)}
 \end{array} \quad (22.11)$$

(to make reading easier, we have omitted the subscript ‘dots’ for μ in this equation).

22.5 Degrees of freedom and sums of squares

The degrees of freedom, the comprehension formulas for the sums of squares, and their computational formulas can be obtained from the name of the sources of variation. We start with the degrees of freedom, because they are themselves used to derive the formulas for the sums of squares.

22.5.1 Degrees of freedom

The numbers of degrees of freedom are obtained from the name of the source by following two simple substitution rules.

1. Each letter *not* in parentheses is replaced by its name minus 1. So, for example, the term $\mathcal{AB}$ will give the following number of degrees of freedom: $(A - 1)(B - 1)$.
2. Each letter *in* parentheses is left unchanged (and so we can remove its parentheses). So for example, $\mathcal{B}(\mathcal{C})$ becomes $C(B - 1)$; $\mathcal{S}(\mathcal{ABC})$, becomes $ABC(S - 1)$.

22.5.2 Sums of squares

The comprehension and computational formulas for the sums of squares are obtained by developing the formulas of the degrees of freedom into what is called ‘*developed degrees*’

of freedom'.³ Let us take the example of the source $S(ABC)$ from the $S(\mathcal{A} \times \mathcal{B}(\mathcal{C})) \times \mathcal{D}(\mathcal{F})$ design used in this example. The number of degrees of freedom of this source is $ABC(S-1)$. When developed, the number of degrees of freedom becomes: $ABCS - ABC$.

22.5.2.1 Comprehension formulas

The developed degrees of freedom give the name of the active indices in the comprehension formula of the sums of squares. These indices correspond to a basic score (i.e. the letter Y) when all the letters of the name of the design are present in the indices. These indices also correspond to a mean (i.e. the letter M with some dots) when at least one letter of the name of the design is missing in the indices.

For the example of the $S(\mathcal{A} \times \mathcal{B}(\mathcal{C})) \times \mathcal{D}(\mathcal{F})$ design, the formula for the sums of squares of the source $S(ABC)$ is obtained from its developed degrees of freedom (equal to $ABCS - ABC$) as

$$SS_{S(ABC)} = \sum_{abdfs} (M_{abc..s} - M_{abc...})^2 = DF \sum_{abcs} (M_{abc..s} - M_{abc...})^2. \quad (22.12)$$

The pattern of indices corresponding to the digit 1 is a set of dots. So for example, the source $\mathcal{AD}$ of our design has $(A-1)(D-1) = AD - A - D + 1$ degrees of freedom. The comprehension formula of its sums of squares is

$$\begin{aligned} SS_{AD} &= \sum_{abdfs} (M_{a..d..} - M_{a.....} - M_{...d..} + M_{.....})^2 \\ &= BCFS \sum_{ad} (M_{a..d..} - M_{a.....} - M_{...d..} + M_{.....})^2. \end{aligned} \quad (22.13)$$

22.5.2.2 Computational formulas

The developed degrees of freedom also give the computational formulas for the sums of squares. The procedure is to replace each set of letters (assuming that the number '1' in these formulas is a letter) by the corresponding 'numbers in the square'.

For the example of the $S(\mathcal{A} \times \mathcal{B}(\mathcal{C})) \times \mathcal{D}(\mathcal{F})$ design, the formula for the sums of squares of the source $S(ABC)$ is obtained from its developed degrees of freedom (equal to $ABCS - ABC$) as:

$$\begin{aligned} SS_{S(ABC)} &= \sum_{abdfs} (M_{abc..s} - M_{abc...})^2 \\ &= DF \sum_{abcs} (M_{abc..s} - M_{abc...})^2 \\ &= \boxed{ABCS} - \boxed{ABC}. \end{aligned} \quad (22.14)$$

The computation of the 'numbers in the square' is detailed in the following section.

22.5.2.3 Computing in a square

The 'numbers in the squares' are simply a notational shortcut. They are obtained from a four-step process, which is better understood with an example. Suppose that we want to develop the

³ How original!

following ‘number in a square’: $ABCS$ from our $S(\mathcal{A} \times \mathcal{B}(\mathcal{C})) \times \mathcal{D}(\mathcal{F})$ design. We proceed as follows.

1. Write down the expression Y^2 .
2. Write down as *active* indices of Y^2 all the letters present in the square. Replace the missing letters by dots. If ‘1’ is in the square, there is no letter and therefore no active index, and so all the indices are replaced by dots. For our example, we have the following letters ‘ $abcs$ ’, which gives the following pattern of indices: $abc..s$. Therefore, we obtain the following term: $Y_{abc..s}^2$.
3. If there are inactive indices (i.e. if there are dots) put the sum sign before the Y^2 term. For our example we get: $\sum Y_{abc..s}^2$.
4. If there are inactive indices (i.e. if there are dots) divide the $\sum Y$ term by the letters corresponding to all the dots. In our example we get $\frac{\sum Y_{abc..s}^2}{DF}$.

So to recap, the ‘number in the square’ is developed as

$$\boxed{ABCS} = \frac{\sum Y_{abc..s}^2}{DF} = \frac{1}{DF} \sum Y_{abc..s}^2. \quad (22.15)$$

Some additional examples, taken from the usual $S(\mathcal{A} \times \mathcal{B}(\mathcal{C})) \times \mathcal{D}(\mathcal{F})$ design, follow:

$$\boxed{ABCD} = \frac{\sum Y_{abcd..}^2}{FS} = \frac{1}{FS} \sum Y_{abcd..}^2$$

$$\boxed{ABC} = \frac{\sum Y_{abc...}^2}{DFS} = \frac{1}{DFS} \sum Y_{abc...}^2$$

$$\boxed{1} = \frac{Y_{.....}^2}{ABCDFS}$$

$$\boxed{ABCDEF} = \sum Y_{abcdefs}^2 \quad (22.16)$$

22.6 Example

Let us take the example of an $S(\mathcal{A}) \times \mathcal{B}(\mathcal{C})$ design with $\mathcal{A}$, $\mathcal{B}(\mathcal{C})$, $\mathcal{C}$ and $S(\mathcal{A})$ being random factors.

We obtain the following sources of variation: $\mathcal{A}$, $\mathcal{C}$, $\mathcal{AC}$, $\mathcal{B}(\mathcal{C})$, $S(\mathcal{A})$, $\mathcal{AB}(\mathcal{C})$, $\mathcal{CS}(\mathcal{A})$, and $\mathcal{BS}(\mathcal{AC})$. This gives the following score model:

$$Y_{abcs} = \mu... + a_a + c_c + ac_{ac} + b_{b(c)} + s_{s(a)} + ab_{ab(c)} + cs_{cs(a)} + bs_{bs(ac)} + e_{bs(ac)}. \quad (22.17)$$

To illustrate the comprehension formula, let us take the source $\mathcal{AB}(\mathcal{C})$. We get the following degrees of freedom:

$$(A-1)(B-1)C = ABC - AC - BC + C. \quad (22.18)$$

These degrees of freedom give, in turn, the comprehension formula for the sum of squares of $\mathcal{AB}(\mathcal{C})$

$$\begin{aligned} SS_{AB(C)} &= \sum_{abcs} (M_{abc.} - M_{a.c.} - M_{.bc.} + M_{..c.})^2 \\ &= S \sum_{abc} (M_{abc.} - M_{a.c.} - M_{.bc.} + M_{..c.})^2. \end{aligned} \quad (22.19)$$

22.7 Expected values

In order to perform the statistical analysis, we need to have the expected values of the mean squares corresponding to each source of variation of the design.

These expected values are used to find the test mean squares (either directly or by combining several of them as for a quasi- F). They are also needed to compute some of the coefficients of intensity of effect (e.g. ω^2 , ρ^2).

Appendix F on expected values gives the detail of the principles and rules used to derive the expected values of the mean squares. These methods can be used to derive the expected values for the mean squares corresponding to all designs covered in this book. However, even for simple designs, that approach is rather long and tedious. For complex designs, this approach (even though always correct) can become taxing (even for the most simple minded statistician). As statisticians are no more masochistic⁴ than most people, they have developed algorithms to simplify and automatize the process of finding expected values of mean squares. These rules depend heavily upon the assumptions of the score model (such as $\sum \alpha_a = 0$, and homoscedasticity). The rules detailed here and adapted to our notation, are presented in several sources (cf. among others, Bennet and Franklin, 1954; Schultz, 1955; Cornfield and Tukey, 1956; Lee, 1975—our personal favorite; Honeck *et al.* 1983).

Our algorithm uses four steps. We illustrate it with the example of an $S(\mathcal{A}) \times \mathcal{B}$ design with $\mathcal{A}$ and $\mathcal{S}$ being random factors. As an exercise, we can check that we have five sources of variation for this design,

$$\mathcal{A}, \mathcal{B}, S(\mathcal{A}), \mathcal{AB}, BS(\mathcal{A}).$$

(22.20)

The score model for this design is

$$Y_{abs} = \mu + a_a + \beta_b + s_{s(a)} + a\beta_{ab} + \beta s_{bs(a)} + e_{bs(a)}.$$

(22.21)

22.7.1 Expected value: Step 1

The first step is to build a square table with as many rows and columns as the design has sources of variation. Label the rows and the columns with the name of the sources. Leave the first row and first column empty.

So, we obtain the following table:

	$BS(\mathcal{A})$	$\mathcal{AB}$	$S(\mathcal{A})$	$\mathcal{B}$	$\mathcal{A}$
$\mathcal{A}$					
$\mathcal{B}$					
$S(\mathcal{A})$					
$\mathcal{AB}$					
$BS(\mathcal{A})$					

⁴ In fact, we have been told that it is standard folklore to refer to statistics classes as *sadistics* classes!

22.7.2 Expected value: Step 2

Fill in the first row of the table with variance terms (i.e. σ^2 or ϑ^2). This is done according to the following rules:

Rule 1. If all the factors outside the parentheses are fixed then the source of variation is fixed. The variance term is ϑ^2 for a fixed source. In all other cases, the source is random (i.e. a source is random if at least one of the factors outside the parentheses is random). The variance term for a random source is σ^2 .

Rule 2. Write down the name of the source as a subscript of the variance term. Use lower case roman letters for the subscript. Multiply the variance term by all the upper case letters that are not in the name of this source.

Rule 3. Take care of the heading of the first column of the table. If the design is a completely random design (i.e. if the factor S is nested in all other factors), leave this column empty and write down the variance term corresponding to the factor S as σ_e^2 . In all other cases, put the term σ_e^2 as the heading of the first column.

With these rules, we obtain the following table:

	$BS(\mathcal{A})$	$\mathcal{AB}$	$S(\mathcal{A})$	$\mathcal{B}$	$\mathcal{A}$	
	σ_e^2	$\sigma_{bs(a)}^2$	$S\sigma_{ab}^2$	$B\sigma_{s(a)}^2$	$AS\vartheta_b^2$	$BS\sigma_a^2$
$\mathcal{A}$						
$\mathcal{B}$						
$S(\mathcal{A})$						
$\mathcal{AB}$						
$BS(\mathcal{A})$						

22.7.3 Expected value: Step 3

Fill in the table with 1s according to the following rules:

Rule 1. Put a 1 in all the rows of the first column.

For the other columns, use Rule 2 below.

Rule 2. Put a 1 if the name of the row source is included in the name of the column source and if *all* the remaining letters are either random factors or are within parentheses. As a particular case, put 1 if the row name is the same as the column name. In other words, *do not* put a 1 if at least one of the remaining factors outside the parentheses is fixed.

For example, consider the cell at the intersection of row B and column $BS(A)$. The letter B is included in the name $BS(A)$. The remaining letters are S and A . These two letters correspond to factors that are either random or within parentheses (or both for A). Therefore, we put a 1 at the intersection of row B and column $BS(A)$.

Another example, consider the cell at the intersection of row A and column $BS(A)$. The remaining letters are S and B . At least one of the letters outside the parentheses (i.e. B) corresponds to a fixed factor. Therefore, we *do not* put a 1 at the intersection of row A and column $BS(A)$.

Using these two rules, we obtain the following table.

	$BS(\mathcal{A})$		$\mathcal{AB}$	$S(\mathcal{A})$	$\mathcal{B}$	$\mathcal{A}$
	σ_e^2	$\sigma_{bs(a)}^2$	$S\sigma_{ab}^2$	$B\sigma_{s(a)}^2$	$AS\vartheta_b^2$	$BS\sigma_a^2$
$\mathcal{A}$	1			1		1
$\mathcal{B}$	1	1	1		1	1
$S(\mathcal{A})$	1			1		
$\mathcal{AB}$	1	1	1			
$BS(\mathcal{A})$	1	1				

22.7.4 Expected value: Step 4

Find the expected values. For each row, the expected value of the mean square is composed of all the terms having a value of 1 in this row. The final result is displayed in Table 22.1

From Table 22.1, we can find that the experimental sources of variation $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{AB}$ can be tested using F ratios. Specifically, the sources are tested with

$$F_A = \frac{MS_A}{MS_{S(A)}}, \quad F_B = \frac{MS_B}{MS_{AB}}, \quad F_{AB} = \frac{MS_{AB}}{MS_{BS(A)}}.$$

(22.22)

22.8 Two additional exercises

As additional practice, we will compute the expected values of the mean squares for an $S(\mathcal{A} \times \mathcal{B}(\mathcal{C}))$ design for two different patterns of choices for the fixed and random factors. The sources of variation for an $S(\mathcal{A} \times \mathcal{B}(\mathcal{C}))$ design are:

$$\mathcal{A}, \mathcal{B}(\mathcal{C}), \mathcal{C}, S(\mathcal{ABC}), \mathcal{AB}(\mathcal{C}), \mathcal{AC}.$$

(22.23)

	$BS(\mathcal{A})$		$\mathcal{AB}$	$S(\mathcal{A})$	$\mathcal{B}$	$\mathcal{A}$	
	σ_e^2	$\sigma_{bs(a)}^2$	$S\sigma_{ab}^2$	$B\sigma_{s(a)}^2$	$AS\vartheta_b^2$	$BS\sigma_a^2$	$\mathbf{E}\{MS\}$
$\mathcal{A}$	1			1		1	$\sigma_e^2 + B\sigma_{s(a)}^2 + BS\sigma_a^2$
$\mathcal{B}$	1	1	1		1	1	$\sigma_e^2 + \sigma_{bs(a)}^2 + S\sigma_{ab}^2 + AS\vartheta_b^2$
$S(\mathcal{A})$	1			1			$\sigma_e^2 + B\sigma_{s(a)}^2$
$\mathcal{AB}$	1	1	1				$\sigma_e^2 + \sigma_{bs(a)}^2 + S\sigma_{ab}^2$
$BS(\mathcal{A})$	1	1					$\sigma_e^2 + \sigma_{bs(a)}^2$

Table 22.1 Final stage of the table used to derive the expected values of the mean squares for an $S(\mathcal{A}) \times \mathcal{B}$ design with $\mathcal{A}$, S being random factors, and $\mathcal{B}$ being a fixed factor.

22.8.1 $\mathcal{S}(\mathcal{A} \times \mathcal{B}(C)) : \mathcal{A}$ and $\mathcal{B}$ fixed, $\mathcal{C}$ and $\mathcal{S}$ random

The score model is

$$Y_{abcs} = \mu + \alpha_a + \beta_{b(c)} + c_c + \alpha\beta_{ab(c)} + \alpha c_{ac} + e_{s(abc)}. \quad (22.24)$$

The results are given in Table 22.2.

22.8.2 $\mathcal{S}(\mathcal{A} \times \mathcal{B}(C)) : \mathcal{A}$ fixed, $\mathcal{B}, \mathcal{C}$, and $\mathcal{S}$ random

The score model is

$$Y_{abcs} = \mu + \alpha_a + b_{b(c)} + c_c + \alpha b_{ab(c)} + \alpha c_{ac} + e_{s(abc)}. \quad (22.25)$$

The results are given in Table 22.3.

	$\mathcal{S}(\mathcal{ABC})$	$\mathcal{AC}$	$\mathcal{AB}(C)$	$\mathcal{C}$	$\mathcal{B}(C)$	$\mathcal{A}$	
	σ_e^2	$BS\sigma_{ac}^2$	$S\vartheta_{ab(c)}^2$	$ABS\sigma_c^2$	$AS\vartheta_{b(c)}^2$	$BCS\vartheta_a^2$	$\mathbf{E}\{MS\}$
$\mathcal{A}$	1	1				1	$\sigma_e^2 + BS\sigma_{ac}^2 + BCS\vartheta_a^2$
$\mathcal{B}(C)$	1				1		$\sigma_e^2 + AS\vartheta_{b(c)}^2$
$\mathcal{C}$	1			1			$\sigma_e^2 + ABS\sigma_c^2$
$\mathcal{AB}(C)$	1		1				$\sigma_e^2 + S\vartheta_{ab(c)}^2$
$\mathcal{AC}$	1	1					$\sigma_e^2 + BS\sigma_{ac}^2$
$\mathcal{S}(\mathcal{ABC})$	1						σ_e^2

Table 22.2 Final stage of the table used to derive the expected values of the mean squares for an $\mathcal{S}(\mathcal{A} \times \mathcal{B}(C))$ design with $\mathcal{A}, \mathcal{B}$ being fixed factors, and $\mathcal{C}$ and $\mathcal{S}$ being random factors.

	$\mathcal{S}(\mathcal{ABC})$	$\mathcal{AC}$	$\mathcal{AB}(C)$	$\mathcal{C}$	$\mathcal{B}(C)$	$\mathcal{A}$	
	σ_e^2	$BS\sigma_{ac}^2$	$S\vartheta_{ab(c)}^2$	$ABS\sigma_c^2$	$AS\vartheta_{b(c)}^2$	$BCS\vartheta_a^2$	$\mathbf{E}\{MS\}$
$\mathcal{A}$	1	1	1			1	$\sigma_e^2 + BS\sigma_{ac}^2 + S\vartheta_{ab(c)}^2 + BCS\vartheta_a^2$
$\mathcal{B}(C)$	1				1		$\sigma_e^2 + AS\vartheta_{b(c)}^2$
$\mathcal{C}$	1			1	1		$\sigma_e^2 + AS\vartheta_{b(c)}^2 + ABS\sigma_c^2$
$\mathcal{AB}(C)$	1		1				$\sigma_e^2 + S\vartheta_{ab(c)}^2$
$\mathcal{AC}$	1	1	1				$\sigma_e^2 + S\vartheta_{ab(c)}^2 + BS\sigma_{ac}^2$
$\mathcal{S}(\mathcal{ABC})$	1						σ_e^2

Table 22.3 Final stage of the table used to derive the expected values of the mean squares for an $\mathcal{S}(\mathcal{A} \times \mathcal{B}(C))$ design with $\mathcal{A}$ being a fixed factor, and $\mathcal{B}, \mathcal{C}$ and $\mathcal{S}$ being random factors.

Appendices

A	Descriptive statistics	411
B	The sum sign: Σ	424
C	Elementary probability: a refresher	426
D	Probability distributions	443
E	The binomial test	470
F	Expected values	485

A

Descriptive statistics

A.1 Introduction

Psychologists and other researchers often want to organize and summarize the data that they have collected. Reformatting and summarizing data can provide insights, and this is also useful for communicating the results of research to others. The branch of statistics that provides ways to describe the characteristics of a set of data is called descriptive statistics and in this appendix, we review basic descriptive statistics.

A.2 Histogram

We will use a simple example to illustrate some of the basic techniques of descriptive statistics. Suppose we have a set of 20 scores on a final examination that has a possible high score of 100. One way of presenting this information is to simply provide a list of all the scores as in Table A.1

However, there are other ways, which may be more meaningful, of describing these *raw data*. One of these ways is a frequency table (as shown in Table A.2) which lists categories of scores along with their corresponding frequencies. To construct a frequency table, the scores are first divided into categories or classes. Usually, the range of the scores included in each class is the same, although some rounding may be necessary. Each score belongs to exactly one class, and there is a class for every score. One way to construct a frequency table for our sample examination data is to decide on 5 classes, each with a width of 10.

The visual device of graphing is also often used to describe data. A type of bar graph called a *histogram* is used to illustrate a frequency table. In a histogram, the vertical scale (*Y* axis) delineates frequencies, and the horizontal scale (*X* axis) shows the values of the data being represented. The data values may be given as boundary points or midpoints of each class. Each bar of the histogram represents a class of the frequency table. The frequency distribution of our sample data depicted in a histogram would appear as shown in Figure A.1.

In this histogram, we can see, for example, that two observations fall in the 50–59 range, three in the 60–69 range, seven in the 70–79 range, five in the 80–89 range, and three in the 90–100 range.

An alternative graphical way to represent the frequency distribution of data is a frequency polygon. In this type of graph the axes are like those in a histogram, but the frequency of

50	55	63	65	68	70	72	73	75	77	78	79	83	85	86	87	89	93	96	98

Table A.1 A data set

Class	Frequency
50–59	2
60–69	3
70–79	7
80–89	5
90–100	3

Table A.2 A table from the data of Table A.1

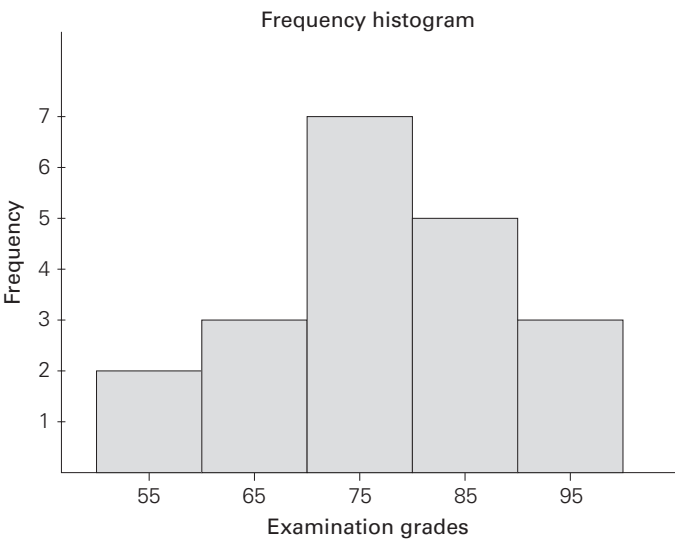


Figure A.1 Histogram from data in Table A.2 (20 exam grades).

scores in each category is drawn as a single dot. The dots are then connected with straight lines. Dots representing zero frequency are placed one interval below and one interval above the observed data. The data from our example can be graphed as a frequency polygon as shown in Figure A.2.

Distribution information may also be presented by giving the proportion of scores in each category. The proportion is calculated by dividing the number of scores in a category by the total number of scores. This manner of describing data is useful when we wish to compare the scores from two or more distributions that do not have an equal number of scores. The data in our example can be summarized in the proportion table shown in Table A.3.

This proportional information can be graphically displayed in a *proportion histogram*. In this type of histogram the vertical axis provides a scale for proportion. The data in our example can be shown as a proportion histogram and a proportion polygon in Figure A.3.

Before we continue this overview of descriptive statistics, it is helpful to introduce some notation that will be used in our calculations.

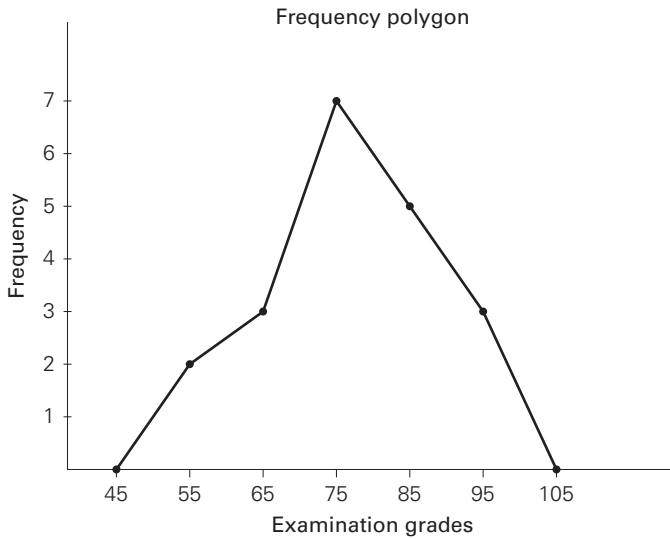


Figure A.2 Frequency polygon from data in Table A.2 (20 exam grades).

Class	Proportion
50–59	.10
60–69	.15
70–79	.35
80–89	.25
90–100	.15

Table A.3 Proportion table (compare with Table A.2)

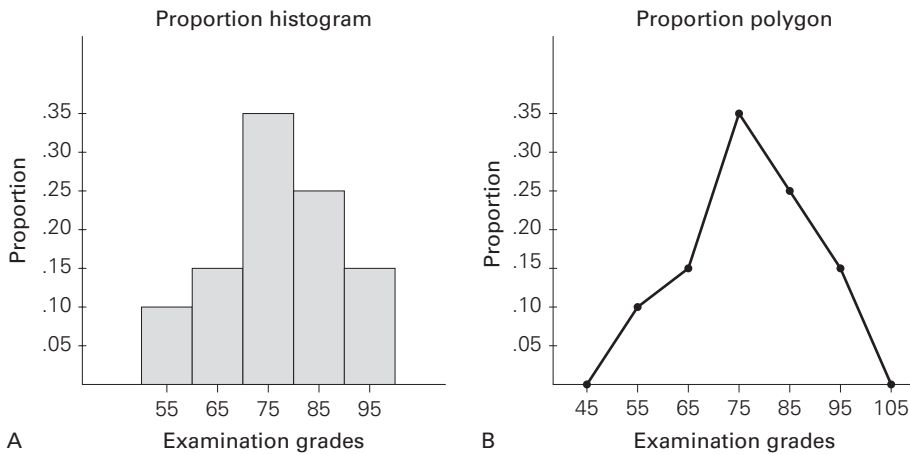


Figure A.3 Proportion histogram (A) and proportion polygon (B) [data from Table A.3].

A.3 Some formal notation

We use the following notations to describe data.

A.3.1 Notations for a score

The value of a score is Y (or Y_n , or Y_s). It will be our convention to use Y to represent a score (also called an observation) that the experimenter measures. As we shall see later on, we will also call Y the dependent variable. Because we often have more than one observation, we will distinguish between the observations by using a subscript on Y . For example, Y_1 will represent observation number one, Y_2 observation number two, and so on. To state this in a more general fashion, Y_n represents the n th score out of a total of N scores.

When the observations are scores attained by subjects, we use S to denote the number of subjects receiving a particular treatment and s as the subscript for Y . Therefore, the score of a particular subject is denoted Y_s . For example, the score of the 6th subject is expressed as Y_6 .

A.3.2 Subjects can be assigned to different groups

When we conduct an experiment, subjects are often assigned to different groups. One of the groups may receive a certain treatment, and the other group may serve as a control, receiving no treatment. It is also possible to provide different levels of the same treatment to different groups of subjects, for example a drug dose of 10 mg, 20 mg, and 30 mg. This treatment, which is manipulated by the experimenter, is also called an independent variable. The number of levels of the treatment is denoted A . In the example of the three levels of drug dosage $A = 3$.

A.3.3 The value of a score for subjects in multiple groups is $Y_{a,s}$

When subjects are assigned to different groups for particular levels of a treatment, we denote scores with a double subscript to indicate both the treatment group and the subject number. The general notation for a score value in this case is $Y_{a,s}$ where a denotes the level (experimental group) and s denotes the subject. For example the score of the third subject in the second treatment group is $Y_{2,3}$.

A.3.4 The summation sign is $\sum$

In mathematics, the function of summation (adding up) is symbolized using $\sum$, the Greek capital letter *sigma*. The sum of all the scores for S subjects is represented as:

$$\sum_{s=1}^S Y_s .$$

The expression $s = 1$ underneath the sigma indicates that the first value to be assigned to the subscript of Y is 1. The S above the sigma indicates that the last value to be assigned to the subscript of Y is the total number of subjects S . Values assigned to the subscript of Y are integers ranging from 1 up to and including S .

$$\sum_{s=1}^S Y_s = Y_1 + Y_2 + \cdots + Y_S .$$

If five subjects ($S = 5$) receive the following scores:

Subject	1	2	3	4	5
Score	80	85	70	60	92

the sum of their scores is:

$$\sum_{s=1}^S Y_s = Y_1 + Y_2 + Y_3 + Y_4 + Y_5 \quad (\text{A.1})$$

$$= 80 + 85 + 70 + 60 + 92$$

$$= 387. \quad (\text{A.2})$$

For subjects assigned to different levels of treatment $\mathcal{A}$, the sum of all scores is represented as

$$\sum_{a=1}^A \sum_{s=1}^S Y_{a,s}.$$

This notation indicates summation of the scores of all of the subjects in the first treatment group, plus the scores of all of the subjects in the second group, and so on, until a grand sum is found for all subjects in all groups.

$$\begin{aligned} \sum_{a=1}^A \sum_{s=1}^S Y_{a,s} &= Y_{1,1} + \cdots + Y_{1,S} \\ &\quad + Y_{2,1} + \cdots + Y_{2,S} + \cdots \\ &\quad + Y_{A,1} + \cdots + Y_{A,S}. \end{aligned} \quad (\text{A.3})$$

In the following example there are two levels of treatment ($A = 2$) and three subjects ($S = 3$) are assigned to each level. If the following scores are obtained:

Subject	1	2	3
Group 1 scores	65	85	80
Group 2 scores	80	85	92

the sum of the scores is calculated as

$$\begin{aligned} \sum_{a=1}^A \sum_{s=1}^S Y_{a,s} &= Y_{1,1} + Y_{1,2} + Y_{1,3} + Y_{2,1} + Y_{2,2} + Y_{2,3} \\ &= 65 + 85 + 80 + 80 + 85 + 92 \\ &= 487. \end{aligned} \quad (\text{A.4})$$

A.4 Measures of central tendency

Frequency tables and graphs are useful for conveying information about data, but there are also other, more quantitative, ways to describe data. These measures are very useful when one wants to compare several sets of data and make inferences. Measures of central tendency are

used to calculate a value that is central to a set of scores. There are several ways to arrive at a value that is central to a collection of raw data. The three most common are the *arithmetic mean*, the *median*, and the *mode*.

A.4.1 Mean

The mean corresponds to what most people call an *average*. It is calculated by summing all the scores and then dividing this sum by the number of scores. Continuing with our example of 20 examination scores, adding them up we find a value of 1,542. Because there are 20 scores we divide this sum by 20 to obtain the mean of 77.1.

The population mean of Y scores is denoted μ_Y , the Greek letter *mu*, which is read as ‘mew’. We will denote the mean of a sample of Y scores as M_Y . A sample is a subset (portion) of a population. An alternative notation is $\bar{Y}$, read ‘Y bar’. Using the summation notation the formula for calculating the mean of a sample is

$$M_Y = \bar{Y} = \frac{1}{N} \sum_{n=1}^N Y_n .$$

A.4.2 Median

The median of a set of scores is the value of the middle score when the scores are arranged in order of increasing magnitude. Given the set of scores [20, 45, 62, 85, 90], the median is 62. If the number of scores is even, then the median is the mean of the two scores in the middle. In the example given previously of 20 examination scores the median is the mean of 77 and 78, which is 77.5.

50 55 63 65 68 70 72 73 75 77 78 79 83 85 86 87 89 93 96 98

$$M_Y = \frac{1}{N} \sum_{n=1}^N Y_n = \frac{77 + 78}{2} = 77.5 .$$

When some of the scores have extreme values, the median can be a more meaningful measure of central tendency than the mean. Consider the following set of scores: 10, 80, 83, 88, 90. The mean of these scores is 70.2.

$$\begin{aligned} M_Y &= \frac{1}{N} \sum_{n=1}^N Y_n \\ &= \frac{10 + 80 + 83 + 88 + 90}{5} \\ &= \frac{351}{5} \\ &= 70.2 \end{aligned} \tag{A.5}$$

By contrast, the median is 83, which seems to be more representative of the scores.

A.4.3 Mode

The *mode* is the score that occurs the most frequently. In our first example no score occurs more than once. Therefore, there is no mode for this set of data. In the set of scores

$$50, 60, 75, 75, 80, 95,$$

the mode is 75 because it occurs twice and all the other scores occur only once.

A set of data is termed *bimodal* if there are two scores, not next to each other, that occur with the same greatest frequency. The set of scores:

$$50, 60, 60, 75, 80, 80, 90,$$

is bimodal because both 60 and 80 occur most frequently. If more than two scores occur with the greatest frequency the set of data is said to be *multimodal*.

A.4.4 Measures of central tendency recap

Although a measure of central tendency, such as a mean, provides an economical way to describe a set of data, it does not, by itself, always convey as much information as we would like. Consider the following three frequency distribution histograms (see Figure A.4), each based on ten scores. They all have the same mean, $M_Y = 75$, but they are obviously not alike in their shapes. It would be good practice for you to calculate the mean of each of these distributions at this point.

There are several ways to quantify the ‘shape’ of a distribution, or the way that the scores are spread around the mean. These ways are termed *measures of dispersion*.

A.5 Measures of dispersion

A.5.1 Range

One of the simplest ways to indicate the dispersion of a set of scores around their mean is to give the *range*. The range is the difference between the largest and the smallest scores:

$$Y_{\text{largest}} - Y_{\text{smallest}} .$$

In the example shown above, the range is the same for all three frequency distributions:

$$Y_{\text{largest}} - Y_{\text{smallest}} = 95 - 55 = 40 .$$

Therefore, in the case of this example, the range does not help us to distinguish between the three distributions.

A.5.2 Sum of squares

Another measure of dispersion (scatter) is called the *sum of squares* and is denoted as SS_Y for a set of Y scores. This is the sum of squared deviations (distances) from the mean, which is calculated as:

$$SS_Y = \sum_{n=1}^N (Y_n - M_Y)^2 .$$

This formula instructs us to find first the mean of the set of scores, M_Y , and then to subtract the mean from each of the Y scores to obtain a distance called a deviation from the mean.

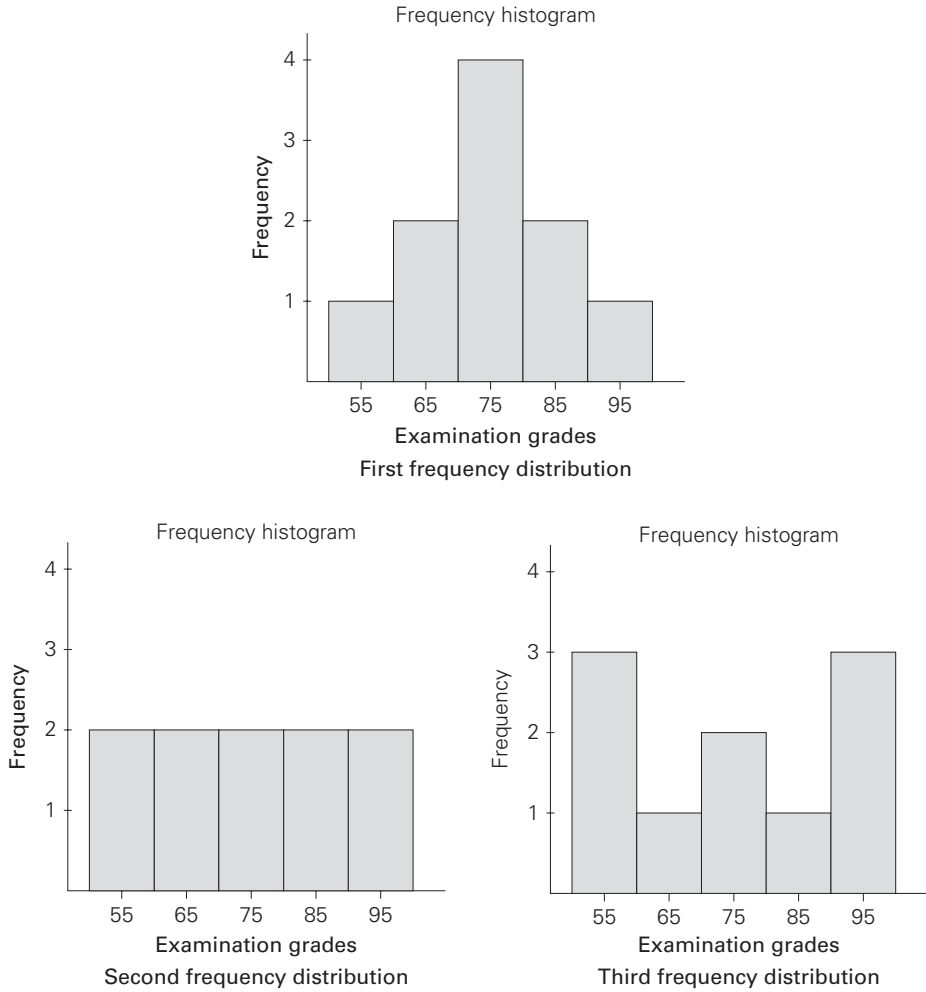


Figure A.4 Three different frequency distributions.

Each of these deviations is then squared, and then the squared deviations are summed. We can calculate the sum of squares for each of the three distributions in our example.

- For the first distribution:

$$\begin{aligned}
 SS_Y &= \sum_{n=1}^N (Y_n - M_Y)^2 \\
 &= (55 - 75)^2 + (65 - 75)^2 + (65 - 75)^2 + (75 - 75)^2 + (75 - 75)^2 \\
 &\quad + (75 - 75)^2 + (75 - 75)^2 + (85 - 75)^2 + (85 - 75)^2 + (95 - 75)^2 \\
 &= (-20)^2 + (-10)^2 + (-10)^2 \\
 &\quad + (0)^2 + (0)^2 + (0)^2 + (0)^2 + (10)^2 + (10)^2 + (20)^2 \\
 &= 400 + 100 + 100 + 0 + 0 + 0 + 0 + 100 + 100 + 400 \\
 &= 1,200.00 .
 \end{aligned}
 \tag{A.6}$$

- For the second distribution:

$$\begin{aligned}
 SS_Y &= \sum_{n=1}^N (Y_n - M_Y)^2 \\
 &= (55 - 75)^2 + (55 - 75)^2 + (65 - 75)^2 + (65 - 75)^2 + (75 - 75)^2 \\
 &\quad + (75 - 75)^2 + (85 - 75)^2 + (85 - 75)^2 + (95 - 75)^2 + (95 - 75)^2 \\
 &= (-20)^2 + (-20)^2 + (-10)^2 \\
 &\quad + (-10)^2 + (0)^2 + (0)^2 + (10)^2 + (10)^2 + (20)^2 + (20)^2 \\
 &= 400 + 400 + 100 + 100 + 0 + 0 + 100 + 100 + 400 + 400 \\
 &= 2,000.00 .
 \end{aligned} \tag{A.7}$$

- For the third distribution:

$$\begin{aligned}
 SS_Y &= \sum_{n=1}^N (Y_n - M_Y)^2 \\
 &= (55 - 75)^2 + (55 - 75)^2 + (55 - 75)^2 + (65 - 75)^2 + (75 - 75)^2 \\
 &\quad + (75 - 75)^2 + (85 - 75)^2 + (95 - 75)^2 + (95 - 75)^2 + (95 - 75)^2 \\
 &= (-20)^2 + (-20)^2 + (-20)^2 \\
 &\quad + (-10)^2 + (0)^2 + (0)^2 + (10)^2 + (20)^2 + (20)^2 + (20)^2 \\
 &= 400 + 400 + 400 + 100 + 0 + 0 + 100 + 400 + 400 + 400 \\
 &= 2,600.00 .
 \end{aligned} \tag{A.8}$$

As you can see the sum of squares measure of dispersion does distinguish between the three frequency distributions. The sum of squares grows larger as more scores are further from the mean.

A.5.3 Variance

A measure of dispersion that is often used in statistics is the *variance*. The variance of a population is an average of the squared deviations around the mean. To compute the variance of a population, the sum of squares is calculated as we have shown above, and then this sum is divided by the number of scores. Because the sum of squared deviations from the mean is divided by a value based on the number of scores in the sample, it is useful when comparing several samples of different sizes. The symbol used to indicate the variance of a population is σ^2 , a small Greek sigma that is squared. The population variance of a set of scores Y is denoted as σ_Y^2 , and measures the dispersion of scores around the population mean μ_Y . The formula for calculating the variance of a population is

$$\sigma_Y^2 = \frac{SS_Y}{N} = \frac{1}{N} \sum_{n=1}^N (Y_n - \mu_Y)^2 .$$

The symbol used to represent the variance of a sample is denoted $\hat{\sigma}^2$. The ‘hat’ over the sigma indicates that $\hat{\sigma}^2$ is an estimate of the population variance. The variance of a sample of Y scores is symbolized $\hat{\sigma}_Y^2$. To compute the estimate of the variance of a population from a

sample, the sum of squares is calculated as we have shown above, and then this sum is divided by the number of scores minus 1. The formula to calculate a sample variance is

$$\hat{\sigma}_Y^2 = \frac{SS_Y}{N-1} = \frac{1}{N-1} \sum_{n=1}^N (Y_n - M_Y)^2.$$

We can apply the above formula for sample variance to the three distributions in our example. We have already calculated the sum of squares for each distribution and can use those results in the variance formula.

- Variance of the first distribution:

$$\begin{aligned} \hat{\sigma}_Y^2 &= \frac{SS_Y}{N-1} = \frac{1}{N-1} \sum_{n=1}^N (Y_n - M_Y)^2 \\ &= \frac{1,200}{N-1} \\ &= \frac{1,200}{10-1} \\ &= \frac{1,200}{9} \\ &= 133.30. \end{aligned} \tag{A.9}$$

- Variance of the second distribution:

$$\begin{aligned} \hat{\sigma}_Y^2 &= \frac{SS_Y}{N-1} = \frac{1}{N-1} \sum_{n=1}^N (Y_n - M_Y)^2 \\ &= \frac{2,000}{N-1} \\ &= \frac{2,000}{10-1} \\ &= \frac{2,000}{9} \\ &= 222.20. \end{aligned} \tag{A.10}$$

- Variance of the third distribution:

$$\begin{aligned} \hat{\sigma}_Y^2 &= \frac{SS_Y}{N-1} = \frac{1}{N-1} \sum_{n=1}^N (Y_n - M_Y)^2 \\ &= \frac{2,600}{N-1} \\ &= \frac{2,600}{10-1} \\ &= \frac{2,600}{9} \\ &= 288.90. \end{aligned} \tag{A.11}$$

A.5.4 Standard deviation

The final measure of dispersion that we shall present is the *standard deviation*. The standard deviation is the square root of the variance. Therefore, it is measured in the same unit as the scores. (Recall that to obtain the sum of squares we squared all the distances from the mean. Taking the square root will ‘undo’ this squaring operation.) The standard deviation of a population is represented as σ . For a population of Y scores the standard deviation is σ_Y . The formula for the population standard deviation is

$$\sigma_Y = \sqrt{\frac{1}{N} \sum_{n=1}^N (Y_n - \mu_Y)^2}.$$

The standard deviation for a sample is symbolized $\hat{\sigma}$. For a set of Y scores, the standard deviation is indicated as $\hat{\sigma}_Y$. Again, the ‘hat’ over the sigma indicates that the sample standard deviation is an estimate of the population standard deviation. The formula to calculate the sample standard deviation is

$$\hat{\sigma}_Y = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (Y_n - M_Y)^2}.$$

Because we have already calculated the variance for each of the distributions in our example, we can simply take the square root of that variance to obtain the standard deviation.

- Standard deviation of the first distribution:

$$\begin{aligned} \hat{\sigma}_Y &= \sqrt{\frac{1}{N-1} \sum_{n=1}^N (Y_n - M_Y)^2} \\ &= \sqrt{133.3} \\ &\approx 11.50. \end{aligned} \tag{A.12}$$

- Standard deviation of the second distribution:

$$\begin{aligned} \hat{\sigma}_Y &= \sqrt{\frac{1}{N-1} \sum_{n=1}^N (Y_n - M_Y)^2} \\ &= \sqrt{222.2} \\ &\approx 14.90. \end{aligned} \tag{A.13}$$

- Standard deviation of the third distribution:

$$\begin{aligned} \hat{\sigma}_Y &= \sqrt{\frac{1}{N-1} \sum_{n=1}^N (Y_n - M_Y)^2} \\ &= \sqrt{288.9} \\ &\approx 17.00. \end{aligned} \tag{A.14}$$

We can use the descriptive statistics of mean and standard deviation to describe the three different distributions in our example. For the first distribution the mean is equal to 75.00, $M_Y = 75.00$, and the standard deviation is 11.50, $\hat{\sigma}_Y = 11.50$. The second distribution can be described as having $M_Y = 75.00$ and $\hat{\sigma}_Y = 14.90$, and the third distribution as having $M_Y = 75.00$ and $\hat{\sigma}_Y = 17.00$. Notice that when scores cluster near the mean the standard deviation is smaller than when scores are more distant from the mean.

A.6 Standardized scores alias Z-scores

The mean and the standard deviation are useful to compare scores that are contained in different distributions. Notice that a score of 85 on the examination in the above example can be interpreted differently, depending upon the distribution that contains the score. By examining the three histograms we can see that for the first distribution a score of 85 is lower than only one other score in the distribution, for the second distribution 85 is lower than two scores, and in the third distribution lower than three scores. Thus a score of 85 does not itself tell us about the ‘position’ of this grade in relation to the other scores on the examination.

Scores from different distributions, such as the ones in our example, can be *standardized* in order to provide a way of comparing them that includes consideration of their respective distributions. Prior to the comparison, scores are standardized by subtracting the mean and then dividing by the standard deviation of the distribution. The resulting standardized score is termed a Z-score. The formula for calculating a Z-score is

$$Z = \frac{Y - M_Y}{\hat{\sigma}_Y} .$$

We say that subtracting the mean *centers* the distribution, and that dividing by the standard deviation *normalizes* the distribution. We could have used, alternatively, the population standard deviation (σ) instead of the estimation of the standard deviation ($\hat{\sigma}$) in the previous formula. This choice is, in general, of no real importance as long as one compares only Z-scores computed with the same formula.

The interesting properties of the Z-scores are that they have a zero mean (effect of ‘centering’) and a variance and standard deviation of 1 (effect of ‘normalizing’). This is because all distributions expressed in Z-scores have the same mean (0) and the same variance (1) that we can use Z-scores to compare observations coming from different distributions.

The fact that Z-scores have a zero mean and a unitary variance can be shown by developing the formulas for the sum of Z-scores and for the sum of the squares of Z-scores. This is done in the following digression that you can skip on the first reading or if you are already convinced.

A.6.1 Z-scores have a mean of 0, and a variance of 1

In order to show that the mean of the Z-scores is 0, it suffices to show that the sum of the Z-scores is 0. This is shown by developing the formula for the sum of the Z-scores:

$$\begin{aligned} \sum Z &= \sum \frac{Y - M_Y}{\hat{\sigma}_Y} \\ &= \frac{1}{\hat{\sigma}_Y} \sum (Y - M_Y) \\ &= \frac{1}{\hat{\sigma}_Y} (\sum Y - NM_Y) \\ &= \frac{1}{\hat{\sigma}_Y} (NM_Y - NM_Y) \\ &= 0 . \end{aligned} \tag{A.15}$$

In order to show that the variance of the Z-scores is 1, it suffices to show that the sum of the squared Z-scores is $(N - 1)$.¹ This is shown by developing the formula for the sum of the squared Z-scores:

$$\begin{aligned}\sum Z^2 &= \sum \left(\frac{Y - M_Y}{\hat{\sigma}_Y} \right)^2 \\ &= \frac{1}{\hat{\sigma}_Y^2} \sum (Y - M_Y)^2 .\end{aligned}$$

But $(N - 1)\hat{\sigma}_Y^2 = \sum (Y - M_Y)^2$, hence:

$$\begin{aligned}\sum Z^2 &= \frac{1}{\hat{\sigma}_Y^2} \times (N - 1)\hat{\sigma}_Y^2 \\ &= (N - 1) .\end{aligned}\tag{A.16}$$

A.6.2 Back to the Z-scores

Applying the formula for a Z-score to a score of 85 from each of the distributions gives the following standardized scores.

- For the first distribution:

$$Z_1 = \frac{Y - M_Y}{\hat{\sigma}_Y} = \frac{85 - 75}{11.5} = \frac{10}{11.5} = .87 .$$

- For the second distribution:

$$Z_1 = \frac{Y - M_Y}{\hat{\sigma}_Y} = \frac{85 - 75}{14.9} = \frac{10}{14.9} = .67 .$$

- For the third distribution:

$$Z_1 = \frac{Y - M_Y}{\hat{\sigma}_Y} = \frac{85 - 75}{17} = \frac{10}{17} = .59 .$$

Subtracting the mean gives the distance of a score from the mean. Because the mean is 75 for all three distributions, this *deviation from the mean* is 10 in each case. Dividing by the standard deviation, which is different for each distribution, rescales this distance to express it in terms of the number of standard deviations. This rescaling provides a score measure that enables us to compare a score from one distribution with a score from another distribution. By standardizing a score we convert it to a measure that places that score in the context of its distribution. As an exercise, you can draw the histogram for each of the three distributions using Z-scores. You will see that the *shape* of the distributions is unchanged but that they have been scaled to a common unit.

¹ If we use the population variance σ we need to show that the sum of the squared Z-scores is equal to N , instead of $(N - 1)$.

B

The sum sign: $\sum$

B.1 Introduction

In this appendix we review the essential properties of the sum sign.

B.2 The sum sign

The $\sum$ sign (read: sigma) is used to denote the sum of a set of numbers, for example:

$$Y_1 + Y_2 + \cdots + Y_s + \cdots + Y_S = \sum_{s=1}^S Y_s$$

The initial value of the index is 1 and the terminal value is S . The formula is read as: ‘sum for small s going from 1 to big S of Y sub s ’.

When context allows it, this notation can be simplified. Specially, if the initial value is not given, it will be set to 1. Likewise, the default value of the terminal value is the same letter as the subscript but in uppercase letter, e.g. if the subscript is s the terminal value is S . So,

$$\sum_{s=1}^S Y_s = \sum_s^S Y_s = \sum_s Y_s = \sum Y_s$$

From this basic definition we can derive a handful of nice properties (in general, to prove them it suffices to develop the formulas).

1. If C is a constant, then

$$\sum_a C = \underbrace{C + C + \cdots + C}_A \text{ times} = AC$$

2. This property is equally true when dealing with several $\sum$ signs:

$$\sum_a \sum_s C = ASC$$

3. If C is a constant, then,

$$\sum_s CY_s = C \sum_s Y_s$$

4. The $\sum$ sign is commutative,

$$\sum_a \sum_s Y_{as} = \sum_s \sum_a Y_{as}$$

5. With several $\sum$ signs, some quantities can be constant for only one $\sum$ sign,

$$\sum_s \sum_a C_a Y_{as} = \sum_a \sum_s C_a Y_{as} = \sum_a (C_a \sum_s Y_{as}) = \sum_a C_a \sum_s Y_{as}$$

6. The $\sum$ sign and the + sign are distributive,

$$\sum_s (Y_s + Z_s) = \sum_s Y_s + \sum_s Z_s$$

7. 'Nice' equalities,

$$\left(\sum_{s=1}^2 Y_s\right)^2 = Y_1^2 + Y_2^2 + 2Y_1 Y_2 = \sum_{s=1}^2 Y_s^2 + 2Y_1 Y_2$$

8. More generally (this property will be quite often used for the computation of expected values),

$$\left(\sum_{s=1}^S Y_s\right)^2 = \sum_{s=1}^S Y_s^2 + \sum_{s \neq s'} Y_s Y_{s'} = \sum_{s=1}^S Y_s^2 + 2 \sum_{s < s'} Y_s Y_{s'}$$

Note, in passing, the important distinction (sometimes forgotten) between

$$\left(\sum_s Y_s\right)^2 \quad \text{and} \quad \sum_s Y_s^2$$

9. *Summation without subscript*

The subscript under the $\sum$ sign can be omitted. In this case, it is assumed that the summation is to be carried out only for the subscript present in the variables to be summed:

$$\sum M_{a\cdot} = \sum_a M_{a\cdot}; \quad \sum Y_{as} = \sum_a \sum_s Y_{as}$$

10. *Implicit summation*

The $\sum$ sign can be omitted. It suffices to replace the subscript used for the summation by a dot or a plus sign.

$$\sum_s Y_s = Y_{\cdot}; \quad \sum_s Y_{as} = Y_{a\cdot}; \quad \sum Y_{as} = Y_{\cdot\cdot}$$

$$\sum_a \left(\sum_s Y_{as}\right)^2 = \sum_a Y_{a\cdot}^2 = \sum Y_{a\cdot}^2$$

Or, for the + notation:

$$\sum_s Y_s = Y_{+}; \quad \sum_s Y_{as} = Y_{a+}; \quad \sum Y_{as} = Y_{++}$$

$$\sum_a \left(\sum_s Y_{as}\right)^2 = \sum_a Y_{a+}^2 = \sum Y_{a+}^2$$

C

Elementary probability: a refresher

C.1 Introduction

In this appendix we will introduce the basic notions of probability needed for understanding the statistical procedures that are covered in this book.

Probability theory gives the foundations of most statistical procedures. Even though the field of probability theory is vast and sophisticated, we can understand most of what is needed to master its most important contributions to methodology and statistics with a handful of rather simple notions. This chapter gives a rapid survey of these notions. If you are not familiar with them, you will realize that they consist essentially of learning a new vocabulary and some new notations. It is important that you learn these new words and master these new notations because we will use them in this book.

Incidentally, probability theory was initiated by Pascal (the French mathematician and philosopher) and Fermat (of Fermat's 'last theorem' fame) *circa* 1654, in order to solve problems dealing with games of dice and cards. This may explain why so many examples in probability are concerned with games (if you want to know more about the origins of probability theory, you can read the delicious recent book of Devlin, 2008).

C.2 A rough definition

A rough, working description of probability states that the goal of probability is to assign numbers to events such that these numbers reflect or express how likely is a given event.

For example, we want to be able to find a number expressing how likely it is to win a state lottery when buying a ticket. We know that without buying a ticket we have 'no chance' of winning the lottery, hence we want the probability of winning the lottery without buying a ticket to be zero. Even if we do not know how likely it is to win the lottery when buying a ticket, we know that it is more likely than without buying a ticket. Hence, this probability should be greater than zero (i.e. it should be a positive number). If we want to be able to compare the probability of two events, we need to have a common scale that allows these comparisons.

This way, we could make statements such as ‘it is more likely to win a state lottery than to be struck by lightning’ (I do not know if this statement is true or not by the way!). We need also to have a maximum value for the probability which corresponds to an event that we are certain will occur. For example, if we buy *all* the tickets for a state lottery, we are sure to win. By convention, this maximum value is set to the practical value of 1. We shall need, however, a more formal definition of probability in order to be able to work with it. Before we can do this, a small *détour* is necessary.

C.3 Some preliminary definitions

Before being able to give a more precise definition of probability, we need to define some other notions that are used in the definition. These notions of *experiment*, *event*, and *sample space*, are defined in the following section.

C.3.1 Experiment, event and sample space

In the vocabulary of probability, the term of experiment takes a different meaning than in psychology (so we will give another meaning to the term experiment in the following chapters when we refer to ‘scientific experiments’). As far as probability theory is concerned, an experiment is *any well-defined act or process which has a single well-defined outcome*. This outcome is called an *elementary event* or simply an *event* (or also a *sample point*). The set of all possible outcomes (or events) of an experiment is called the *sample space*. You can find two different notations for the sample space Ω (the capital Greek letter Omega) or $\mathcal{S}$ (capital S in a funny font). Therefore, the sample space of an experiment is the set of all possible events for this experiment. Elementary events can be combined to form *compound events* (also named *complex events* or *composed events*). Formally, a compound event is any subset (with more than one element) of the sample space.

For example, if you toss a coin and record if the coin falls on a Head or on a Tail, this is an experiment, because tossing a coin is a well-defined process, and there is only one well-defined outcome, namely: Head or Tail. Be careful to note that by a ‘single well-defined outcome’ we mean that there is only one *actual* outcome, even though there are several *possible* outcomes (i.e. two in the coin example). The sample space of this experiment is composed of two elementary events: {Head, Tail}.

Another example of an experiment: you measure the time a subject needs to read a word. This is a well-defined process and you have only one time per subject, and hence one well-defined outcome. Now, the sample space is more complex than in the previous example. Suppose that our measurement apparatus has a precision of one millisecond (a millisecond is one thousandth of a second, it is abbreviated ‘ms’), and that we know that a subject will always need less than 2 seconds to read a word. Then, the sample space is composed of all the values between 0 and 2000: $\{0, 1, \dots, 1999, 2000\}$, because these are the possible values in milliseconds that we can observe as an outcome for this experiment. Any of these numbers represents an *elementary event*. If we consider the answers between 500 ms and 1000 ms, then this subset of the sample space is a *compound event*.

As you can see, the notion of an experiment is a very broad one, and practically, any process or procedure that can be described without ambiguity is an experiment for probability theory.

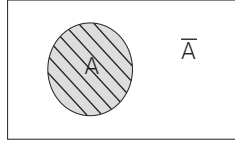


Figure C.1 A Venn diagram showing that the union of A and $\bar{A}$ is $\mathfrak{S}$.

C.3.2 More on events

As we have seen previously, events are subsets of the sample space. As such, the notions of set theory are relevant to describe them. Some of these notions are of particular interest here.

Because the sample space for a given experiment represents the set of all possible outcomes (i.e. events), it is equivalent to the *Universal set*¹ in set theory.

Now, if we call A a (simple or compound) event, the *complementary event* or *complement of A* , noted $\bar{A}$, is the set of events of the sample space that does not belong to A . Formally,

$$\bar{A} = \mathfrak{S} - A. \quad (\text{C.1})$$

Equivalently,

$$A \cup \bar{A} = \mathfrak{S}. \quad (\text{C.2})$$

where the symbol $\cup$ (read ‘cup’ or ‘union’) represents the union of two events.²

It is often very practical when dealing with sets (and hence with probability), to picture relations between sets or subsets with *Venn diagrams* like that in Figure C.1. The name ‘Venn’ is a tribute to the British mathematician John Venn who invented them in the late nineteenth century as a tool to make easy proofs. In a Venn diagram, the sample space (or universal set) is represented by a rectangle. Events (i.e. sets and subsets) are represented by ellipsoidal or circular regions.³ Regions of the space are very often shaded with different hues to reveal relationships between sets.

For example, suppose we toss a die (one of a pair of dice), and record the number on the top of the die. The sample space is made of the possible values of the die, namely the set:

$$\mathfrak{S} = \{1, 2, 3, 4, 5, 6\}. \quad (\text{C.3})$$

The event ‘obtaining an even number’ is a compound event. It corresponds to the subset noted

$$A = \{2, 4, 6\}. \quad (\text{C.4})$$

The complement of A is

$$\bar{A} = \mathfrak{S} - A = \{1, 3, 5\}. \quad (\text{C.5})$$

It corresponds to the event ‘obtaining an odd number’. We can verify that the union of A and $\bar{A}$ gives, indeed, the sample space:

$$A \cup \bar{A} = \{2, 4, 6\} \cup \{1, 3, 5\} = \{1, 2, 3, 4, 5, 6\} = \mathfrak{S}. \quad (\text{C.6})$$

¹ This is the set containing all the elements of a given domain. It is also called the *Universe of discourse*.

² The union of two sets is composed of the elements that belong either to one set or to both of them.

³ They are sometimes called, tongue in cheek, by mathematicians, *potatoids* when they are drawn by hand because then these regions have roughly the shape of a potato.

Because the union of A and $\bar{A}$ gives back the sample space, we say that these events are *exhaustive*. More generally, exhaustive events are defined as follows:

Definition C.1. We say that a set of events is exhaustive if their union gives the sample space.

With the die example, consider the following events:

- the event ‘obtaining an odd number’

$$A = \{1, 3, 5\} \quad (\text{C.7})$$

- the event ‘obtaining a prime number’

$$B = \{1, 2, 3, 5\} \quad (\text{C.8})$$

- the event ‘obtaining a multiple of 3’

$$C = \{3, 6\} \quad (\text{C.9})$$

- the event ‘obtaining a value of 4’

$$D = \{4\}. \quad (\text{C.10})$$

These events (i.e. A , B , C , and D) constitute an exhaustive set of events because the union of these four events gives back the sample space. Formally

$$A \cup B \cup C \cup D = \{1, 2, 3, 4, 5, 6\} = \mathfrak{S}. \quad (\text{C.11})$$

In contrast, the events A , B , and C are *not* an exhaustive set of events because their union does not give the sample space, namely

$$A \cup B \cup C = \{1, 2, 3, 5, 6\} \neq \mathfrak{S}. \quad (\text{C.12})$$

Another important and cognate notion refers to *exclusive* or *disjoint* events. This is the topic of the following definition.

Definition C.2. Two events are said to be mutually exclusive or disjoint if they cannot occur together. More formally, two events are mutually exclusive (or exclusive for short) if their intersection⁴ is empty.⁵

Figure C.2 illustrates the notion of exclusive elements. For another example, consider the following events (we are still using the die example):

$$A = \{1, 2\}, B = \{3, 6\}. \quad (\text{C.13})$$

These events are mutually exclusive because

$$A \cap B = \{1, 2\} \cap \{3, 6\} = \emptyset; \quad (\text{C.14})$$

where the symbol $\cap$ (read ‘cap’ or ‘intersection’ or ‘inter’) represents the intersection operation and the symbol $\emptyset$ represents the *empty* set (i.e. the set containing no elements).

⁴ The intersection of two sets is composed of the elements belonging to *both* sets.

⁵ The empty set contains *no* elements.

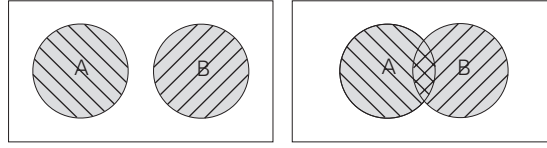


Figure C.2 (1) A and B are exclusive or disjoint events. (2) A and B are not exclusive events because their intersection $A \cap B$ is not empty.

More generally, a set of events is said to be mutually exclusive if all possible pairs of events are mutually exclusive. For example, the following events:

$$A = \{1, 2\}, B = \{3, 6\}, C = \{5\} \quad (\text{C.15})$$

are mutually exclusive because

$$\begin{aligned} A \cap B &= \{1, 2\} \cap \{3, 6\} = \emptyset \\ A \cap C &= \{1, 2\} \cap \{5\} = \emptyset \\ B \cap C &= \{3, 6\} \cap \{5\} = \emptyset. \end{aligned} \quad (\text{C.16})$$

Note that any event and its complement are both exhaustive and exclusive. (Do you see why? If not, check the definition.)

An event is a *joint* event if it is the intersection of two or more events. So, for example, given A and B , we can define a joint event $A \cap B$ (we will need this notion later on to define the notion of *independent* events).

C.3.3 Or, and, union and intersection

We have used the set theory terminology of intersection and union of events. We could have used, as well, the equivalent terminology of logic. In this case, the set notion of union corresponds to the logical function OR (i.e. the ‘inclusive or’), and the set notion of intersection corresponds to the logical function AND. Therefore,

$$A \cup B = A \text{ OR } B \quad (\text{C.17})$$

$$A \cap B = A \text{ AND } B. \quad (\text{C.18})$$

We will use whichever notation makes more sense in a given context, so try to memorize these synonyms.

We now have the necessary vocabulary for a more precise definition of probability.

C.4 Probability: a definition

The probability of an event is a number noted $\Pr(\text{event})$ taking values between 0 and 1. The value 0 means that we are sure that this event will not occur, and the value of 1 means that we are sure that this event will occur. Formally, a probability is a function associating to any event A of a given sample space $\mathfrak{S}$, a number, noted $\Pr(A)$, according to the following axioms (we indicate in parentheses, a ‘word interpretation’ of the axioms):

Axiom C.1. $0 \leq \Pr(A) \leq 1$. (*A probability is always between 0 and 1.*)

Axiom C.2. $\Pr(\mathfrak{S}) = 1$. (We are sure that some event from the sample space will occur.)

Axiom C.3. If two events A and B are mutually exclusive (i.e. $A \cap B = \emptyset$), then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$. (The probability of the union of exclusive events is the sum of their probability.)

From these axioms and from elementary set theory,⁶ we can derive several theorems. A sample of the theorems that we will need later on follows (the derivation of the theorems is left as an exercise). As for the axioms, we state the theorem, and we give in parentheses a word interpretation of the theorem.

Theorem C.1. $\Pr(\emptyset) = 0$. (We are sure that the impossible event will not occur.)

Theorem C.2. If A is an event and $\bar{A}$ is its complement, then $\Pr(A) = 1 - \Pr(\bar{A})$. Equivalently, $\Pr(A) + \Pr(\bar{A}) = 1$. (Either A or $\bar{A}$ will occur.)

Theorem C.3. If A and B are two events, then

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B). \quad (\text{C.19})$$

(The probability of obtaining the outcome A or B is the sum of their respective probabilities minus the probability of obtaining the outcome A AND B .) If we compare with axiom C.3, we see that it represents a particular case of this theorem. This happens because when A and B are exclusive $\Pr(A \cap B) = 0$.

Theorem C.4. If a set of events is exhaustive and exclusive then the sum of their probabilities is 1. Formally, if $\{A, B, \dots, X\}$ is a set of exhaustive and exclusive events, then $\Pr(A) + \Pr(B) + \dots + \Pr(X) = 1$.

(We are sure that some event from a set of exhaustive and exclusive events will occur.)

Theorem C.5. If A is a compound event composed of a set of N elementary events noted $\{A_1, A_2, \dots, A_n, \dots, A_N\}$, then the probability of A is the sum of the probabilities of its elementary events. Formally, if $A_n \cap A_{n'} = \emptyset$ for all pairs of indices such that $n \neq n'$, then

$$\Pr(A) = \sum_{n=1}^N \Pr(A_n). \quad (\text{C.20})$$

Theorem C.5 can be presented in a simplified form when all the elementary events in the space have the same probability (the technical term for this is *equiprobability*).

Theorem C.6. If $\#\{A\}$ denotes the number of elements of A , and $\#\{\mathfrak{S}\}$ the number of elements of $\mathfrak{S}$, and if all the elementary events of $\mathfrak{S}$ are equiprobable, then

$$\Pr(A) = \frac{\#\{A\}}{\#\{\mathfrak{S}\}}. \quad (\text{C.21})$$

This very important theorem is often stated as

$$\Pr(A) = \frac{\text{Number of favorable cases}}{\text{Number of possible cases}}. \quad (\text{C.22})$$

⁶ It is not necessary to know set theory to understand all that. It just makes it easier if you do.

As an illustration, suppose that we toss three coins (a nickel, a dime, and a quarter), and that we record which coin falls on *Head* and which one falls on *Tail*. Because each coin can take two different values, all three of them can give $2 \times 2 \times 2 = 2^3$ different combinations. So, the sample space is composed of $2^3 = 8$ elements which are:

$$\mathfrak{S} = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\} . \quad (\text{C.23})$$

(Make sure that you understand how these elements are obtained.) The first letter in each element represents the nickel, the second letter the dime, the third letter the quarter. For example, *THT* represents the event: ‘the nickel and the quarter fall on Tail, and the dime falls on Head’. Consider the event

$$A = \{\text{Obtaining 2 Heads}\} . \quad (\text{C.24})$$

This event corresponds to the set of following elementary events:

$$A = \{HHT, HTH, THH\} . \quad (\text{C.25})$$

Therefore, the probability of obtaining 2 Heads when tossing 3 coins is

$$\begin{aligned} \Pr(A) &= \frac{\# \{A\}}{\# \{\mathfrak{S}\}} \\ &= \frac{\text{Number of favorable cases}}{\text{Number of possible cases}} \\ &= \frac{\# \{HHT, HTH, THH\}}{\# \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}} \\ &= \frac{3}{8} \\ &= .375 . \end{aligned} \quad (\text{C.26})$$

C.5 Conditional probability

The idea of conditional probability deals with relationships between events. Specifically, using conditional probability we want to evaluate the probability of a specific event occurring *knowing* that another event actually happened.

For example, suppose (once again) that we toss a die. The sample space is:

$$\mathfrak{S} = \{1, 2, 3, 4, 5, 6\} . \quad (\text{C.27})$$

Now consider the event

$$A = \text{‘obtaining a prime number’} = \{1, 2, 3, 5\} , \quad (\text{C.28})$$

and the event

$$B = \text{‘obtaining an even number’} = \{2, 4, 6\} . \quad (\text{C.29})$$

Supposing you are told that the outcome of our experiment was a prime number: what is the probability of obtaining an even number? There are 4 ways of obtaining a prime number

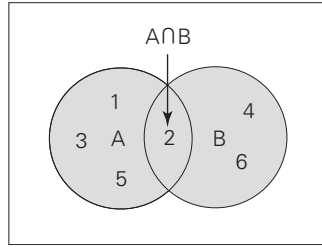


Figure C.3 There is just one even number in the set of prime numbers A . As $\# \{A\} = 4$, the conditional probability of B (obtaining an even number) given that A (obtaining a prime number) occurred is equal to $1/4 = .25$.

(i.e. $\# \{A\} = 4$). Of these 4 prime numbers, only one is even (i.e. $\# \{A \cap B\} = 1$). Therefore, using Equation C.22 we find that the probability of obtaining an even number is

$$\frac{\text{Number of favorable cases}}{\text{Number of possible cases}} = \frac{\# \{A \cap B\}}{\# \{A\}} = \frac{1}{4}. \quad (\text{C.30})$$

This is illustrated in Figure C.3.

From the previous discussion we can derive the following definition of *conditional probability*.

Definition C.3. If A and B are two events such that A is not the empty set, then the conditional probability of B knowing that A happened, denoted $\Pr(B | A)$, is the ratio of the number of elementary elements in the intersection of A and B to the number of elementary elements of A . Formally:

$$\Pr(B | A) = \frac{\# \{A \cap B\}}{\# \{A\}}. \quad (\text{C.31})$$

This equation is often presented in the equivalent form

$$\Pr(B | A) = \frac{\Pr(A \cap B)}{\Pr(A)}. \quad (\text{C.32})$$

In the context of conditional probability, $\Pr(B)$ is often called the *prior probability*, and $\Pr(B | A)$ the *posterior probability*.

So, we will say the probability of obtaining an even number when we toss a die *knowing* that the die gave a prime number is

$$\begin{aligned} & \Pr(\text{obtaining an even number} \mid \text{we have obtained a prime number}) \\ &= \frac{\Pr(\text{obtaining a prime and even number})}{\Pr(\text{obtaining a prime number})} \\ &= \frac{1/6}{4/6} \\ &= \frac{1}{4}. \end{aligned} \quad (\text{C.33})$$

Note that in this example, the *prior* probability of obtaining an even number is equal to $\frac{3}{6} = .5$.

C.5.1 Bayes' theorem

From the definition of conditional probability, we can derive one of the most important theorems of decision theory known as *Bayes' theorem*.⁷ This theorem connects the conditional probability of A given B to the probability of B given A .

To make the point clearer, we start with an example. You have the impression that you have some mild memory problems, and you have read that these memory problems could be caused by a brain tumor requiring urgent and dangerous brain surgery. You visit a neuropsychologist who administers to you a memory test. The neuropsychologist gives you the following information. In this clinic, 2 patients out of every 100 patients have been found to have a brain tumor. The results of the test are presented in a binary form, either you score above or below 50. Of 100 patients with a brain tumor detected one year after the test, 80 had a score below 50. By contrast, of 100 patients without a tumor, 30 had a score below 50. Now, knowing that you score below 50 on the test, how likely is it that you have a brain tumor? Try to make a rough estimation from the data presented. Most people think that it is *very likely* that you have a brain tumor given the results of the test.

To formalize our problem, we need to identify the different events of interest. Let us call A the event 'to have a brain tumor'. The complement of A is 'not to have a brain tumor'. From the numbers given above (2 persons out of 100 have a brain tumor), we can compute the following prior probabilities:

$$\Pr(\text{Have a brain tumor}) = \Pr(A) = \frac{2}{100} = .02 \quad (\text{C.34})$$

$$\Pr(\text{Not to have a brain tumor}) = \Pr(\bar{A}) = 1 - \Pr(A) = 1 - .02 = .98. \quad (\text{C.35})$$

If we call B the event 'to score below 50 on the test', we can find two conditional probabilities from the data given by the neuropsychologist. The statement

Of 100 patients with a brain tumor detected one year after the test, 80 have a score below 50

means that the probability of scoring below 50 given that you have a tumor is

$$\Pr(B | A) = \frac{80}{100} = .80. \quad (\text{C.36})$$

On the other hand, the statement that

Of 100 patients without a tumor, 30 will have a score below 50

means that the probability of scoring below 50, given that you do *not* have a tumor, is

$$\Pr(B | \bar{A}) = \frac{30}{100} = .30. \quad (\text{C.37})$$

So, what we want to find is the probability of A (to have a tumor) knowing B (you score below 50 on the memory test). The information that we have is $\Pr(A)$, $\Pr(B | A)$ and $\Pr(B | \bar{A})$. Bayes' theorem gives the formula to solve this problem.

⁷ The theorem is named from the English clergyman Thomas Bayes. It dates from the second half of the eighteenth century.

Theorem C.7 (Bayes' theorem). *If A and B are two (non-empty) events, the probability of A given B can be computed as*

$$\Pr(A | B) = \frac{\Pr(B | A) \Pr(A)}{\Pr(B | A) \Pr(A) + \Pr(B | \bar{A}) \Pr(\bar{A})} . \quad (\text{C.38})$$

An alternative (but equivalent) formulation is

$$\Pr(A | B) = \frac{\Pr(B | A) \Pr(A)}{\Pr(B)} . \quad (\text{C.39})$$

(the proof is given as a digression in the following section).

Using this theorem, we can compute that the probability of having a tumor, knowing that you have below 50 on the memory test, is

$$\begin{aligned} \Pr(A | B) &= \frac{\Pr(B | A) \Pr(A)}{\Pr(B | A) \Pr(A) + \Pr(B | \bar{A}) \Pr(\bar{A})} \\ &= \frac{.80 \times .02}{.80 \times .02 + .30 \times .98} \\ &= \frac{.016}{.016 + .294} \\ &\approx .0516 , \end{aligned} \quad (\text{C.40})$$

which means that you have actually, roughly *5 chances in one hundred* (i.e. a .05 chance) to have a tumor knowing that you scored below 50 on the memory test. This example shows, in passing, that most of us are quite poor at trying to guess probability from raw data (did you estimate correctly this value from the data at the beginning of this section?). This justifies the use of formal approaches as a remedy for the weaknesses of our intuition.

C.5.2 Digression: proof of Bayes' theorem

Reading this digression is well advised give you some insight into the techniques used in probability, and should convince you that they are rather elementary.

The first step is to express the probability of any event in terms of its intersection with another event.

Theorem C.8. *If A and B are two events, then*

$$\Pr(B) = \Pr(A \cap B) + \Pr(\bar{A} \cap B) . \quad (\text{C.41})$$

This equation comes from the fact that the two events $\Pr(A \cap B)$ and $\Pr(\bar{A} \cap B)$ are exclusive. In other words, what theorem C.8 says is that any element of B either belongs to A or does not belong to A . This is illustrated in Figure C.4.

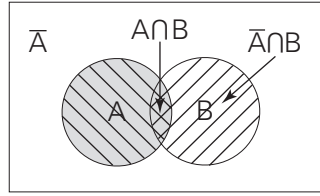


Figure C.4 The set B is always equal to $\{B \cap A\} + \{B \cap \bar{A}\}$. Because $\{B \cap A\}$ and $\{B \cap \bar{A}\}$ are exclusive events, $\Pr(B) = \Pr(B \cap A) + \Pr(B \cap \bar{A})$.

The next step is to rewrite Equation C.32:

$$\Pr(B | A) = \frac{\Pr(A \cap B)}{\Pr(A)}$$

as

$$\Pr(A \cap B) = \Pr(B | A)\Pr(A). \quad (\text{C.42})$$

Changing the roles of A and B we find also from Equation C.32 that

$$\Pr(A \cap B) = \Pr(A | B)\Pr(B). \quad (\text{C.43})$$

Substituting $\bar{A}$ for A in Equation C.42 we find also that

$$\Pr(\bar{A} \cap B) = \Pr(B | \bar{A})\Pr(\bar{A}). \quad (\text{C.44})$$

From Equation C.32 and substituting A and B we find that

$$\Pr(A | B) = \frac{\Pr(A \cap B)}{\Pr(B)}. \quad (\text{C.45})$$

Substituting C.43 in C.45 gives Equation C.39 as:

$$\Pr(A | B) = \frac{\Pr(B | A)\Pr(A)}{\Pr(B)}. \quad (\text{C.46})$$

Replacing $\Pr(B)$ in Equation C.46 by the expression from Equation C.41 gives

$$\Pr(A | B) = \frac{\Pr(B | A)\Pr(A)}{\Pr(A \cap B) + \Pr(\bar{A} \cap B)}. \quad (\text{C.47})$$

Using Equations C.42 and C.44 we finally obtain Bayes' theorem:

$$\Pr(A | B) = \frac{\Pr(B | A)\Pr(A)}{\Pr(B | A)\Pr(A) + \Pr(B | \bar{A})\Pr(\bar{A})}. \quad (\text{C.48})$$

C.6 Independent events

Intuitively, when we say that two events are independent of each other, we mean that there is no relationship whatsoever between these events. Equivalently, using the vocabulary of conditional probability, we can say that, when two events are independent, knowing that one of them has occurred does not give us any information about the occurrence of the other one. This means that when two events are independent, then the conditional probability of one event, given the other, should be the same as the probability of the first event. This leads to the following formal definition.

Definition C.4 (Independence 1). *Two (non-empty) events A and B are independent if*

$$\Pr(A | B) = \Pr(A) . \quad (\text{C.49})$$

Equation C.49 may be rewritten in order to give an equivalent but more practical and more frequent definition. Recall that

$$\Pr(A | B) = \frac{\Pr(A \cap B)}{\Pr(B)} , \quad (\text{C.50})$$

which implies together with Equation C.49 that when A and B are independent

$$\Pr(A) = \frac{\Pr(A \cap B)}{\Pr(B)} . \quad (\text{C.51})$$

This equation, rearranged, gives the alternative definition of independence.

Definition C.5 (Independence 2). *If, and only if, A and B are two non-empty independent events, then*

$$\Pr(A \cap B) = \Pr(A) \Pr(B) . \quad (\text{C.52})$$

Let us go back to our favorite die for an example. The sample space is (as usual):

$$\mathfrak{S} = \{1, 2, 3, 4, 5, 6\} . \quad (\text{C.53})$$

Consider the event ‘obtaining an even number’:

$$A = \{2, 4, 6\} , \quad (\text{C.54})$$

and the event ‘obtaining a number smaller than 3’:

$$B = \{1, 2\} . \quad (\text{C.55})$$

With the first definition of independence (i.e. Definition C.4), in order to check the independence of A and B , we need to prove that $\Pr(A | B)$ is equal to $\Pr(A)$.

We start by evaluating the first term

$$\begin{aligned} \Pr(A | B) &= \frac{\#\{A \cap B\}}{\#\{B\}} \\ &= \frac{1}{2} ; \end{aligned} \quad (\text{C.56})$$

we now evaluate the second term

$$\begin{aligned} \Pr(A) &= \frac{\#\{A\}}{\#\{\mathfrak{S}\}} \\ &= \frac{3}{6} \\ &= \frac{1}{2} . \end{aligned} \quad (\text{C.57})$$

Because $\Pr(A | B)$ is equal to $\Pr(A)$ we conclude that A and B are independent.

With the second definition for independence (i.e. Definition C.5) we need to show that $\Pr(A \cap B)$ is equal to $\Pr(A)\Pr(B)$. We start by evaluating the term $\Pr(A \cap B)$:

$$\begin{aligned}\Pr(A \cap B) &= \frac{\#\{A \cap B\}}{\#\{\mathfrak{S}\}} \\ &= \frac{1}{6} .\end{aligned}\tag{C.58}$$

Then we proceed to evaluate the term $\Pr(A)\Pr(B)$:

$$\begin{aligned}\Pr(A)\Pr(B) &= \frac{\#\{A\}}{\#\{\mathfrak{S}\}} \times \frac{\#\{B\}}{\#\{\mathfrak{S}\}} \\ &= \frac{1}{2} \times \frac{1}{3} \\ &= \frac{1}{6} .\end{aligned}\tag{C.59}$$

Because $\Pr(A \cap B)$ is equal to $\Pr(A)\Pr(B)$, we conclude that A and B are independent (we will, indeed, always reach the same conclusion using either the first or the second definition of independence).

C.7 Two practical counting rules

From the previous definitions and theorems, we can derive two practical rules for evaluating probabilities of compound events. These rules are the *product* rule and the *addition* rule. These two rules are quite often used to compute probabilities of compound events.

C.7.1 The product rule

From the second definition of independence (Definition C.5), we know that if A and B are independent events, then $\Pr(A \cap B) = \Pr(A \text{ AND } B)$ is equal to $\Pr(A)\Pr(B)$. This can be generalized to several independent events to give the product rule.

Rule C.1 (Product Rule). *The probability of the joint occurrence of several independent events is the product of their respective probabilities. With an equation, the rule states that if $\{A, B, \dots, X\}$ are a set of independent events, their joint probability is*

$$\begin{aligned}\Pr(A \cap B \cap \dots \cap X) &= \Pr(A \text{ AND } B \text{ AND } \dots \text{ AND } X) \\ &= \Pr(A) \times \Pr(B) \times \dots \times \Pr(X) .\end{aligned}\tag{C.60}$$

For example, if we toss two (fair) coins and count the number of Heads, the first coin has a probability of landing Head of $\frac{1}{2}$, and the second coin has a probability of landing Head of $\frac{1}{2}$. These two events are independent (because they cannot influence each other). Therefore, the probability of having the first coin landing on Head *and* the second coin landing on Head is

$$\begin{aligned}\Pr(2 \text{ Heads}) &= \Pr(\text{first coin lands Head AND second coin lands Head}) \\ &= \Pr(\text{first coin lands Head}) \times \Pr(\text{second coin lands Head}) \\ &= \frac{1}{2} \times \frac{1}{2} .\end{aligned}\tag{C.61}$$

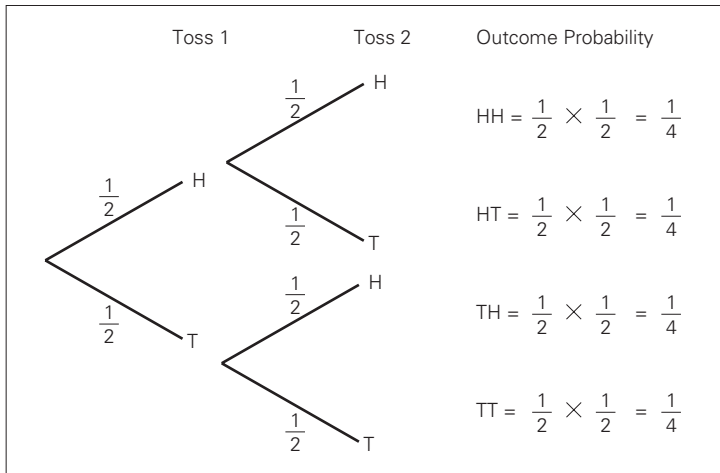


Figure C.5 A tree diagram describing the possible outcomes of tossing two independent coins along with the probability assigned to each outcome. The probability of an event represented by a leaf of the tree is the product of the probabilities assigned to each of its branches [e.g. $\Pr(HH) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$].

A convenient way of describing the probabilities of a set of joint events is to use a ‘tree diagram’ as illustrated in Figure C.5. A tree diagram lists all the possible outcomes of experiments composed of joint events. For example, when tossing two coins, each outcome is a joint event representing the result of the experiment for the first coin **AND** the result of the experiment for the second coin. In order to represent all the possible outcomes of this experiment, we start by drawing two branches corresponding to the outcomes for the first coin, and we assign its probability (i.e. $\frac{1}{2}$) to each branch. Then, for each outcome of the first coin, we draw two new branches representing the outcomes for the second coin, and, again, we assign to each branch its probability. This way, we obtain 4 different final events that we will call ‘leaves’. Each of these leaves corresponds to an elementary event of the experiment ‘tossing two coins’. In order to obtain the probability of each elementary event, we just multiply the probability assigned to each branch corresponding to this event.

C.7.2 Addition rule

From Axiom C.3, on page 431 and from Theorem C.5, on page 431 we can derive the *addition rule* for the probability of the union of *exclusive events*.

Rule C.2 (Addition Rule for exclusive events). *The probability of the union of several exclusive events is the sum of their respective probabilities. With an equation, the rule states that if $\{A, B, \dots, X\}$ is a set of exclusive events, the probability of their union is:*

$$\begin{aligned} \Pr(A \cup B \cup \dots \cup X) &= \Pr(A \text{ OR } B \text{ OR } \dots \text{ OR } X) \\ &= \Pr(A) + \Pr(B) + \dots + \Pr(X) . \end{aligned} \tag{C.62}$$

For example, using again the two coins of Figure C.5, the event ‘obtaining at most one Head with two coins’ is the union of three exclusive events: $\{TT, HT, TH\}$. Therefore, the

probability of obtaining at most one Head is:

$$\begin{aligned}\Pr(\text{at most 1 Head}) &= \Pr(TT \cup HT \cup TH) \\ &= \Pr(TT \text{ OR } HT \text{ OR } TH) \\ &= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} \\ &= \frac{3}{4}.\end{aligned}$$

There is also a more general version of the addition rule for non-exclusive events. We just give here the version for 2 and 3 non-exclusive events.

Rule C.3 (Addition Rule for 2 non-exclusive events). *The probability of two non-exclusive events is the sum of their respective probabilities minus the probabilities of their intersection (this rule is just a restatement of Theorem C.3). With an equation the rule states that*

$$\begin{aligned}\Pr(A \cup B) &= \Pr(A \text{ OR } B) \\ &= \Pr(A) + \Pr(B) - \Pr(A \cap B).\end{aligned}\tag{C.63}$$

This rule can be extended to more than 2 events; we give here the version for three events (extension to the general case is left as an exercise).

Rule C.4 (Addition Rule for 3 non-exclusive events). *The probability of three non-exclusive events noted A, B and C is given by the following formula:*

$$\begin{aligned}\Pr(A \cup B \cup C) &= \Pr(A \text{ OR } B \text{ OR } C) \\ &= \Pr(A) + \Pr(B) + \Pr(C) \\ &\quad - \Pr(A \cap B) - \Pr(A \cap C) - \Pr(B \cap C) \\ &\quad + \Pr(A \cap B \cap C).\end{aligned}\tag{C.64}$$

Chapter summary

C.8 Key notions of the chapter

Below are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

Experiment	Exhaustive events and exclusive
Event—elementary and complex (or compound)	(or disjoint) events
Sample space	Probability

Conditional probability

Independent events

Product rule

Addition rule

C.9 New notations

Below are the new notations introduced in this chapter. Test yourself on their meaning.

Sample space: $\mathfrak{S}$ Event: A Complementary event: $\overline{A}$ Union of 2 events: $\cup$ or **OR**Intersection of 2 events: $\cap$ or **AND**Empty set: $\emptyset$ Probability of an event: $\Pr(\text{event})$ Number of elements of A : $\#\{A\}$ Number of elements of $\mathfrak{S}$: $\#\{\mathfrak{S}\}$ Probability of B knowing that A happened: $\Pr(B | A)$

C.10 Key formulas of the chapter

Below are the main formulas introduced in this chapter: try to go through them and understand what they mean.

Probability definition:

Axioms:

1. $0 \leq \Pr(A) \leq 1$.
2. $\Pr(\mathfrak{S}) = 1$.
3. If $A \cap B = \emptyset$, then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.

Theorems:

1. $\Pr(\emptyset) = 0$.
2. If A is an event and $\overline{A}$ is its complement, then $\Pr(A) = 1 - \Pr(\overline{A})$.
3. If A and B are two events, then

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

4. If $\{A, B, \dots, X\}$ is a set of exhaustive events, then

$$\Pr(A) + \Pr(B) + \dots + \Pr(X) = 1$$

5. If $A_n \cap A_{n'} = \emptyset$ for all pairs of indices such that $n \neq n'$ then

$$\Pr(A) = \sum_{n=1}^N \Pr(A_n)$$

6. If all the elementary events of $\mathfrak{S}$ are equiprobable, then

$$\Pr(A) = \frac{\#\{A\}}{\#\{\mathfrak{S}\}}$$

or equivalently:

$$\Pr(A) = \frac{\text{Number of favorable cases}}{\text{Number of possible cases}}$$

Conditional probability

$$\Pr(B | A) = \frac{\#(A \cap B)}{\#(A)} \quad \Pr(B | A) = \frac{\Pr(A \cap B)}{\Pr(A)}$$

Independent events

$$\Pr(A | B) = \Pr(A)$$

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B)$$

Product rule

$$\begin{aligned} \Pr(A \cap B \cap \dots \cap X) &= \Pr(A \text{ AND } B \text{ AND } \dots \text{ AND } X) \\ &= \Pr(A) \times \Pr(B) \times \dots \times \Pr(X) \end{aligned}$$

Addition rule

- exclusive events

$$\begin{aligned} \Pr(A \cup B \cup \dots \cup X) &= \Pr(A \text{ OR } B \text{ OR } \dots \text{ OR } X) \\ &= \Pr(A) + \Pr(B) + \dots + \Pr(X) \end{aligned}$$

- non-exclusive events

$$\begin{aligned} \Pr(A \cup B) &= \Pr(A \text{ OR } B) \\ &= \Pr(A) + \Pr(B) - \Pr(A \cap B) \end{aligned}$$

$$\begin{aligned} \Pr(A \cup B \cup C) &= \Pr(A \text{ OR } B \text{ OR } C) \\ &= \Pr(A) + \Pr(B) + \Pr(C) \\ &\quad - \Pr(A \cap B) - \Pr(A \cap C) - \Pr(B \cap C) \\ &\quad + \Pr(A \cap B \cap C) . \end{aligned} \tag{C.65}$$

C.11 Key questions of the chapter

Below are some questions about the content of this chapter. All the answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- * Why is a probability always between 0 and 1?
- * What condition(s) should be satisfied for the product rule to apply?
- * What condition(s) should be satisfied for the addition rule to apply?
- * What is the set equivalent of the AND and OR operations?
- * Give some examples of the use of Bayes' theorem.
- * What do we mean by 'it is probable that it will rain tomorrow'?
- * What do we mean when we say it is probable that Gustav Mahler composed his first symphony before the 20th century? Can we use the notion of 'in the long run' here?

D

Probability distributions

D.1 Introduction

In this appendix we will expand the notions of elementary probability in order to describe, in a convenient way, the set of outcomes of a given experiment. The main idea is to assign a number to each possible outcome of the experiment (for example, if we toss 5 coins, we can count the number of Heads), and then assign a probability to this number. We call this number a *random variable*. The set of all the possible values of a random variable, each of them being assigned a probability, is called a *probability distribution*.

Probability distributions can be plotted, and they can be described like any set of numbers by their mean (which we call the *expected value*), and by their variance. We will also introduce our first important probability distribution named the *binomial distribution*. This probability distribution is used to describe the outcome of experiments in which several independent processes can take one of two values (i.e. tossing 5 coins, when each of them can land Head or Tail).

D.2 Random variable

The term ‘random variable’ may seem a bit abstruse but it is, in fact, a rather simple notion. The term ‘variable’ is used because a random variable can take several different values (hence it varies), and ‘random’ means that how it varies has something to do with probability (random means ‘occurring by chance’). A random variable is a way of associating a number with each possible elementary event of an experiment. More formally we have the following definition.

Definition D.1 (Random variable). *A random variable Y is a function that associates one and only one number to each elementary event of some sample space $\mathfrak{S}$. Note that from this definition, two different elementary events can have the same number associated to each of them.*

Event	<i>HHH</i>	<i>HHT</i>	<i>HTH</i>	<i>HTT</i>	<i>THH</i>	<i>THT</i>	<i>TTH</i>	<i>TTT</i>
<i>Y</i>	3	2	2	1	2	1	1	0

Table D.1 Values of the random variable $Y = \{\text{Number of Heads}\}$ for the 8 possible elementary events when tossing three coins.

For example, if we toss (again) a nickel, a dime, and a quarter and if we record which coin falls on *Head* and which one falls on *Tail*, the sample space is composed of $2^3 = 8$ elements which are:

$$\mathfrak{S} = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}. \quad (\text{D.1})$$

The first letter in each element represents the nickel, the second letter the dime, the third letter the quarter. For example, *THT* represents the event: ‘the nickel and the quarter land on Tail, and the dime lands on Head’. Now we can define the random variable Y as the ‘number of Heads per event’. Because we have three coins, the possible values for Y are $\{0, 1, 2, 3\}$.

- The event $\{HHH\}$ corresponds to a value of $Y = 3$ [formally we should write something like $Y(\{HHH\}) = 3$, but the context is clear enough not to overdo formalism].
- The events $\{HHT, HTH, THH\}$ correspond to a value of $Y = 2$.
- The events $\{HTT, THT, TTH\}$ correspond to a value of $Y = 1$.
- The event $\{TTT\}$ corresponds to a value of $Y = 0$.

We can summarize these results in Table D.1.

D.3 Probability distributions

From the previous example (cf. Table D.1), we can find the probability of obtaining any given value of the random variable $Y = \{\text{Number of Heads}\}$. In order to do so, we first remark that the elementary events are all equiprobable, and that we can have zero, one, or more elementary events corresponding to a given value of Y . Hence, at any value of Y there corresponds a compound event of $\mathfrak{S}$. Note, in passing, that this event can be the empty event. For example, there is no way to obtain 4 Heads with three coins and the empty event corresponds to the value of $Y = 4$ for our present example. Because values of Y correspond to compound events, the probability of obtaining any value of Y , which is recalled hereunder, can be computed from Formula C.22 (from Appendix C, page 431) as

$$\Pr(\text{Event}) = \frac{\text{Number of favorable cases}}{\text{Number of possible cases}}. \quad (\text{D.2})$$

With this formula, we can compute the probability of obtaining any given value of Y as follows:

- For $Y = 0$,

$$\Pr(Y = 0) = \frac{\#\{TTT\}}{\#\{\mathfrak{S}\}} = \frac{1}{8} = .125. \quad (\text{D.3})$$

- For $Y = 1$,

$$\Pr(Y = 1) = \frac{\#\{HTT, THT, TTH\}}{\#\{\mathfrak{S}\}} = \frac{3}{8} = .375. \quad (\text{D.4})$$

- For $Y = 2$,

$$\Pr(Y = 2) = \frac{\#\{THH, HTH, HHT\}}{\#\{\mathcal{S}\}} = \frac{3}{8} = .375. \quad (\text{D.5})$$

- For $Y = 3$,

$$\Pr(Y = 3) = \frac{\#\{HHH\}}{\#\{\mathcal{S}\}} = \frac{1}{8} = .125. \quad (\text{D.6})$$

Actually, we have just computed the probability distribution of the random variable Y . Formally we have the following definition:

Definition D.2 (Probability distribution). *A probability distribution is a list of all the possible values of a random variable with a probability assigned to each possible value of the random variable.*

There are several ways to represent a probability distribution. The first one is to gather these values as in Table D.2. Another very convenient way is to plot them as histograms or probability polygons as in Figure D.1. When plotting a probability distribution, the abscissa (i.e. the horizontal axis) always represents the values of the random variable and the ordinate (i.e. the vertical axis) always represents the probability assigned to the values of the random variable.

In addition to plotting a probability distribution, we want to be able to *describe* it in a more concise way than by listing all its possible values with their probability. As for many distributions, we will use two indices to describe a probability distribution: the mean, and the variance. The mean gives the most likely value of the probability distribution and the variance gives an indication of the dispersion of the values of the random variable around their mean. The concept of *expected value* or *mathematical expectation* is an important notion related to the mean, and the variance of a probability distribution. We will introduce this concept first and then use it in relation to the mean and variance.

Values of Y	0	1	2	3
$\Pr(Y)$	.125	.375	.375	.125

Table D.2 Probability distribution for the random variable $Y = \{\text{Number of Heads}\}$ when tossing three coins (compare with Table D.1).

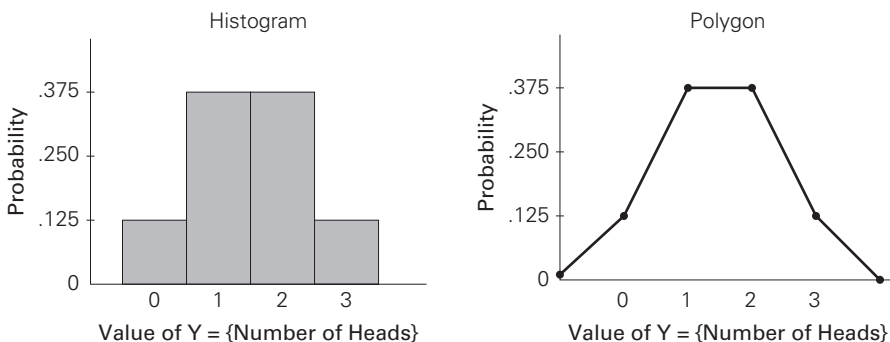


Figure D.1 Histogram and polygon of the probability distribution for the random variable $Y = \{\text{Number of Heads}\}$ when tossing three coins (compare with Table D.2).

D.4 Expected value and mean

The notion and the name of expected value betrays, in a way, the frivolous origin of probability theory (i.e. describing games and their outcomes). Consider the example of tossing three coins. Suppose now that we decide to use these three coins as a way of gambling.

- If we obtain 3 Heads, you'll give me \$10,
- if we obtain 2 Heads, you'll give me \$20,
- if we obtain 1 Head, I'll give you \$10,
- if we obtain 0 Heads, I'll give you \$40.

The question is *how much could you expect to win or lose* if you play this game. First, because we are assigning some numbers (i.e. the money you'll win or lose) to any possible outcome of the experiment, we are actually defining a random variable. Let us call it W (for a change, because we have been using Y in the same example to count the number of Heads). If we remember that losing some money corresponds to a *negative* value, then W , from your perspective, can take the values

$$\{-\$10, -\$20, \$10, \$40\} . \quad (\text{D.7})$$

We can compute the probability associated with each of these values as we did previously:

- for $W = -\$10$,

$$\Pr(W = -\$10) = \frac{\#\{HHH\}}{\#\{\mathfrak{S}\}} = \frac{1}{8} = .125 \quad (\text{D.8})$$

- for $W = -\$20$,

$$\Pr(W = -\$20) = \frac{\#\{HHT, HTH, THH\}}{\#\{\mathfrak{S}\}} = \frac{3}{8} = .375 \quad (\text{D.9})$$

- for $W = \$10$,

$$\Pr(W = \$10) = \frac{\#\{HTT, THT, TTH\}}{\#\{\mathfrak{S}\}} = \frac{3}{8} = .375 \quad (\text{D.10})$$

- for $W = \$40$,

$$\Pr(W = \$40) = \frac{\#\{HHH\}}{\#\{\mathfrak{S}\}} = \frac{1}{8} = .125. \quad (\text{D.11})$$

So, for example, if we play long enough, you can expect to win $-\$10$ (i.e. to lose \$10) 12.5% of the time, to win $-\$20$, 37.5% of the time, to win \$10, 37.5% of the time, and to win \$40, 12.5% of the time. Therefore, on the average, the amount of money you can *expect* to win is:

$$\begin{aligned} & (-\$10 \times .125) + (-\$20 \times .375) + (\$10 \times .375) + (\$40 \times .125) \\ &= -\$1.25 - \$7.50 + \$3.75 + \$5.00 \\ &= -\$8.75 + \$8.75 \\ &= 0 . \end{aligned} \quad (\text{D.12})$$

Your expected gain (or loss) is zero. Therefore, you can expect neither to win nor to lose anything *in the long run* if you play this game. The amount you can expect to win is the *expected value* or *mathematical expectation* of the random variable W .

We can now proceed to a more formal definition of the expected value of a random variable.

Definition D.3 (Expected value). *If Y is a random variable taking N values*

$$Y_1, Y_2, \dots, Y_N$$

with the probability assigned to each value being

$$\Pr(Y_1), \Pr(Y_2), \dots, \Pr(Y_N) ,$$

then, the expected value of Y , or mathematical expectation of Y , is noted $\mathbf{E}\{Y\}$, or μ_Y , and is computed as

$$\begin{aligned} \mathbf{E}\{Y\} &= \mu_Y \\ &= \Pr(Y_1) \times Y_1 + \Pr(Y_2) \times Y_2 + \dots + \Pr(Y_N) \times Y_N \\ &= \sum_n^N \Pr(Y_n) \times Y_n . \end{aligned} \tag{D.13}$$

The expected value of Y is also called the mean of the random variable.

From the definition, it is apparent that the expected value of a random variable is simply a *weighted* average for which each value of the random variable is weighted by its probability. Therefore, the expected value gives the value most likely to be obtained, *in the long run*, if we were to repeat the experiment. In addition to knowing the expected value, it is also important to know how much the values obtained can differ or vary from their mean. This index is called the *variance* of the random variable. We define it in the following section.

D.5 Variance and standard deviation

The essential idea behind the concept of variance of a random variable is to compute an index reflecting the dispersion of the values of the random variable around their mean.

A quantity that reflects the dispersion of a value around its mean is the *distance* or *deviation* from this value to the mean. To compute the distance from value Y_i to its mean, we simply subtract the mean from this value:

$$\text{distance from } Y_i \text{ to } \mu_Y = Y_i - \mu_Y . \tag{D.14}$$

The more dispersed the values of Y , the larger, on average, are the magnitudes of the deviations. Therefore, we want our index of dispersion to reflect the average magnitude of the deviations of the values of the Y_i around their mean. A first idea could be to use the average deviation, but the sum (and hence the average) of the deviations of a set of scores around their mean is *always* zero. (This is shown in Appendix A, Section A.6.1, page 422.) Actually, the easiest way to eliminate this problem and to take into account the *magnitude* of the deviations, is to square each of the deviations before summing them. This way, we always deal with positive numbers. We use squared numbers because they yield a lot of interesting properties (such the Pythagorean theorem) which are particularly relevant for the chapters dealing with analysis of variance.

So, our index of dispersion, named the *variance*, is the weighted average of the square deviations of the scores around their mean. More specifically, the *variance* of a random variable is the *expected value of the squared deviations of the values of the random variable from their mean*. We will formalize the concept of variance in the next definition.

Definition D.4 (Variance). *If Y is a random variable taking N values*

$$Y_1, Y_2, \dots, Y_N$$

with the probability assigned to each value being

$$\Pr(Y_1), \Pr(Y_2), \dots, \Pr(Y_N),$$

and with μ_Y being its expected value (i.e. mean), the variance of Y is denoted σ_Y^2 or $\text{Var}(Y)$ and is computed as

$$\begin{aligned} \sigma_Y^2 &= \text{Var}(Y) = \mathbf{E}\{(Y - \mu_Y)^2\} \\ &= \Pr(Y_1) \times (Y_1 - \mu_Y)^2 + \Pr(Y_2) \times (Y_2 - \mu_Y)^2 \\ &\quad + \dots + \Pr(Y_N) \times (Y_N - \mu_Y)^2 \\ &= \sum_n^N \Pr(Y_n) \times (Y_n - \mu_Y)^2. \end{aligned} \quad (\text{D.15})$$

The variance is the expected value of the squared deviation of the values of Y from their mean. The square root of the variance is called the standard deviation of the random variable. It is denoted σ_Y .

A large variance indicates a large dispersion, a small variance indicates that the values of the random variable are, on the average, close to the mean. Recall that the main reason for defining the standard deviation in addition to the variance is to have a measure of dispersion that is expressed in the same unit as the random variable (the variance is expressed in squared units). As an illustration, for the gambling example, the standard deviation is in dollars, but the variance is expressed in squared dollars (\$²) (quite an inflated type of currency!).

As an illustration, we compute now the variance for the dollar game example that we have used for illustrating the computation of the expected value. Recall that the expected value of W was $\mathbf{E}\{W\} = 0$. Table D.3 details the different quantities needed to compute σ_W^2 :

$$\begin{aligned} \sigma_W^2 &= \sum_n^N \Pr(Y_n) \times (Y_n - \mu_Y)^2 \\ &= .125 \times (-10)^2 + .375 \times (-20)^2 + .375 \times (10)^2 + .125 \times (40)^2 \\ &= .125 \times 100 + .375 \times 400 + .375 \times 100 + .125 \times 1600 \\ &= 12.5 + 150 + 37.5 + 200 \\ &= \$^2 400. \end{aligned} \quad (\text{D.16})$$

The standard deviation of W is the square root of the variance. Its value is

$$\sigma_W = \sqrt{\sigma_W^2} = \sqrt{\$^2 400} = \$20. \quad (\text{D.17})$$

Values of Y	0	1	2	3
$\Pr(Y)$	.125	.375	.375	.125
Values of W	40	10	-20	-10
$(W - \mu_W)^2$	1600	100	400	100
$\Pr(W) \times (W - \mu_W)^2$	200	37.5	150	12.5

Table D.3 Probability distribution for the random variable $Y = \{\text{Number of Heads}\}$ when tossing three coins. Probability distribution of the random variable W . The different steps to compute the variance of W are indicated in the Table. $\sigma_W^2 = 200 + 37.50 + 150 + 12.50 = \$^2400$ (\$^2 means ‘squared dollars’, see text for explanation); $\sigma_W = \$20$.

It is expressed in dollars, like the mean of W . Often, when the context is sufficiently clear, the unit of measurement of the variance and the standard deviation can be omitted. For instance, we could refer to our example as showing an expected value or mean of zero dollars with a variance of 400 (‘dollars squared’ being implied by the context).

D.6 Standardized random variable: Z-scores

Often we need to compare different random variables, or values obtained from different random variables. For example, suppose that you use your favorite coin to gamble and that you report having obtained 5 Heads. Is this value more indicative of ‘good luck’ than the value of 4 Heads obtained by your friend Aïoli who is playing with another coin? It is impossible to answer this question because the number of trials for each case has not been provided, the type of coins (i.e. biased or fair) is not known, etc. Essentially, the lack of a common scale makes it impossible to compare these two values. Therefore, in order to compare two random variables, we need to express them on a common scale. Recall that the mean and the variance of a random variable characterize this variable. Consequently, to be able to compare two different variables, we need to transform them such that they will have the same mean and the same variance. There are, indeed, many possible choices for the new mean and variance. But, as long as we can choose, the simplest numbers will be the best. For the mean, zero is a very practical number because, then, we will know automatically that a *negative* number is *smaller* than the mean, and that a *positive* number is *larger* than the mean. For the variance, we know that it *must* be a positive number (why?¹). The ‘nicest’ positive number is clearly one. So we want the variance of the transformed variables to be equal to one (hence the standard deviation of the transformed variable will also be equal to one because $1 = \sqrt{1}$). We call this transformation *standardizing* or *Z-score transformation*. The precise definition follows.

Definition D.5 (Z-scores, standardized variables). *If Y is a random variable with mean μ_Y and variance σ_Y^2 , then it is transformed into a Z-score, or standardized as:*

$$Z = \frac{Y - \mu_Y}{\sigma_Y} . \quad (\text{D.18})$$

The standardized random variable Z has zero mean and unitary variance (i.e. $\mu_Z = 0$ and $\sigma_Z^2 = 1$, the proof is given in Appendix A, Section A.6.1, page 422.)

¹ Because it is a weighted sum of *squared* deviations, and squared numbers are *always* positive.

Values of W	-10	-20	10	40
$\Pr(W)$	.125	.375	.375	.125
$\Pr(Z) = \Pr(W)$	.125	.375	.375	.125
Values of Z_W	$-\frac{1}{2}$	-1	$\frac{1}{2}$	2
$\Pr(Z_W) \times Z_W$	-.0625	-.375	.1875	.25
Z_W^2	$\frac{1}{4}$	1	$\frac{1}{4}$	4
$\Pr(Z_W) \times Z_W^2$	.03125	.375	.09375	.5

Table D.4 Probability distribution of the random variable W . $\mu_W = 0$, $\sigma_W = 20$. $Z_W = \frac{W - \mu_W}{\sigma_W} = \frac{W}{20}$. Quantities showing that $\mu(Z) = 0$ and $\sigma_Z^2 = 1$ are detailed. Please note that $\Pr(Z) = \Pr(W)$ (see text for explanation).

Note that the probability function of W and Z_W is the same (i.e. it has the same shape). It is the values of the random variables that change. Actually, the Z -transformation *substitutes* standardized values for raw values as Table D.4 makes clear. Precisely, we have the following important relationship:

$$\Pr(Y_n) = \Pr(Z_{Y_n}) \quad \text{for all } n. \quad (\text{D.19})$$

As an exercise, we will use the random variable W described in Table D.4. Recall that $\mu_W = 0$ and $\sigma_W = 20$ (we will omit the dollar sign for simplicity). The random variable takes the values

$$\{-10, -20, 10, 40\}$$

with respective probabilities of

$$\{.125, .375, .375, .125\}.$$

Table D.4 gives the different quantities needed to compute the Z -scores and to check that their mean is zero and their variance one. Using Equation D.18, we compute the following Z -scores:

- for $W_1 = -10$

$$Z_{W_1} = \frac{W_1 - \mu_W}{\sigma_W} = \frac{-10 - 0}{20} = -\frac{1}{2} \quad (\text{D.20})$$

- for $W_2 = -20$

$$Z_{W_2} = \frac{W_2 - \mu_W}{\sigma_W} = \frac{-20 - 0}{20} = -1 \quad (\text{D.21})$$

- for $W_3 = 10$

$$Z_{W_3} = \frac{W_3 - \mu_W}{\sigma_W} = \frac{10 - 0}{20} = \frac{1}{2} \quad (\text{D.22})$$

- for $W_4 = 40$

$$Z_{W_4} = \frac{W_4 - \mu_W}{\sigma_W} = \frac{40 - 0}{20} = 2. \quad (\text{D.23})$$

For simplicity, in what follows we will denote by Z_n what we should, in fact, denote Z_{W_n} .

We can now check that the mean of the random variable W is zero and that its variance is one. For the mean of Z_W (cf. Definition D.3) we obtain:

$$\begin{aligned}
 \mu_Z &= \sum_n^N \Pr(Z_n) \times Z_n \\
 &= \left(.125 \times -\frac{1}{2} \right) + (.375 \times -1) + \left(.375 \times \frac{1}{2} \right) + (.125 \times 2) \\
 &= -.0625 - .375 + .1875 + .25 \\
 &= 0 .
 \end{aligned} \tag{D.24}$$

For the variance of Z_W we obtain (cf. Definition D.4):

$$\begin{aligned}
 \sigma_Z^2 &= \sum_n^N \Pr(Z_n) \times (Z_n - \mu_Z)^2 \\
 &= \sum_n^N \Pr(Z_n) \times Z_n^2 \quad (\text{because } \mu_Z = 0) \\
 &= \left[.125 \left(-\frac{1}{2} \right)^2 \right] + [.375 (-1)^2] + \left[.375 \left(\frac{1}{2} \right)^2 \right] + [.125 (2)^2] \\
 &= \left(.125 \times \frac{1}{4} \right) + (.375 \times 1) + \left(.375 \times \frac{1}{4} \right) + (.125 \times 4) \\
 &= .03125 + .375 + .09375 + .500 \\
 &= 1 .
 \end{aligned} \tag{D.25}$$

D.7 Probability associated with an event

The Z -transformation in itself is a very useful tool to compare different random variables because it re-scales their values. Therefore, the Z transformation is useful to compare specific values of different random variables. However, it cannot be used to compare *probabilities*. Let us see why first, and then try to find an index useful for comparing probabilities of events.

Suppose now, that we want to be able to compare not only values, but also probabilities assigned to values. If we do so, we run into a problem, namely that the probability of obtaining a given value depends upon the number of values of the random variable. For example, if we toss one fair coin, the probability of obtaining one Head is $\frac{1}{2}$. If we toss three coins, the probability of obtaining one Head is now $\frac{3}{8}$, if we toss four coins, the probability of obtaining 1 Head is now $\frac{1}{4}$. In general, the larger the number of values of the random variable, the smaller the probability of any given value will be, on the average. Consequently, we cannot compare probabilities obtained with random variables having different number of values.

In order to eliminate this problem, the solution is to compare the probability of obtaining an event or any event *more extreme* than this event. By more extreme we mean that:

- For an event below the mean (i.e. with a negative Z -score), events more extreme correspond to all the values of the random variable less than or equal to this event.

- For an event above the mean (i.e. with a positive Z -score), events more extreme correspond to all the values of the random variable greater than or equal to this event.
- When we consider only the *magnitude* (i.e. the absolute value) of an event, the events more extreme correspond to all the events whose absolute values are greater than or equal to this event.

The probability of obtaining the values more extreme than a given event is generally referred to as the *probability associated* with this event. Previously we have calculated probabilities assigned to specific outcomes of a certain number of coin tosses, for example, the probability of obtaining exactly three heads when five coins are tossed. We could also calculate the probability of obtaining *three heads or more* when five coins are tossed. This probability is obtained by simply adding the probability of obtaining three, four, and five heads. It is an example of the probability associated with the event of obtaining three heads (an event above the mean). In a like manner, we could calculate the probability of obtaining *two or fewer heads* when tossing five coins, by summing the probabilities for zero, one, and two heads. This is an example of the probability associated with the event of obtaining two heads (an event below the mean).

This concept is formalized in the following definition.

Definition D.6 (Probability associated with A : $p(A)$, p). *If A represents an event corresponding to a specific value of the random variable Y , then the probability associated with A is the sum of the probability of the events more extreme than A . It corresponds to the probability of obtaining A or any event more extreme than A . The probability associated with A is denoted $p(A)$, or, when the context is clear, p .*

If $Z_A < 0$ (i.e. A is below the mean of Y):

$$p(A) = \Pr(Y \leq A) \quad (\text{D.26})$$

[$p(A)$ is the sum of the probability of the values of the random variable smaller than or equal to A].

If $Z_A > 0$ (i.e. A is above the mean of Y):

$$p(A) = \Pr(Y \geq A) \quad (\text{D.27})$$

[$p(A)$ is the sum of the probability of the values of the random variable larger than or equal to A].

If $Z_A = 0$ (i.e. A is equal to the mean of Y):

$$p(A) = .5 \quad (\text{D.28})$$

(the probability associated with the mean is .50).

The probability associated with the absolute value of A is the sum of the probabilities associated with $|A|$ and $-|A|$:

$$\text{absolute } \{p(A)\} = p(-Z_A) + p(+Z_A). \quad (\text{D.29})$$

For example, if we toss a coin three times, and if we count the number of Heads, we have already seen that $\mu_Y = 1.5$, and the probability associated with $Y = 2$ (i.e. obtaining 2 Heads) is equal to the probability of obtaining 2 Heads or more than 2 Heads. This is equal to:

$$p(Y = 2) = \Pr(Y \geq 2) = \Pr(2) + \Pr(3) = .375 + .125 = .5. \quad (\text{D.30})$$

D.8 The binomial distribution

You may have realized, by now, that probability theory seems to be very fond of tossing coins: this is a good model of what should happen by chance when we repeat several times the same elementary experiment (also called a *trial* in this context) having only two possible outcomes. For another example, if we ask 20 persons if they prefer the color Blue or the color Red, we have 20 repetitions of the same basic experiment with two possible answers {Red, Blue}. The outcome of each experiment is said to be *binary* or *dichotomous*. A synonym for a binary outcome is a *dichotomy*. The binomial distribution is the probability distribution of the random variable that counts the number of one outcome (i.e. 'Blue' or 'Red') of the dichotomy.

We have, in fact, already met this distribution, when we were looking at the probability distribution of the random variable counting the number of Heads when tossing three coins. In this example, the dichotomy is 'landing on Head or Tail', and each trial is a toss of a coin.

We will explore this distribution a bit deeper for two main reasons. The first reason is to use it as an exercise to practice the important concepts of this Appendix. The second reason for getting acquainted with the binomial distribution is that we will use it to introduce the very important notion of statistical test in a subsequent Appendix.

Consider, for example, the four coins (we have decided to add a new coin for this section). If we toss them, this will correspond to four trials. Because the outcome of any toss cannot influence any other coin, we note that the trials are *mutually independent*. Each trial can take one of two possible values. Therefore, we have

$$2 \times 2 \times 2 \times 2 = 2^4 \text{ different outcomes.} \quad (\text{D.31})$$

Using the tree procedure introduced in Appendix C (Figure C.5, page 439), we can enumerate all the possible outcomes. Figure D.2 shows the results. The outcomes can be gathered in the following display:

$$\begin{array}{l} \text{HHHH HHHT HHHT HHTT} \\ \text{HTHH HTHT HTTH HTTT} \\ \text{TTHH THTT THTH THTT} \\ \text{TTHH TTHT TTTH TTTT} \end{array} \quad (\text{D.32})$$

If we define Y as being the random variable counting the number of Heads, we can rearrange the outcomes of this experiment in order to put together the events corresponding to the same value of Y .

$$\begin{array}{l} Y = 0 \quad \text{TTTT} \\ Y = 1 \quad \text{TTHH TTHT THTT HTTT} \\ Y = 2 \quad \text{TTHH THTH THHT HTTH HTHT HHTT} \\ Y = 3 \quad \text{TTHH HHTH HHHT} \\ Y = 4 \quad \text{HHHH} \end{array} \quad (\text{D.33})$$

From these values, we can derive the probability distribution described by Table D.5, and by Figure D.3.

For a binomial distribution, it is customary to denote by P or $\Pr(A)$ the probability of obtaining the outcome we are counting (i.e. Head) on one trial. We will refer, occasionally, to this outcome as a '*success*'. So, if we obtain 3 Heads out of 5 coins, we will say that we have obtained 3 successes. For fair coins, $\Pr(\text{Head})$ or $P = \frac{1}{2}$. Hence, we could say that, for fair coins, the probability of a success is one half or fifty percent.

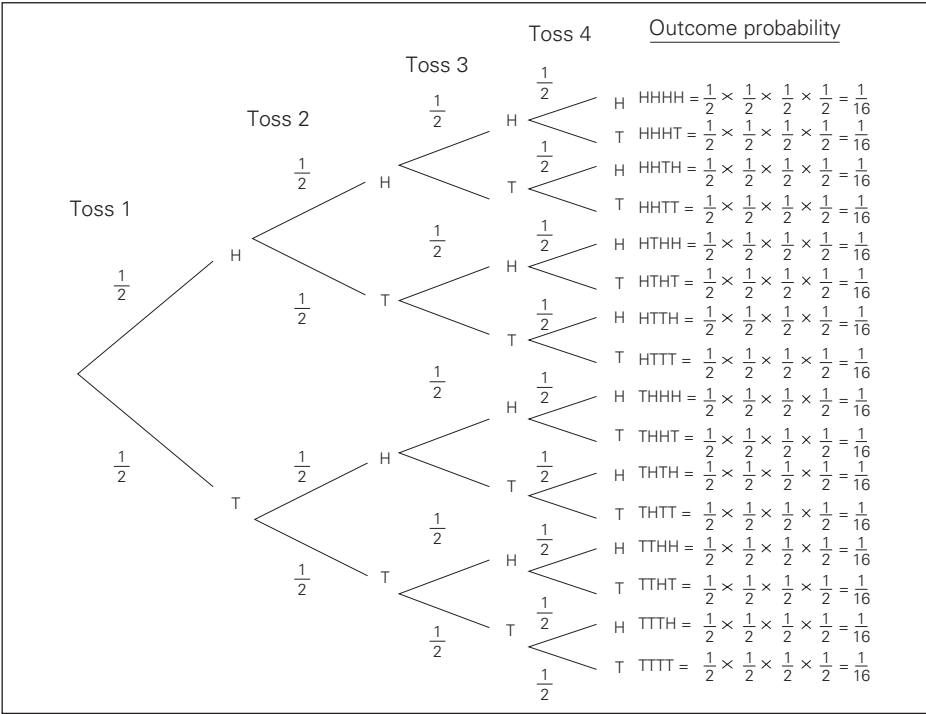


Figure D.2 A tree diagram describing the possible outcomes of tossing four independent coins along with the probability assigned to each outcome. The coins are assumed to be fair coins. The probability of an event represented by a ‘leaf’ of the tree is the product of the probabilities assigned to each of its branches [e.g. $\Pr(HHHH) = \left(\frac{1}{2}\right)^4 = \frac{1}{16}$].

Values of Y	0	1	2	3	4
# {Y}	1	4	6	4	1
Pr(Y)	.0625	.25	.375	.25	.0625

Table D.5 Binomial probability distribution for the random variable $Y = \{\text{Number of Heads}\}$ when tossing four fair coins ($\Pr(\text{Head})$, or $P = \frac{1}{2}$).

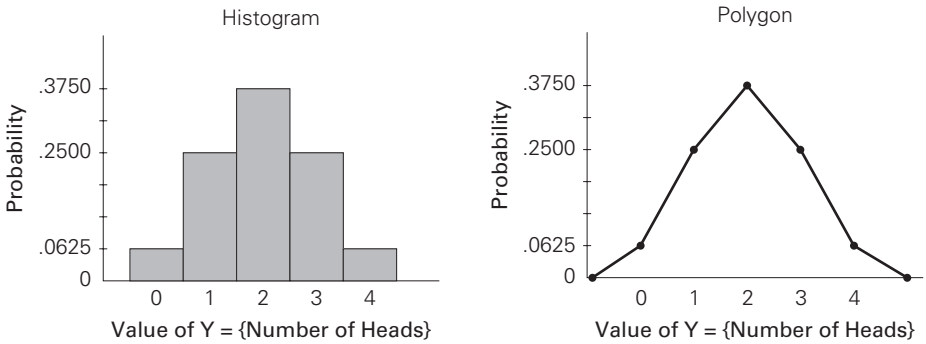


Figure D.3 Histogram and polygon of the probability distribution for the random variable $Y = \{\text{Number of Heads}\}$ when tossing four coins (compare with Table D.5).

Let us start by computing the expected value of the binomial distribution when N , the number of trials, is equal to four. From Table D.5, and formula D.13 we can compute the mean of the binomial distribution:

$$\begin{aligned}
 \mathbf{E}\{Y\} &= \mu_Y \\
 &= \Pr(Y_1) \times Y_1 + \Pr(Y_2) \times Y_2 + \cdots + \Pr(Y_N) \times Y_N \\
 &= \sum_n^N \Pr(Y_n) \times Y_n \\
 &= (.0625 \times 0) + (.25 \times 1) + (.375 \times 2) + (.25 \times 3) + (.0625 \times 4) \\
 &= 0 + .25 + .75 + .75 + .25 \\
 &= 2 .
 \end{aligned} \tag{D.34}$$

So, the expected value of the number of Heads when tossing 4 coins is 2.

A nice property of the binomial distribution is that its mean or expected value can be computed with a simple formula.

Rule D.1 (Expected value of the binomial). *The mean of a binomial distribution of N trials, each of them having a probability of ‘success’ of P , is equal to*

$$\mathbf{E}\{Y\} = \mu_Y = N \times P . \tag{D.35}$$

As a check, we verify that when tossing $N = 4$ coins, each of them having a probability of $P = \frac{1}{2}$ to be landing on Head, the expected value should be

$$\mu_Y = 4 \times \frac{1}{2} = 2 , \tag{D.36}$$

which, indeed, corresponds to the result previously found with Equation D.34.

We now proceed to compute the variance of the binomial distribution when N , the number of trials, is equal to four. From Table D.6, and formula D.15, we can compute the variance of the binomial distribution:

$$\begin{aligned}
 \sigma_Y^2 &= \text{Var}(Y) \\
 &= \Pr(Y_1) \times (Y_1 - \mu_Y)^2 + \Pr(Y_2) \times (Y_2 - \mu_Y)^2 \\
 &\quad + \cdots + \Pr(Y_N) \times (Y_N - \mu_Y)^2 \\
 &= \sum_n^N \Pr(Y_n) \times (Y_n - \mu_Y)^2 \\
 &= [.0625 \times (0 - 2)^2] + [.25 \times (1 - 2)^2] + [.375 \times (2 - 2)^2] \\
 &\quad + [.25 \times (3 - 2)^2] + [.0625 \times (4 - 2)^2] \\
 &= (.0625 \times 4) + (.25 \times 1) + (.375 \times 0) + (.25 \times 1) + (.0625 \times 4) \\
 &= .25 + .25 + 0 + .25 + .25 \\
 &= 1 .
 \end{aligned} \tag{D.37}$$

So, the variance of the number of Heads when tossing 4 coins is 1.

Pr(Y)	.0625	.25	.375	.25	.0625
Y	0	1	2	3	4
$Y - \mu_Y$	-2	-1	0	1	2
$(Y - \mu_Y)^2$	4	1	0	1	4
$\Pr(Y) \times (Y - \mu_Y)^2$	.25	.25	0	.25	.25

Table D.6 Binomial distribution probability. $N = 4$, $P = \frac{1}{2}$, the random variable is $Y = \{\text{Number of Heads}\}$ when tossing four fair coins. Quantities needed to compute the variance of the distribution. $\sigma_Y^2 = \sum \Pr(Y) \times (Y - \mu_Y)^2 = .25 + .25 + 0 + .25 + .25 = N \times \frac{1}{2} \times \frac{1}{2} = 1$.

We have also a nice property of the binomial distribution concerning the variance. It can be computed with a formula much more simple than the general formula.

Rule D.2 (Variance of the binomial). *The variance of a binomial distribution of N trials, each of them having a probability of ‘success’ of P , is equal to*

$$\sigma_Y^2 = \text{Var}(Y) = N \times P \times (1 - P) . \quad (\text{D.38})$$

As a check, we verify that when tossing $N = 4$ coins, each of them having a probability of $P = \frac{1}{2}$ to be landing on Head, the variance should be

$$\sigma_Y^2 = 4 \times \frac{1}{2} \times \left(1 - \frac{1}{2}\right) = 1 , \quad (\text{D.39})$$

which, indeed, corresponds to the result previously found with Equation D.37.

D.9 Computational shortcuts

Finding the complete list of elementary events for the binomial distribution is very simple, at least *in principle*, because the ‘tree-approach’ is always possible. There is, however, a *practical* problem which occurs when N (the number of trials) becomes large (say larger than 10). First of all, because the number of elementary events is a power function (i.e. 2^N) it becomes very large rapidly. As an illustration, the number of elementary events as a function of the number of trials is given hereunder for values of N from 1 to 10:

$$\begin{array}{cccccccccccc} N & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 2^N & 2 & 4 & 8 & 16 & 32 & 64 & 128 & 256 & 512 & 1024 . \end{array} \quad (\text{D.40})$$

Consequently, drawing the tree becomes very cumbersome as N grows (just try to draw the tree for, say, $N = 7$ to have an idea of what we mean).

Fortunately, mathematicians (from Pascal and Bernoulli to Gauss) have derived several computational shortcuts that can make computing probability for the binomial distribution rather easy. These formulas are based on elementary counting rules, and on the probability rules described in Appendix C. Before looking at them, however, we need to introduce two new notations: the *factorial* number and the *binomial* coefficient.

D.9.1 Permutation, factorial, combinations and binomial coefficient

We need to introduce two new notions and notations here.

Definition D.7 (Factorial and number of permutations). *The number of permutations of K elements is also called the factorial number. It is denoted $K!$ (read ‘factorial K ’) and is computed as*

$$K! = 1 \times 2 \times 3 \times \cdots \times (K-1) \times K \quad (\text{D.41})$$

or, equivalently as

$$K! = K \times (K-1) \times \cdots \times 3 \times 2 \times 1. \quad (\text{D.42})$$

Also, by definition $0! = 1$.

The number of permutations of K objects corresponds to the number of ways we can order these K objects. For example, suppose that a test has 5 questions. How many different orders can we create from these 5 questions? We have 5 ways of choosing the first question, 4 ways of choosing the second question (because there are only 4 available questions left *after* the first one has been chosen), 3 ways of choosing the third question, 2 ways of choosing the second question, and finally only one way of choosing the last question (because this is the only one left, by now!). Therefore, the total number of orders, or number of permutations, of $K = 5$ questions is ‘factorial 5’:

$$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120. \quad (\text{D.43})$$

The number of *combinations* of a subset of objects from a larger set, refers to the number of ways we can choose a subset from the set. Suppose that you have a set of 20 questions available and that a test is made of 5 questions. How many different tests can you make assuming that you consider that the order of the questions of the test is irrelevant (i.e. two tests with the same questions in a different order are the same test). To obtain this formula, we first observe that we can choose the first question 20 different ways, the second question 19 different ways, the third question 18 different ways, the fourth question 17 different ways, and, finally, the fifth question 16 different ways. Hence, we obtain

$$20 \times 19 \times 18 \times 17 \times 16 \quad (\text{D.44})$$

different *ordered* tests (by ‘ordered’ we mean that two tests with the same questions in a different order are considered different).

However, with this procedure, a given set of 5 questions in one order is considered different from the same set in a different order. This implies that we are counting the $5!$ tests ($5!$ is the number of different orders of 5 elements) as being different, even though they are the same (only the order of the questions differs). Therefore, if we are not interested in the order of the questions, we need to divide the number of ‘ordered tests’ given in Expression D.44 by the number of permutations of these five questions. Hence, the number of different tests made with 5 questions out of 20 is:

Number of combinations of 5 questions from 20

$$\begin{aligned} &= \frac{20 \times 19 \times 18 \times 17 \times 16}{5 \times 4 \times 3 \times 2 \times 1} \\ &= \frac{19 \times 18 \times 17 \times 16}{3 \times 2 \times 1} \\ &= 19 \times 3 \times 17 \times 16 \\ &= 15,504. \end{aligned} \quad (\text{D.45})$$

We could rewrite this equation into a more convenient form, using the factorial notation, as:

Number of combinations of 5 questions from 20

$$\begin{aligned}
 &= \frac{20!}{5!(20-5)!} \\
 &= \frac{20!}{5! \times 15!} \\
 &= \frac{20 \times 19 \times 18 \times 17 \times 16}{5 \times 4 \times 3 \times 2 \times 1} \\
 &= 15,504.
 \end{aligned} \tag{D.46}$$

If we generalize this example, we obtain the definition of the number of combinations.

Definition D.8 (Number of combinations, binomial coefficient). *The number of combinations of C elements from an ensemble of N elements is denoted $\binom{N}{C}$ (read ‘binomial of N , C ’ or ‘number of combinations of C elements from N elements’) and is computed as*

$$\binom{N}{C} = \frac{N!}{C!(N-C)!} \tag{D.47}$$

(note that this definition assumes that C is always smaller than or equal to N).

If we come back to the binomial probability, we can see that if Y is the random variable ‘counting the number of Heads obtained when tossing 4 coins,’ then the number of events giving C Heads out of $N = 4$ trials is given by the binomial coefficient $\binom{N}{C}$. As an exercise, let us check that we obtain the same result with our new formula as the value that we have obtained in Equation D.33 and Table D.5.

- For $Y = 0$,

$$\binom{N}{C} = \binom{4}{0} = \frac{N!}{C!(N-C)!} = \frac{4!}{0!4!} = 1. \tag{D.48}$$

- For $Y = 1$,

$$\binom{N}{C} = \binom{4}{1} = \frac{N!}{C!(N-C)!} = \frac{4!}{1!3!} = 4. \tag{D.49}$$

- For $Y = 2$,

$$\binom{N}{C} = \binom{4}{2} = \frac{N!}{C!(N-C)!} = \frac{4!}{2!2!} = 6. \tag{D.50}$$

- For $Y = 3$,

$$Y = 3: \quad \binom{N}{C} = \binom{4}{3} = \frac{N!}{C!(N-C)!} = \frac{4!}{3!1!} = 4. \tag{D.51}$$

- For $Y = 4$,

$$Y = 4: \quad \binom{N}{C} = \binom{4}{4} = \frac{N!}{C!(N-C)!} = \frac{4!}{4!0!} = 1. \tag{D.52}$$

Note that if there are $Y = C$ ‘successes’, each of them having the probability P , then there are also $(N - C)$ failures, each of them having the probability $(1 - P)$ (as a trial gives either a success or a failure). Because all these elementary events (i.e. ‘successes’ and ‘failures’) are *independent*, we can use the *product rule* of Appendix C that states that the probability of

$A \text{ AND } B \text{ AND } \dots Z$ is equal to the product of their respective probabilities (when the events are mutually independent). Therefore, the probability of obtaining C ‘successes’ each with a probability of P AND $(N - C)$ ‘failures’ each of them with a probability of $(1 - P)$ is

$$\begin{aligned} & \Pr(C \text{ successes AND } [N - C] \text{ failures}) \\ &= \underbrace{P \times P \times \dots \times P}_{C \text{ times}} \times \underbrace{(1 - P) \times (1 - P) \times \dots \times (1 - P)}_{(N - C) \text{ times}} \\ &= P^C \times (1 - P)^{(N - C)}. \end{aligned} \quad (\text{D.53})$$

We have seen before that there are $\binom{N}{C}$ ways of obtaining the event ‘obtaining C successes out of N trials’. Using now, the addition rule of Appendix C, and taking into account that each of these events has the same probability of $P^C \times (1 - P)^{(N - C)}$ that we have just computed, we find that the probability of obtaining C ‘successes’ out of N trials is

$$\Pr(Y = C) = \binom{N}{C} \times P^C \times (1 - P)^{N - C}. \quad (\text{D.54})$$

As an exercise, we are now going to compute the complete probability distribution for the random variable counting the number of Heads when tossing 4 coins.

- For $Y = 0$,

$$\begin{aligned} \Pr(Y = 0) &= \binom{N}{C} \times P^C \times (1 - P)^{N - C} \\ &= \binom{4}{0} P^0 (1 - P)^{4 - 0} \\ &= 1 \times .5^0 \times (1 - .5)^4 = 1 \times .5^4 = .0625. \end{aligned} \quad (\text{D.55})$$

- For $Y = 1$,

$$\begin{aligned} \Pr(Y = 1) &= \binom{N}{C} \times P^C \times (1 - P)^{N - C} \\ &= \binom{4}{1} P^1 (1 - P)^{4 - 1} \\ &= 4 \times .5^1 \times (1 - .5)^3 = 4 \times .5^4 = .2500. \end{aligned} \quad (\text{D.56})$$

- For $Y = 2$,

$$\begin{aligned} \Pr(Y = 2) &= \binom{N}{C} \times P^C \times (1 - P)^{N - C} \\ &= \binom{4}{2} P^2 (1 - P)^{4 - 2} \\ &= 6 \times .5^2 \times (1 - .5)^2 = 6 \times .5^4 = .3750. \end{aligned} \quad (\text{D.57})$$

- For $Y = 3$,

$$\begin{aligned} \Pr(Y = 3) &= \binom{N}{C} \times P^C \times (1 - P)^{N - C} \\ &= \binom{4}{3} P^3 (1 - P)^{4 - 3} \\ &= 4 \times .5^3 \times (1 - .5)^1 = 4 \times .5^4 = .2500. \end{aligned} \quad (\text{D.58})$$

- For $Y = 4$,

$$\begin{aligned}
 \Pr(Y = 4) &= \binom{N}{C} \times P^C \times (1 - P)^{N-C} \\
 &= \binom{4}{4} P^4 (1 - P)^{4-4} \\
 &= 1 \times .5^4 \times (1 - .5)^0 = 1 \times .5^4 = .0625.
 \end{aligned}
 \tag{D.59}$$

Indeed, this result agrees with what we previously found.

D.10 The 'normal approximation'

The previous section shows how it is possible to use the binomial coefficients to make computing probabilities for the binomial distribution easier than computing by using the standard tree procedure. Despite a clear improvement, however, even the formula using the binomial coefficient can become cumbersome when N becomes large (say larger than 50). To make this point clearer, suppose that you are asked on a test to compute the probability associated with $Y = 80$ (e.g. the number of Heads) when $N = 200$ (number of tosses of the coin). This means that you have to compute 81 binomial coefficients (one from each value from 0 to 80). Even though these coefficients are rather straightforward to compute, computing that many of them may not be the most thrilling intellectual experience.

As usual, our mathematician friends (de Moivre in the 18th century, and Gauss in the 19th century) have thought about a solution. Their idea starts with the observation that the shapes of the different binomial distributions corresponding to different values of N are rather similar (this is especially true when P is close to $\frac{1}{2}$). Consider Figure D.4, which shows several binomial distributions for values of N from 5 to 100. The larger N becomes, the smoother the binomial distribution becomes. At the same time, the smaller the probability of finding any *specific* value becomes [which justifies the use of the *associated* probability $p(A)$ rather than the assigned probability, $\Pr(A)$].

De Moivre and Gauss showed that as N increases, the binomial gets closer to a bell-shaped distribution which we call, nowadays the *Gaussian* distribution or also the *normal* distribution. Like all probability distributions, the normal distribution can be rescaled or standardized. It is customary in psychology to consider only the so-called *standardized normal distribution*.

The shape of the normal distribution is given in Figure D.5. If you want to know more about its characteristics you should read the following section. If you feel satisfied with knowing that there is an equation describing this distribution, you may skip it.

D.10.1 Digression: the equation of the normal distribution

Because the normal distribution is continuous, the ordinate is called $f(X)$, which is a function of X , instead of $\Pr(X)$ (this is just a technicality without practical importance, here). Let us look first at the general formula of the normal distribution (i.e. before we re-scale it), and then we will examine the simplified formula for the standardized normal distribution.

The normal (also known as Gaussian) distribution is defined as

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} \times e^{-\frac{(X-\mu)^2}{2\sigma^2}}
 \tag{D.60}$$

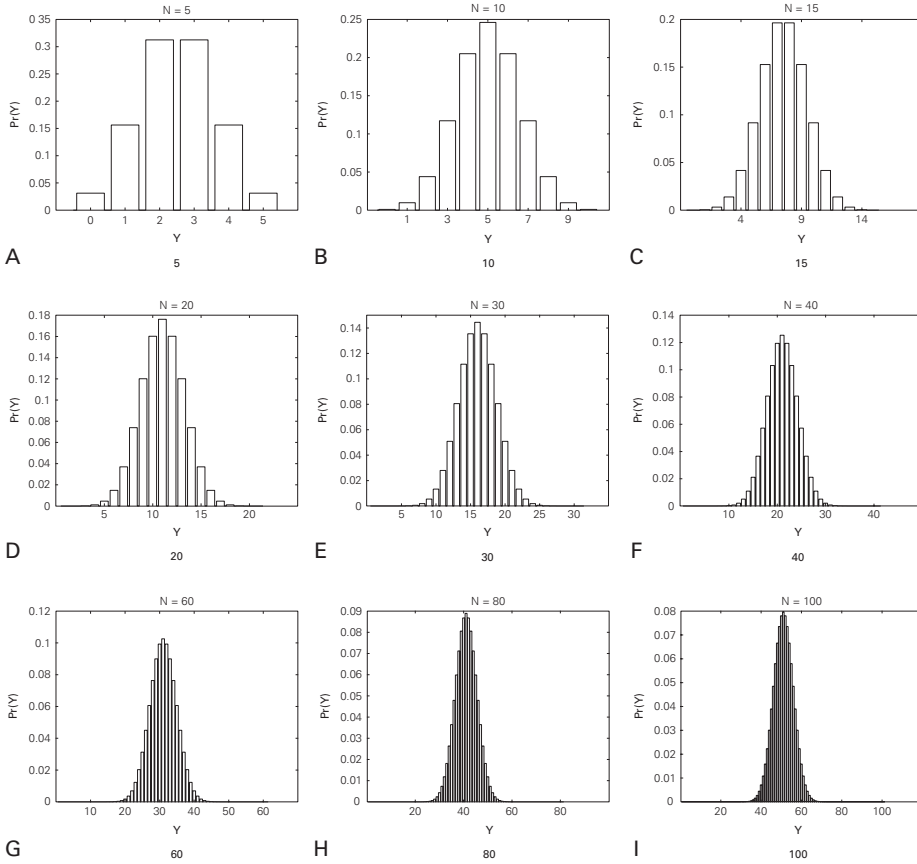


Figure D.4 Some binomial distributions, with $P = \frac{1}{2}$, and N varying from 5 to 100. As N becomes larger, the binomial distribution becomes smoother and looks more and more like the *Normal* distribution (see text for explanation).

where π is the usual $\pi = 3.14159\dots$, e is the so-called *Euler*² number (the basis of the 'natural logarithms', $e = 2.719\dots$), μ is the mean of the distribution and σ^2 is the variance of the distribution. Because we do not change the *shape* of a probability distribution, we can consider the simplified formula when the normal distribution is standardized. Recall that we use the notation Z for a random variable with mean $\mu_Z = 0$, and standard deviation $\sigma_Z = 1$. Substituting Z for X in Equation D.60, and simplifying for $\mu_Z = 0$, and $\sigma_Z = 1$, we find that the standardized normal distribution is defined as

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}}. \quad (\text{D.61})$$

(We use Z because it is the equation of the *standardized* normal distribution.) At first glance, if you are not familiar with calculus, this equation may look formidable. With a closer look, we see that some of its important properties can be derived by inspection. Clearly, the term $\frac{1}{\sqrt{2\pi}}$ is a scaling factor (we need to have an area of 1 for a probability distribution). The number e

² Joseph Euler (1707–1783), born in Switzerland, was a superb mathematician.

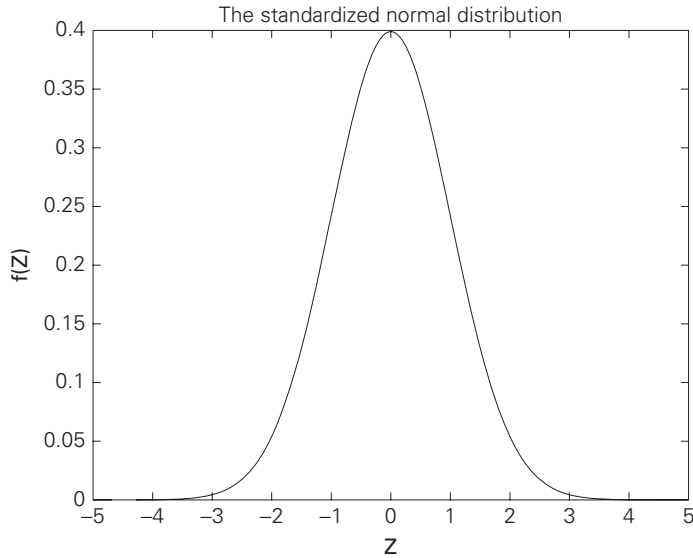


Figure D.5 A picture of the Gaussian distribution, also known as the normal distribution. The distribution is standardized (i.e. $\mu = 0$, $\sigma^2 = 1$).

is a positive constant, so most of the shape of the distribution will be created by the (negative) power term: $-\frac{1}{2}Z^2$. Because it involves a squared term, this implies that the distribution should be symmetrical, and that it should reach its highest value for $Z = 0$.

D.11 How to use the normal distribution

The important characteristic of *the* standardized normal distribution is that there is only *one* distribution (as opposed to an infinity of different binomial distributions). All the standardized binomial distributions will converge towards the normal distribution. We will say also that the normal distribution *approximates* most binomial distributions. When $P = \frac{1}{2}$, for example, the approximation is already fairly good when $N = 10$. The larger N , the better the approximation becomes.

This means that a very *large number* of binomial distributions can be approximated by *only one* distribution. This may explain the popularity of the normal distribution.

Table D.6 shows the graph of the normal distribution superimposed on the binomial distribution when $N = 10$. The normal distribution and the binomial distribution drawn here have the same mean ($\mu = N \times P = 10 \times .5 = 5$) and same variance [$\sigma^2 = N \times P \times (1 - P) = 10 \times .5 \times .5 = 2.5$]. As you can see, the normal distribution is very close to the binomial and seems to be a smoothed version of the binomial distribution.

Take a look at Figure D.7, and suppose that we want to use the normal distribution to approximate the binomial probability associated with the value 3. As you can see on the graph of Figure D.7, the value of 3 for the binomial distribution is the center of a bar that goes from 2.5 to 3.5. Therefore, the probability associated with 3 [$p(3)$], corresponds to the area *on the left of the value 3.5*. Consequently, if we want to approximate the discrete binomial distribution by

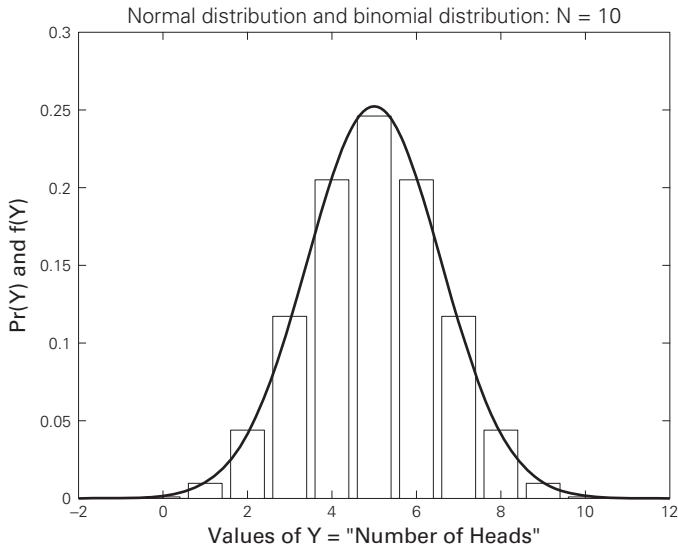


Figure D.6 The binomial distribution for $N = 10$, and $P = \frac{1}{2}$; $\mu_Y = 5$, $\sigma_Y^2 = 2.5$. The normal distribution with same mean and variance is superimposed on the binomial distribution.

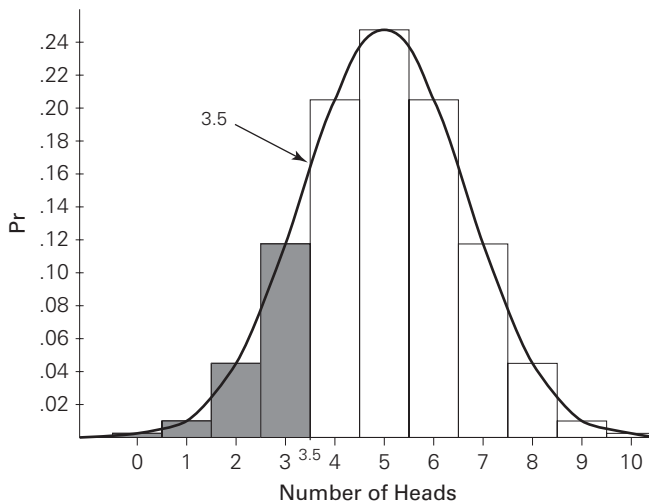


Figure D.7 The binomial distribution for $N = 10$, and $P = \frac{1}{2}$; $\mu_Y = 5$, $\sigma_Y^2 = 2.5$. The normal distribution with same mean and variance is superimposed on the binomial distribution. A value of 3 of the binomial distribution corresponds to a value of 3.5 for the normal distribution.

the continuous³ normal distribution we need to use a value of 3.5 for the normal distribution. In other words, to find the probability associated with a value of Y smaller than the mean μ_Y , we *add* a value of .5 to this score. If the value Y for which we search $p(Y)$ is *larger* than

³ In statistics and mathematics, the opposite of discrete is not indiscrete, or even indiscreet, but continuous!

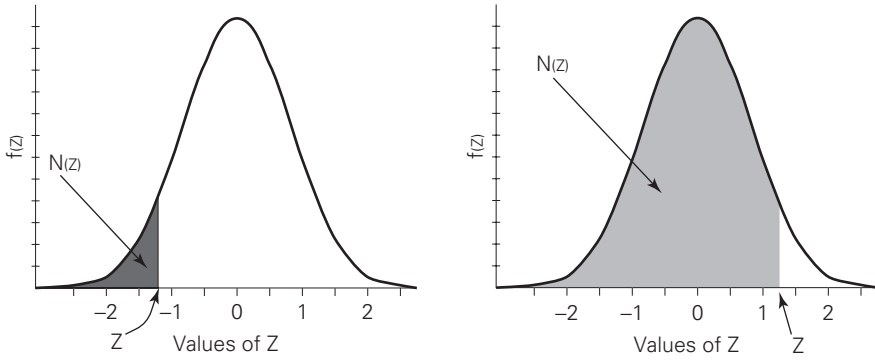


Figure D.8 The table of the normal distribution gives Z (values of the abscissa) and $\mathcal{N}(Z)$ [shaded area corresponding to the probability $\mathcal{N}(Z) = \Pr(\text{finding a value equal to or less than } Z)$]. Note that $\mathcal{N}(Z)$ gives the probability associated with Z only when $Z \leq 0$. Otherwise, when $Z > 0$, then the probability associated with Z is computed as $p(Z) = 1 - \mathcal{N}(Z)$.

the mean μ_Y , then we *subtract* the value .5 from this score (in case of doubt, draw a graph and look). The fact of using 3.5 for the normal when approximating a value of 3 is called the *correction for continuity*.

The next step is to transform the score of 3.5 into a Z -score. In order to do so, we use Equation D.18:

$$Z = \frac{Y - \mu_Y}{\sigma_Y} = \frac{3.5 - 5}{\sqrt{2.5}} \approx -.95. \quad (\text{D.62})$$

And now, what do we do with this value? As it happens, the values of Z , along with the probability associated with them, can be derived from tables. There are several ways of presenting those tables and you will find all possible variations in the literature.

The table of the normal distribution in the Appendix (see Table 1, p. 497) gives for any value of Z the probability corresponding to an equal or *smaller value* of Z (this is equivalent to the probability associated with Z , *only for Z negative!*). The area on the left of Z is called $\mathcal{N}(Z)$. Figure D.8 shows the relationship between Z and $\mathcal{N}(Z)$. If we are only interested in the probability associated with Z , it is sufficient always to look in the table for $-|Z|$. But we will see, later on, that this table has several other functions.

Looking at the table, we find that for $Z = -.95$, $p(Z) = \mathcal{N}(Z) = .171$. So according to the normal approximation, the probability associated with 3 Heads is

$$p(Y = 3) \approx p(Z = -.95) = \mathcal{N}(-.95) = .171. \quad (\text{D.63})$$

How does the ‘normal approximation’ compare with computing the exact values? As an exercise (that should convince you of the interest of using such an easy approximation), let us try to compute the probability associated with $Y = 3$ by using the binomial coefficients. We find that

$$\begin{aligned} p(Y = 3) &= \Pr(0) + \Pr(1) + \Pr(2) + \Pr(3) \\ &= \binom{10}{0} \times P^0 \times (1 - P)^{10} + \binom{10}{1} \times P^1 \times (1 - P)^9 \\ &\quad + \binom{10}{2} \times P^2 \times (1 - P)^8 + \binom{10}{3} \times P^3 \times (1 - P)^7 \end{aligned}$$

$$\begin{aligned}
 &= .0009765625 + .009765625 + .0439453125 + .1171875 \\
 &\approx .1719,
 \end{aligned}
 \tag{D.64}$$

which is indeed very close to the approximated value of $p(Y=3) \approx .171$ given by Equation D.63.

D.12 Computers and Monte-Carlo

A lot of the properties of statistics and probability have to deal with repeating experiments a large number of times. For example, when we mention that the probability of obtaining one Head when flipping one coin is $\frac{1}{2}$ we mean that, *in the long run*, the proportion of Heads should be $\frac{1}{2}$. One way of verifying that assumption could be to toss a coin for a long time. For example, Kerrich (1948) tossed a coin 10,000 times and recorded the outcome of each experiment. He reported that the proportion of Heads that he found was .5067. The great advantage of computers is that they are ideally suited for these repetitive tasks. For example, it takes barely 5 minutes for a modern computer to simulate the flipping of 10,000 coins, record the proportion of Heads, and plot a graph as in Figure D.9.

Because of their speed (and their patience, or some people would say stupidity) computers can be used to *estimate* or *approximate* probabilities. This approach is getting more and more popular, and is known under the name of the *Monte-Carlo* procedure.⁴

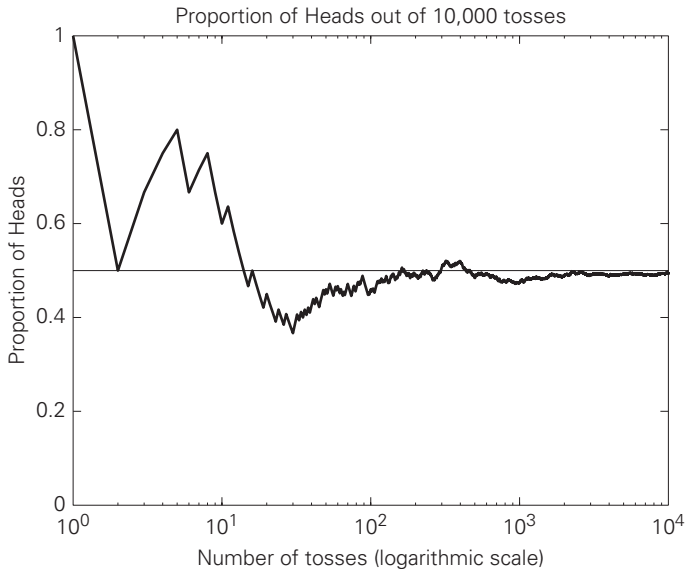


Figure D.9 The proportion of Heads when tossing a fair coin 10,000 times. As the number of trials increases, the proportion of coins gets closer and closer to its expected value of $\Pr(\text{Head}) = \frac{1}{2}$.

⁴ Once again, the connection between probability and gambling is to be blamed for the name. Monte Carlo is a well-known place for gambling in Europe.

The general idea behind a Monte-Carlo procedure is to *simulate* the experiment for which we want to estimate some probabilities, and then run a very large number of simulations in order to derive the probability distribution empirically. An obvious advantage of this method is to be able to handle cases for which we do not know how to compute the probability distribution. As an illustration, we will approximate the probability distribution of the random variable $Y = \text{number of Heads when tossing 3 coins}$. Indeed, in this case, we do know the correct distribution, so we will check that the results of both methods concur.

To create a Monte-Carlo estimation of the binomial distribution, we just need 10 coins (even one coin will do, if we repeat 10 times the tossing). We toss these 10 coins and record the value of Y (i.e. we count the number of Heads). We repeat this experiment once again and record again the value of Y . We do it again, and again ... Let's say that we repeat this basic experiment 10,000 times. Or, unless you really enjoy this type of activity, we can ask a computer to do it for us.⁵ At the end of the series of experiments we have an

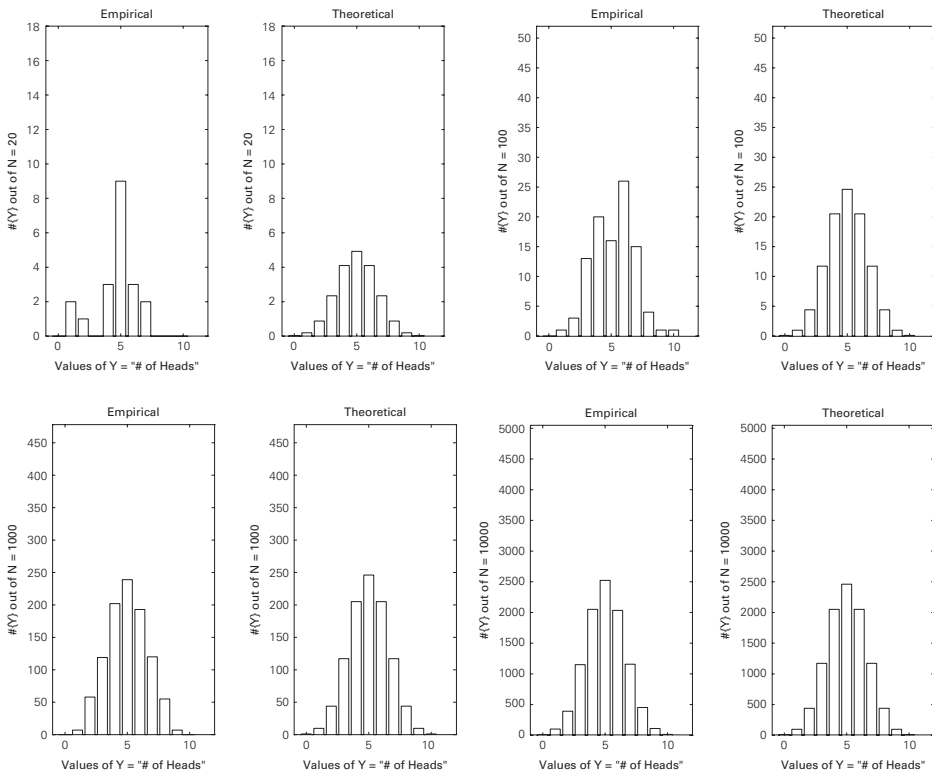


Figure D.10 Results of some Monte-Carlo simulations. The different panels show the frequency distributions of the number of Heads when tossing 10 coins. For each panel, the empirical and the theoretical distributions are shown side by side with the same scale. The number of experiments is (a) top left panel: 20; (b) top right panel: 100; (c) bottom left panel: 1000; (d) bottom right panel: 10,000. The Monte-Carlo method distribution is very similar to the theoretical binomial distribution, even with a relatively small number of experiments. For a number of experiments of 10,000, the two distributions are virtually indistinguishable.

⁵ Most modern computer languages have built-in functions that simulate tossing a coin, or drawing random numbers.

empirical distribution. For each possible value of Y (i.e. from zero to 10 Heads) we have recorded the number of times we have observed that event. So, we know the number of times we have obtained 0 Heads, 1 Head, dots, 10 Heads. This gives a frequency distribution. If we divide each of the values of the frequency distribution by the number of trials, we obtain an estimation of the probability distribution of interest.

The results of some Monte-Carlo simulations are given in Figure D.10. We have varied the number of the series of Monte-Carlo experiments to show the empirical *law of large numbers*. As the number of trials increases, the empirical Monte-Carlo distribution gets closer and closer to the theoretical distribution. (To make the graphs easier to label, we have plotted the frequency distribution, but the probability distributions will appear identical.)

So, supposing that we cannot recall the formula for the binomial distribution, we can derive an excellent approximation of it using computer simulations. Actually, we strongly suspect that if the problems that gave rise to probability theory had arisen in our computer age, most of the theory would be the Monte-Carlo approach! This approach is increasingly popular since computers are increasingly becoming faster and cheaper. An additional advantage of the Monte-Carlo approach is its applicability to cases where the traditional mathematical approach fails to specify probability distributions analytically. So you can expect to see more and more of the Monte-Carlo methodology in applied statistics.

Chapter summary

D.13 Key notions of the chapter

Below are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

Random variable	Permutation
Probability distribution	Factorial
Expected value or mathematical expectation	Combinations
Mean and variance of a probability distribution	Binomial coefficient
Binomial distribution	Z-score
	Normal approximation
	Monte-Carlo approach

D.14 New notations

Below are the new notations introduced in this chapter. Test yourself on their meaning.

Random variable: Y	Factorial K : $K!$
Expected value of Y : $E\{\}$	Binomial coefficient of N , C : $\binom{N}{C}$

D.15 Key formulas of the chapter

Below are the main formulas introduced in this chapter: try to go through them and understand what they mean.

Expected value (or mathematical expectation) of Y

$$\begin{aligned}\mathbf{E}\{Y\} &= \mu_Y \\ &= \Pr(Y_1) \times Y_1 + \Pr(Y_2) \times Y_2 + \cdots + \Pr(Y_N) \times Y_N \\ &= \sum_n^N \Pr(Y_n) \times Y_n\end{aligned}$$

Variance of Y

$$\begin{aligned}\sigma_Y^2 = \text{Var}(Y) &= \mathbf{E}\{(Y - \mu_Y)^2\} \\ &= \Pr(Y_1) \times (Y_1 - \mu_Y)^2 + \Pr(Y_2) \times (Y_2 - \mu_Y)^2 \\ &\quad + \cdots + \Pr(Y_N) \times (Y_N - \mu_Y)^2 \\ &= \sum_n^N \Pr(Y_n) \times (Y_n - \mu_Y)^2\end{aligned}$$

Expected value of the binomial

$$\mathbf{E}\{Y\} = \mu_Y = N \times P$$

Variance of the binomial

$$\sigma_Y^2 = \text{Var}(Y) = N \times P \times (1 - P)$$

Factorial

$$K! = K \times (K - 1) \times \cdots \times 3 \times 2 \times 1$$

Binomial coefficient

$$\binom{N}{C} = \frac{N!}{C!(N - C)!}$$

The probability of obtaining C ‘successes’, each of them with a probability of P , AND $(N - C)$ ‘failures’, each of them with a probability of $(1 - P)$, is

$$\begin{aligned}\text{Pr}(C \text{ successes AND } [N - C] \text{ failures}) \\ &= \underbrace{P \times P \times \cdots \times P}_C \times \underbrace{(1 - P) \times (1 - P) \times \cdots \times (1 - P)}_{(N - C) \text{ times}} \quad (\text{D.65}) \\ &= P^C \times (1 - P)^{(N - C)}.\end{aligned}$$

D.16 Key questions of the chapter

Opposite are some questions about the content of this chapter. All the answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- * Why is the notion of expected value linked to gambling?
- * What does the expected value of a random variable reveal about the outcome of a long series of repeated experiments?
- * What is the information that the variance gives and that the mean of the random variable does not give?
- * Why is it useful to standardize variables?
- * Why is it interesting to compute the probability *associated* with an event? How does it compare with the probability *assigned* to an event?
- * What is the sign of the Z -score corresponding to a score below the mean?
- * What is the probability of finding a Z -score smaller than 2 with a standardized normal distribution?
- * Why do we prefer using the binomial coefficients to compute probabilities with the binomial rather than the tree approach?
- * What are the pros and cons of the Monte-Carlo approach, as opposed to the classical approach of computing the probability distribution?
- * What do we mean by saying that the binomial distribution converges towards the normal distribution as N increases?
- * What is the difference between $\mathcal{N}(Z)$ and $p(Z)$?
- * What is the difference between tossing 10,000 times 20 coins yourself and having the computer doing it? More seriously, what is the difference between tossing 10,000 times 20 coins and recording the frequency of each event, and computing the binomial probability distribution?

E

The binomial test

E.1 Introduction

In this appendix we describe a statistical test based on the binomial probability distribution. We also introduce the important notion of Monte-Carlo procedures and we show how these procedures can be used to derive sampling distributions.

E.2 Measurement and variability in psychology

Measurement, especially measurement in psychology, is not quite as simple as one might at first suppose. The notion that first springs to mind is simply that if you want data, you just go out and measure things. It should be possible to just go and make observations independent of any theory. In fact, how can you test a theory without independent observations to test it with? However, it is easy to claim here that *no* observations are independent of theory. This is true not only because at least some naive, common-sense theory is necessary to guide observations, to *decide* what to observe, but also because we collect data in order to *support*, *defend*, or *attack* a point of view. The scientific process is essentially a *controversial* one, and no scientists can waste their time collecting data without a purpose. In general, we collect data to show that either a theory is able to predict some hitherto unknown behavior, or that some theory is unable to predict some effect. We use the first technique to find support for a theory, the second one to show that a theory needs to be amended or discarded.

When we measure behavior in psychology—whether the measurement involves questionnaire responses, number of correct answers on a recognition test, response times to target stimuli, number of responses before the extinction of a habit, etc.—not all of the subjects will give the same score, and the same subject will give different scores on different occasions. Subjects are to some degree inconsistent, and their data are what we would call ‘noisy’.

Because of the controversial aspect of science, and because of the noisy aspect of data from life and human sciences, the first point to ascertain is the very existence of the pattern that we think exists. Indeed, it is always possible that some pattern that we believe we discern in our data could, in fact, be an illusion. In general, colleagues who disagree with our pet theories (and favor others) will be very eager to dismiss our results as being an illusion. Therefore, the first thing to prove is that our results *cannot* be attributed to chance factors.

Also, since the aim of psychology is to uncover consistencies in behavior and try to understand them, we need some way to identify those consistencies when they occur. In other words, we need some means of determining when the subjects are behaving in a systematically consistent manner, and when they are behaving inconsistently or randomly.

One way of doing that is by means of a *statistical test* that compares the subject's behavior with random behavior, and assesses the likelihood that the subject is behaving randomly. To the degree that this possibility is unlikely, we can be confident that the subject is behaving consistently. So, essentially, in order to be able to show that an effect exists, we want to show that it is very *unlikely* that our results could be due to chance. This procedure is called a statistical test.

The procedure behind most statistical tests is essentially the same. In this appendix, we will introduce a test appropriate when we collect *binary* measures (e.g. 'Yes-No', 'Girls-Boys'). The goal of the test is to assess if we can believe that processes are responsible for making one of the two possible answers more likely than the other one (e.g. are you more likely to say 'Yes' than 'No'?). This test, which uses the binomial distribution that we studied in Appendix D, is called the *binomial* test.

E.2.1 *Kiwi* and *Koowoo*: a binomial problem

As an example of a simple kind of behavioral pattern that we might want to assess, let us consider a question explored by Chastaing (1958): Are the meanings of words in language entirely arbitrary, or are we able to assign meanings consistently to novel, unknown words simply on the basis of their sound? Precisely, Chastaing believed that in certain circumstances we are able to assign (some) meaning to words based on some of their phonological properties, and this is what he wanted to *prove*. This was, actually, a rather controversial topic because the conventional linguistic wisdom asserts that meanings are assigned to words by languages in a completely arbitrary way.

In order to support his theory (which, indeed, was more sophisticated than this brief account), Chastaing decided to show that even children will use properties of words to assign meaning in some tasks. Chastaing theorized that the sound 'eee' should symbolize, among other properties, '*smallness*' whereas the sound 'ooo' should, by contrast, symbolize '*broadness*'. Thus, the theory predicts that we should attribute (as a 'connotation') some properties of smallness to words with the sound 'eee'. To test his theory and its predictions, Chastaing designed an experiment using children as subjects. He used children, because he wanted to show that assignment of meanings was a *basic* property of language, and consequently should appear very early in human development.

Chastaing presented each child with two dolls, a large doll and a small doll, and told them that the names of the dolls were *Kiwi* and *Koowoo*. Please note that the only difference in the stimuli (i.e. the dolls) is the *size*, whereas the only difference in the words was the opposition 'eee' sound versus 'ooo'. The child's task was to tell which name went with which doll. Chastaing expected that the children, if they were responsive to the sounds of the words, would name the large doll *Koowoo* and the small doll *Kiwi*.¹

Out of 10 children, 9 agreed with Chastaing's expectation. This appears to support his notion that there is something about the sound of *Koowoo* that makes it appropriate for the large doll, and *Kiwi* for the small doll.

Now, suppose that, as a stern and traditional linguist, you basically disagree with Chastaing's theory, and hence with his interpretation of his data. You cannot deny that

¹ Actually Chastaing asked an experimenter, who was unaware of his predictions, to run the experiment. This procedure is called a *double blind* (*double* because, presumably, the children also did not know his predictions), and its goal is to insure that neither the subjects nor the experimenter will answer in a particular way just to please a nice theoretician.

9 Nchildren out of 10 behaved like Chastaing predicted, but you could deny that this result is *due to* a systematic preference of children for using ‘eee’ to symbolize smallness. Specifically, you would be tempted to attribute these results to chance. You would say something like:

These results do not prove anything. Children choose randomly *Kiwi* or *Koowoo*. The results obtained reflect *only* random fluctuations. They are just a fluke that reflects no systematic *bias* of the children.

This point of view asserts that the children were actually responding randomly — guessing with no systematic tendency toward one answer or the other. Whereas Chastaing’s interpretation would be that the behavior of the children reflects a genuine preference for *Kiwi* to symbolize the property ‘being small’ and for *Koowoo* to symbolize the property ‘being large’.

In sum, we are faced with two contradictory interpretations or *statistical hypotheses* to explain the data: either the results reflect a systematic bias of the children or they reflect only chance factors. The last hypothesis amounts to saying that the experimental effect is non-existent. We call it the *null hypothesis* (i.e. it says that the effect is null). This hypothesis is abbreviated as H_0 (read ‘H-zero’). The first hypothesis says that there is *some systematic factor* explaining the data. It essentially says that the *null hypothesis is false* (i.e. ‘there is *some* effect’ cannot be true at the same time as ‘there is *no* effect’). Because this hypothesis is cast as an *alternative* to the null hypothesis, it is called the *alternative hypothesis*. It is abbreviated as H_1 (read ‘H-one’). Note that the null hypothesis specifies a precise value (it says ‘the effect is *zero*’), it is a *precise statistical hypothesis*. By contrast, the alternative hypothesis does *not* specify any precise value (there are an infinite number of ways to be different from zero!), so it is an *imprecise statistical hypothesis*. In order to prove that the alternative hypothesis is true, the technique is to show that the null hypothesis is *unlikely* to be true given the data. This amounts to saying that we will consider that the null hypothesis is not true if the probability of obtaining our data is very small *if* the null hypothesis is true.

So, in order to eliminate the hypothesis that chance is responsible for the results of an experiment (this is the null hypothesis), we need to show that it is extremely unlikely to obtain our data if the null hypothesis is true. This is equivalent to saying that we will not believe in chance, if the results that we have obtained have a small probability of occurring by chance. This procedure is the basis of a *statistical test*.

So now the problem boils down to characterizing what should be the behavior of the subjects if they were to respond randomly. If the subjects respond randomly, each subject is as likely to decide that *Koowoo* goes with the small doll as to decide that it goes with the big doll. Consequently, if subjects show no preference, the response of each of them could be likened to flipping a coin and recording the number of Heads. Recording a Head would be equivalent to choosing *Koowoo* for the large doll. Hence, recording the responses of 10 subjects is equivalent to flipping 10 coins, at least if the subjects respond randomly. We actually happen to know how to evaluate the probability of any combination of Heads and Tails when tossing coins. This is described by the binomial distribution discussed in Appendix D. Figure E.1 shows the binomial distribution for $N = 10$ and $P = \frac{1}{2}$.

E.2.2 Statistical test

Now we can answer the question: How likely is it that we would get a response pattern as strong as 9 H and 1 T (or 9 *Koowoo* = large and 1 *Koowoo* = small) as a result of random

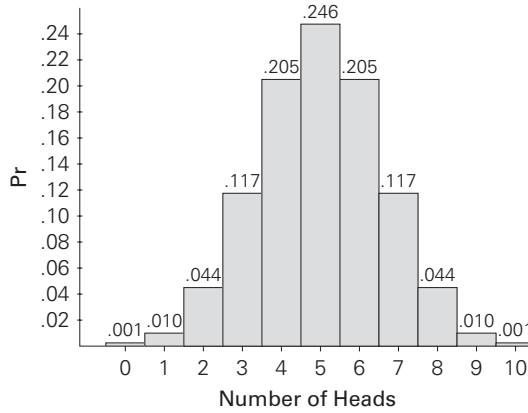


Figure E.1 The binomial distribution with $N = 10$.

guessing? We can see from Figure E.1 and from Appendix D that the probability of obtaining a 9|1 split is

$$\begin{aligned}
 \Pr(9 \text{ Heads out of } 10) &= \binom{10}{9} \times P^9 \times (1 - P)^{10-9} \\
 &= 10 \times .5^9 \times .5^1 \\
 &\approx .009766.
 \end{aligned}
 \tag{E.1}$$

But here we are interested in response patterns of 9|1 *or stronger*, since we would have been just as satisfied, or even more so, with 10|0. Thus, we are interested in the chances of getting a result *at least as strong as* 9|1 from random responding with equal probability. This means that we are interested in the *probability associated*² with the value 9. Namely, we want to evaluate the value of $p(A)$:

$$\begin{aligned}
 p(A) &= \Pr(9 \text{ Heads out of } 10) + \Pr(10 \text{ Heads out of } 10) \\
 &= \binom{10}{9} \times P^9 \times (1 - P)^{10-9} + \binom{10}{10} \times P^{10} \times (1 - P)^0 \\
 &= (10 \times .5^9 \times .5^1) + (1 \times .5^{10} \times .5^0) \\
 &= .009766 + .000977 \\
 &\approx .01074.
 \end{aligned}
 \tag{E.2}$$

So, we have roughly *one chance in one hundred* to obtain 9 or more Heads out of 10 coins when *chance alone* is responsible for the outcome of the experiment. (This probability is represented by the shaded area to the right of the vertical line drawn in Figure E.2).

One chance in one hundred is rather unlikely. Therefore it seems unlikely that the children's responses could have arisen from responding randomly. In more formal terms, this probability

² Remember that the probability of obtaining A or an event more extreme is called the *probability associated with A* . It is denoted $p(A)$.

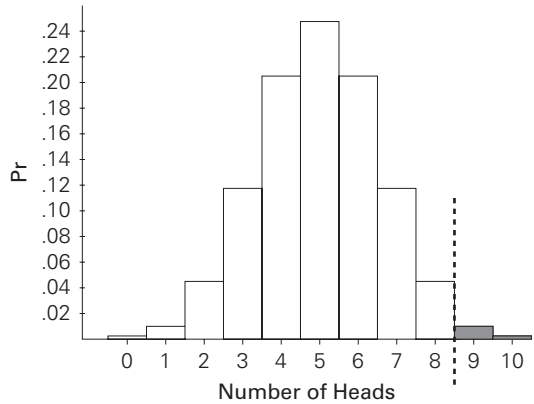


Figure E.2 The binomial distribution for $N = 10$ and $P = .5$. The shaded area represents the probability associated with the event ‘obtaining 9 Heads’.

is sufficiently low so that most psychologists would confidently reject the hypothesis that subjects were responding randomly (i.e. the *null hypothesis*: H_0), and accept the hypothesis that subjects were able to pair the names with the dolls with greater-than-chance consistency (i.e. the *alternative hypothesis*: H_1).

Implicit in our discussion is the existence of a threshold for deciding that something is unlikely. This threshold is denoted by the Greek letter alpha: α . It is, indeed, highly subjective, but, let us note that tradition favors values like ‘one chance in one hundred’ (.01 or 1%), and ‘five chances in one hundred’ (.05 or 5%). Suppose that we have settled on a threshold of $\alpha = .05$. Hence, we would feel that we can reject the null hypothesis because the probability associated with the outcome of our experiment under the null hypothesis is smaller than $\alpha = .05$.³

In this experiment, we started with the notion that if our hypothesis concerning the effects of the sounds of novel words was true, then *Koowoo* would go with the large doll. That is, the H_1 we wanted to test was only concerned with one direction of the result. We were only concerned with one tail of the distribution—namely the one where the children paired *Koowoo* with the large doll with greater-than-chance consistency. The test we made is thus called a *one-tailed test* because we are only going to accept the alternative hypothesis if the result is at one end of the distribution. If the children had been consistently naming the small doll *Koowoo*, the data would *not* have been supporting our theory, which states that they should name the *large* doll *Koowoo*.

Had we been interested in *any result* showing greater-than-chance consistency, even if the pairing of names and dolls was not what we expected, then we would formulate an alternative hypothesis that we would accept if the children consistently paired *either name* with the large doll. Then, we would be prepared to accept patterns of response at either end of the distribution as grounds to reject the null hypothesis H_0 . In the present case, responses of 9|1 Heads or stronger or 1|9 Head or weaker (‘weaker’ only in the sense of running counter to our original H_1) would lead us to reject H_0 and accept our new H_1 , that the children were

³ In journal articles, we summarize in the following way: *9 children out of 10 prefer Koowoo for naming the large doll, which shows support for a naming preference ($p < .05$).*

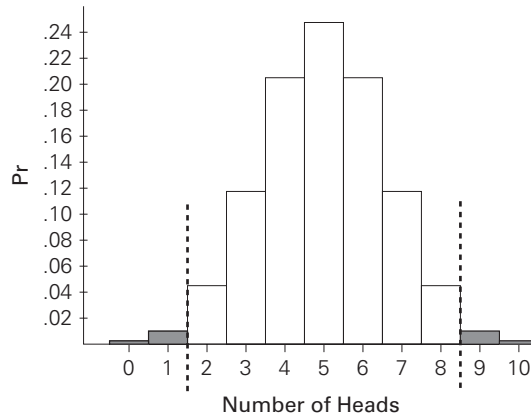


Figure E.3 Two-tailed test: The binomial distribution for $N = 10$ and $P = .5$. The shaded area represents the probability associated with the event ‘obtaining 9 Heads’ or the probability associated with the event ‘obtaining 1 Head’.

responding consistently, pure and simple. This is called a *two-tailed test*. Since the binomial distribution is symmetrical, the probability of obtaining a result that far from chance, taking account of both tails of the distribution, is just twice as great as it was for only one tail: $2 \times .0107 = 0.0215$, which is still small enough for most of us to reject H_0 . This is shown in Figure E.3.

E.3 Coda: the formal steps of a test

At this point, we should summarize formally the steps involved in carrying out a statistical test:

- Decide about a measure. Here, we have chosen the number of children who choose *Koowoo* as being the large doll.
- Specify the null hypothesis H_0 and the alternative hypothesis H_1 . The null hypothesis states that the results can be attributed to chance alone. The alternative hypothesis states that the results cannot be attributed to chance alone (e.g. they reflect some systematic trend).
- Select a statistical model and compute the probability associated with each possible outcome of the experiment if the null hypothesis is true. This is called constructing the sampling distribution of the experiment *under* H_0 .
- Choose α , which is the threshold probability for rejecting H_0 , and determine the region in the sampling distribution where H_0 will be rejected (in fact, formulate a *decision rule*).
- Carry out the experiment and evaluate the data in terms of the decision rule.

In the present case, we decided that we wanted to be able to distinguish between random responding (H_0) and consistent responding in the direction of pairing *Koowoo* and the large doll (H_1). Then, we selected a statistical model that we thought would represent random responding on each trial (namely that random behavior would be equivalent to flipping a *fair* coin). We use the binomial distribution because it gives the probability associated with

each possible value that we could have obtained in this experiment. We will say that we have constructed the *sampling distribution* of our measure (i.e. number of children choosing *Koowoo* for the large doll).

The next step, choosing a threshold probability for rejecting H_0 , was not explicitly present in the above description of the experiment. There is a long tradition in psychology of using probabilities of .05 or .01 as grounds for rejecting H_0 , but there is something subjective in selecting that threshold. How low a probability do you require, before you would be willing to accept that you can rule out the possibility of chance alone and accept the alternative that the results are systematic? Recall that the threshold probability is called the *alpha level*, symbolized by the lower-case Greek alpha (α). The threshold α refers to the probability of making a specific error: namely rejecting H_0 by mistake, when the response pattern really is random. This mistake can occur because, indeed, rare events do happen. And every time such a rare event happens, we will decide that the null hypothesis is false (when it is, in fact, true). This particular mistake or error is called a *Type I error*. In other words, α gives the probability of erroneously concluding that you have a statistically reliable result when you don't. We also call this error a *false alarm*. (Later we will be concerned with the *Type II error*: the probability that you are *missing* a statistically significant result that's really there, but that your statistical test does not detect.)

We can indeed wait after the end of the experiment and compute, then, the probability associated with its outcome. We can also characterize *before* the experiment the region of the sampling distribution that will lead to the rejection of the null hypothesis. We call the set of the values of the possible outcomes leading to a rejection of the null hypothesis, the *region of rejection*, or the *critical region* of the sampling distribution. We call the region of the sampling distribution for which we cannot reject the null hypothesis, the *region of non-rejection* or the *region of suspension* of judgment. If our result falls in the critical region, we say that we reject H_0 and accept H_1 . On the other hand, if our result falls in the region of suspension, we say that we fail to reject H_0 and suspend judgment (more later about this curious wording). The value at the border of these two regions is called the *critical value*. Precisely, the *critical value* is a value with a probability associated equal to or less than α and such that any value more extreme leads to rejecting the null hypothesis, whereas any value less extreme fails to reject the null hypothesis. For our example, with 10 binary outcomes, the critical value for $\alpha = .01$ is 9.⁴ The critical value for $\alpha = .05$ is 8 (see Figure E.1, and Appendix D, Section D.8, pages 453ff.). This is illustrated in Figure E.4.

Finally, we collect the data and compare them to our statistical model, with the result that we reject H_0 and accept H_1 —namely, that children can consistently assign meanings to novel words just on the basis of their sound.

E.4 More on decision making

To perform a statistical test is to make a decision based upon some data and some rules. Because our data are inherently noisy, we cannot be certain of always making the right decision. In addition, a statistical test is performed as a way of convincing someone (yourself

⁴ We have rounded the value of $p(9) \approx .0107$ to .01 because we keep the same number of decimals as the threshold (i.e. 2 decimals) when comparing. In practice, this type of problem is rarely relevant.

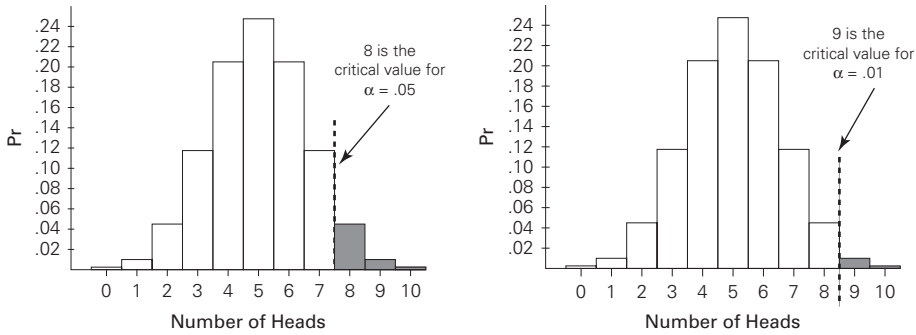


Figure E.4 Sampling distribution (i.e. Binomial distribution) for $N = 10$ binary outcomes. Critical values for $\alpha = .05$ (left panel) and $\alpha = .01$ (right panel). The values in the shaded areas correspond to a rejection of the null hypothesis.

or your colleagues) of the existence of an effect. So, if you are already convinced of the existence of an effect, maybe you will need less evidence to accept it than someone who is not very enthusiastic about this effect.

Clearly, part of the problem is linked to the choice of the α level. The smaller the threshold is, the harder it is to reject wrongly the null hypothesis. By contrast, however, the smaller the threshold is, the harder it is to reject the null hypothesis when it is, actually, false.

In fact, no matter where we place our cutoff point, we face a risk of being wrong. Figure E.5 summarizes what can happen when we make a statistical decision.

Figure E.5 shows that there are two ways that our decision can be wrong. In *reality*, the children either prefer *Koowoo* for naming a large doll or do not. Because we can *decide* that they do prefer *Koowoo* or that they don't for either of these real situations, there are four possible outcomes. If the children do not prefer *Koowoo* and we decide that they do, we will make a *false alarm* (decide that *Koowoo* is preferred when in fact it is not). How can we minimize the risk of making this mistake? We could make our requirements for deciding very stringent, placing the cutoff point so that the area of the tail that is cut off is very small. This positioning of our critical value would require the experiment to result in values that are *very unlikely* to occur if children are choosing randomly (for example zero or ten heads). However, there is a second way to be wrong in our decision. If the children actually prefer *Koowoo* and we decide that they do not, we will *miss* determining that they take the sound of the words to create meaning. We will fail to support our alternative hypothesis even though it was true. Can you see that by decreasing our risk of making a false alarm, we have increased our risk of

		DECISION	
		Do not prefer Koowoo	Do prefer Koowoo
REALITY	Do not prefer Koowoo	Correct 'no' $1 - \alpha$	False alarm Type I error α
	Do prefer Koowoo	Miss Type II error β	Hit Power $1 - \beta$

Figure E.5 The possible outcomes of a statistical test.

missing the discovery that the treatment had an effect? The risks of making the two types of mistakes are inversely related. As the risk of one kind of error decreases, the risk of the other kind increases.

As we have already seen, making a false alarm is called a *Type I* error.⁵ We actually know the probability of making such an error because it is an essential property of the test. The threshold α is chosen in such a way that any event with a probability smaller than α when H_0 is true, will imply rejection of the null hypothesis. Hence the probability of making a Type I error is actually α itself.

The other type of error is to *miss* an effect. An effect of the treatment is missed by failing to reject the null hypothesis when it is in fact false. In other words, a miss is saying that children do not show a preference for *Koowoo* when they, in fact, do prefer it. This mistake is called a *Type II* error.

The risk (probability) of making a Type II error is usually symbolized as β . Unlike α , this risk cannot be established by the experimenter. The value of β depends upon the true state of the world. In particular, β depends upon the intensity of the supposed effect that we want to detect. A strong effect is easier to detect than a small one. It is analogous to trying to detect a signal embedded in noise. The softer the noise, or the stronger the signal, the easier it is to detect the signal. However, the experimenter has *some* control over β . Namely, when α is made smaller by the experimenter, β will increase.

The *power* of a test is the probability of rejecting the null hypothesis when it is in fact false. Precisely, the power of a test is defined as $1 - \beta$. It corresponds to the probability of correctly detecting an existing effect. In our example, the power is the probability that the children indeed prefer *Koowoo* and that we decide that they do. In other words, our decision that there is a systematic preference is correct. There are two ways to increase the power of an experiment. One way is to increase α ,⁶ or in other words, to raise the risk of making a Type I error. Therefore, as α becomes larger, β becomes smaller and consequently $1 - \beta$ becomes larger. Increasing the number of observations (e.g. subjects, coin tosses) will also increase the power.

Using statistical terminology, the four situations that we have just described are summarized in Figure E.6.

		DECISION	
		Fail to reject H_0	Reject H_0
REALITY	H_0 true	Correct 'no' $1 - \alpha$	False alarm Type I error α
	H_0 false	Miss Type II error β	Hit Power $1 - \beta$

Figure E.6 The possible outcomes of a statistical test and their probabilities.

⁵ Yes, there is a Type II. In fact there is even a Type IV, but no Type III, funnily enough. A Type IV is to reach the *wrong* conclusion from a correct statistical analysis. Statisticians decided to call it a Type IV, because several people were arguing about several possible candidates for a Type III.

⁶ Which is equivalent to accept making more false alarms.

E.4.1 Explanation of ‘tricky’ wording

Some of the wording above may seem unnecessarily complex. Why don’t we just say that we ‘accept the null hypothesis’ instead of the round about ‘fail to reject the null hypothesis’? Why can we ‘accept the alternative hypothesis’, but not reject it? There is a reason for this apparently perverse wording. Consider that there is only *one* distribution that occurs when the coin is fair. When the coin is not fair, there are many ways that it could be biased. It could be biased so that the probability of a Head is .9 or .8 or .7 or .6 or .4 or .3 and so on. In fact there are an infinite number of ways to have a biased coin. If you list all the ways you can think of, someone can always name one more way. To accept the null hypothesis (and in conjunction reject the alternative hypothesis) we would have to determine how unlikely it was to get our results under each possible biased distribution. This is not possible because of the infinite number of ways to have a biased coin.

If the above explanation has severely raised your anxiety level, consider the following, more intuitive, example. Take the following situation:

1. You don’t know what city you are in.
2. It rains *A LOT* in Seattle.
3. The null hypothesis is that you are in Seattle.
4. The dependent variable that is measured is the number of consecutive sunny days.

If your measurement shows that it is sunny for 30 days straight you would probably be willing to conclude, with little risk of being wrong, that you were *not* in Seattle. You would therefore reject the null hypothesis. On the other hand, suppose that it has rained for 30 days straight. Would you conclude that you were in Seattle, thus accepting the null hypothesis? Couldn’t you be somewhere else? Aren’t there a lot of places where it rains a lot?⁷ But, then again, isn’t Seattle one of those places? Can you rule out those other places? Therefore, it would be most logical to conclude that although you can’t say for certain that you are in Seattle, you can’t reject that possibility either. In other words, you would fail to reject the null hypothesis and suspend judgment.

E.5 Monte-Carlo binomial test

As we have seen in Appendix D (Section D.12, pages 465ff.), we can use Monte-Carlo methods to estimate probability distributions. Therefore, we can use Monte-Carlo simulations to make statistical decisions. In the *Koowoo* example, we want to estimate the probability associated with obtaining 9 responses out of 10 children stating that ‘*Koowoo* is the large doll’ when the null hypothesis is true. If the children answer randomly, their behavior can be modeled by flipping 10 coins and counting the number of Heads. If we repeat this experiment for a large number of times, we can derive an estimation of the sampling distribution.

We have done just that. We flipped 10,000 times a set of 10 coins and counted the number of coins landing on Heads for each experiment. So, at the end of this procedure, we know how many samples of 10 coins give 0 Heads, 1 Head, ..., 10 Heads. In order to estimate the

⁷ You could even have spent all that time in your shower!

probability associated with a given outcome, we just need to count the number of times this event, or an event more extreme, occurred, and to divide this number by 10,000.

For the *Koowoo* example, we want to find the probability of ‘obtaining 9 or more *Koowoo* answers’ if the responses are chosen randomly. We found that out of 10,000 samples of 10 coins that we tossed, 108 gave 9 or 10 Heads. Therefore, an estimation of the probability associated to the event ‘obtaining 9 *Koowoo* responses out of 10 trials by chance alone’ is

$$\begin{aligned} p(9 \text{ or more } Koowoo) &= \frac{\text{number of samples with 9 or more } Koowoo}{\text{total number of samples}} \\ &= \frac{108}{10,000} \\ &= .0108 . \end{aligned} \tag{E.3}$$

The complete sampling distribution along with its estimations is shown in Figure E.7.

We will indeed reach the same conclusion with the Monte-Carlo test as with the binomial test, namely that it is unlikely that the children respond randomly, and that they do prefer to call the large doll *Koowoo*.

E.5.1 A test for large *N*: normal distribution

Let us now take an additional step elaborating the test we have just made. Here our example will be an experiment by Brown *et al.* (1955). They wanted to investigate a phenomenon closely

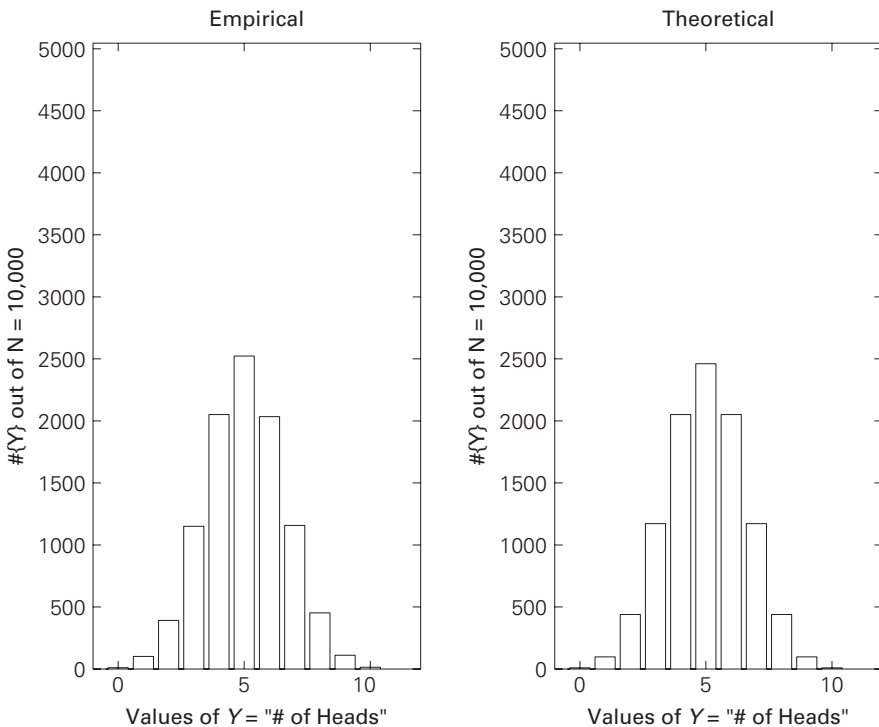


Figure E.7 A Monte-Carlo estimation (left panel) of the sampling distribution under the null hypothesis. Compare with the theoretical binomial distribution (right panel).

related to the one investigated by Chastaing, called ‘phonetic symbolism’. This experiment tests the hypothesis that you can tell what various words from natural languages mean just from the way they sound. For example, we might suppose that pairs of words like ‘large’ and ‘little’ go well with the meanings they have because we open our mouths more in saying one than saying the other, and that this process operates in languages around the world. To provide a fair test of this hypothesis Brown *et al.* chose numerous pairs of words in languages that were not only unknown to the subjects, but quite far removed from any languages their American subjects might have come in contact with: Chinese, Czech, and Hindi. A typical set of word pairs consists of the words for ‘beautiful’ and ‘ugly’ (respectively): *mei*, and *choŭ* (Chinese); *kraša*, *ošklivost* (Czech); and *khubsurat*, *badsurat* (Hindi). Subjects were presented with the English word pair and the foreign word pair, and heard the foreign pair pronounced. Then, they had to decide which foreign word went with which English word. In the case of the words meaning ‘beautiful’ and ‘ugly’, the results from 86 subjects were: Chinese, 88% correct; Czech, 57%; and Hindi, 64%. Clearly, pure guessing—random performance—would lead to an average result of 50%. Did the subjects perform better than chance?

Let us look at the results for the Chinese language. Out of 86 subjects, 76 picked up the correct pairing. Can we conclude that this result shows that people can derive some meaning from the sound of words?

Again, to answer this question we can compare the outcome to the distribution of outcomes to be expected from random responding, and see how unlikely the result is. However, we don’t want to draw a tree diagram as we did in Appendix C, with

$$2^{86} = \text{a BIG number} \quad (\text{E.4})$$

elementary events (the number in question is close to 10 followed by 15 zeros!) Figure E.8 shows the probability distribution for patterns of outcomes from 86 trials. As we have seen in Appendix D, the binomial distribution is becoming smoother as we increase the number of trials, N , and looks more and more like the normal distribution.

The normal distribution is very convenient to use as an approximation to the binomial distribution when N becomes large and the probabilities of the underlying events are close to fifty-fifty (i.e. we try to find the sampling distribution of *fair* coins).⁸ All we need to know are the two parameters that characterize the normal distribution and allow us to fit it to the binomial distribution: the mean of the distribution (μ_Y) and its standard deviation (σ_Y). For the binomial distribution, when $P = .5$, the mean is given by $N \times P = N \times .5$. The standard deviation for the binomial distribution is, in general,

$$\sigma_Y = \sqrt{N \times P \times (1 - P)},$$

and in the case of $P = .5$ this simplifies to $.5 \times \sqrt{N}$.

Hence, we have the following formulas for the specific case of the binomial distribution for $P = .5$:

$$\begin{aligned} \mu_Y &= .5 \times N \\ \sigma_Y &= .5\sqrt{N}. \end{aligned} \quad (\text{E.5})$$

Once we know these two parameters, we can use the convenient properties of the normal distribution to assess the probability of obtaining a result as strong as 76 successes out of 86

⁸ Actually, it begins to work very well for N greater than about 25.

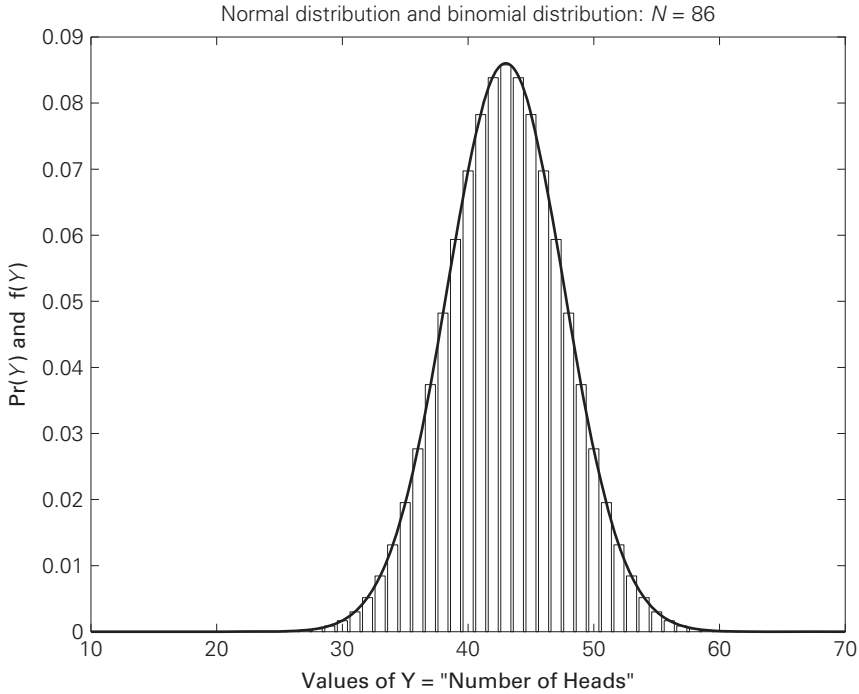


Figure E.8 Binomial distribution with $N = 86$, together with its normal or Gaussian approximation.

(the result obtained by Brown *et al.*) just by chance. In the present case, the two parameters will be

$$\mu_Y = .5 \times 86 = 43, \quad \text{and} \quad \sigma_Y = .5 \times \sqrt{86} \approx 4.6368. \quad (\text{E.6})$$

Knowing those two parameters, we can transform the outcome of this experiment into a Z -score then use the table of the standardized normal distribution given in the appendix to find the probability associated with the Z -score. This probability is the probability of obtaining the outcome of the experiment under the null hypothesis (i.e. by chance alone).

Specifically, we want to know the probability of a result as strong as 76 out of 86 or stronger (i.e. we want to find the probability *associated* with the event ‘76 out of 86’). This is equivalent to finding the area under the normal curve to the right of the criterion that falls at the place corresponding to 76-out-of-86. (This will give us a one-tailed test. What would we do for a two-tailed test?)

We need to find the Z -score corresponding to 76 out of 86 (which is a score coming from a binomial distribution whose mean is $\mu = 86/2 = 43$, and whose standard deviation is $\sigma = \sqrt{86/4} \approx 4.6368$). Recall that a Z -score transforms a score by expressing it as a distance from the mean normalized by the standard deviation. Or with a formula,⁹ the Z -score is

$$Z_Y = \frac{Y - M_Y}{\sigma_Y}.$$

⁹ For simplicity, we didn’t use the *continuity correction*, explained in Appendix D (Section D.11, pages 462ff.). When N is larger than 50, its effect is purely cosmetic, anyway!

Plugging our previously calculated mean and standard deviation into this formula gives:

$$Z_Y = \frac{76 - 43}{4.6368} = 7.117.$$

The table we have for the normal distribution gives the area under the curve *to the left* of the chosen criterion. To obtain the area to the right of the criterion, subtract that area from 100% (=1.00). Note that the table does not provide areas for Z beyond around 5. The area for $Z = 7.117$ is vanishingly small (i.e. smaller than .00001).

Therefore we can reject the H_0 that the subjects were responding randomly, and accept the H_1 that the subjects were responding systematically to features of the foreign words.

As an exercise, try to follow the procedure we have just used to evaluate the results for the Czech and Hindi word pairs (with 57% and 64% responses in the correct direction, respectively).

Chapter summary

E.6 Key notions of the chapter

Below are the main notions introduced in this chapter. If you have problems understanding them, you may want to re-read the part(s) of the chapter in which they are defined and used. One of the best ways is to write down a definition of each of those notions by yourself with the book closed.

Statistical test	False alarm
Null hypothesis	Miss
Alternative hypothesis	Power
Decision rule or criterion	Sampling distribution
Level of significance	One-tail vs two-tail test
Type I error	Monte-Carlo
Type II error	

E.7 New notations

Below are the new notations introduced in this chapter. Test yourself on their meaning.

H_0

H_1

α

β

E.8 Key questions of the chapter

Below are some questions about the content of this chapter. All the answers are to be found in the chapter. If you are in any doubt about your answer, you may want to re-read parts of the chapter.

- * Why do we use statistical tests?
- * When would you use a binomial test?
- * When would you use a one-tail test rather than a two-tail test?
- * Why can't we accept the null hypothesis?
- * What is the probability of making a Type I error?
- * What does it mean to have a level of significance $\alpha = .01$?
- * What is the difference between binomial and normal sampling distributions?

F

Expected values

F.1 Introduction

In this appendix we look at expected values of random variables and how to find them. In experimental design we use the notion of expected value essentially for deriving the expected value of mean squares. This is an essential step for building F and F' ratios. In the previous chapters, the expected values of the mean squares were given and the assumptions used to derive them were merely mentioned. Here we are going to see how these assumptions are used to derive the expected values. Warning: this chapter is a bit more technical than most of the chapters in this book and is full of formulas.¹ But it is still worth taking a look at it, because it makes explicit the techniques used and their limits. The formalization of the techniques and properties used in this chapter are the basis of the rules described in Chapter 22.

The main purpose of this appendix is to show that the derivation of the expected values of the mean squares is a rather straightforward process, albeit tedious. So in this appendix we want to take the mystery (and the magic) out of the expected values.

To make this appendix easier to read, we will deal only with finding expected values for discrete random variables. This is not a real limitation because the same properties found for discrete variables will hold for continuous variables as well (it suffices, in general to replace the sum sign $\sum$ by the integral sign: $\int dx$, and the expression 'sum of' by 'integral of'). We give short intuitive proofs of most of the properties that we use.

F.2 A refresher

We need to remember some definitions (given in Appendix D). They are briefly restated below. If you need a more thorough refresher don't hesitate to re-read the chapters on probability.

Definition F.1 (Random variable). *Y is a random variable if: (1) it can take several different values (hence the term variable); and (2) if we can associate a probability with each value of Y .*

¹ So if you suffer from intense *mathophobia*, take a deep breath and keep on reading. Stiff upper lip!

In other words, Y is a random variable if we can list all its possible values and associate with each value a number called its probability. A probability is a number whose value is between 0 and 1. In addition, the sum of the probabilities of all the values of the random variable is equal to 1. This leads to the next definition.

Definition F.2 (Probability distribution). *A probability distribution is a list of all the possible values of a random variable with a probability assigned to each possible value of the random variable.*

Definition F.3 (Expected value). *If Y is a random variable taking N values*

$$Y_1, Y_2, \dots, Y_N$$

with the probability assigned to each value being

$$\Pr(Y_1), \Pr(Y_2), \dots, \Pr(Y_N),$$

then, the expected value of Y , or mathematical expectation of Y , is noted $\mathbf{E}\{Y\}$, or μ_Y , and is computed as:

$$\begin{aligned} \mathbf{E}\{Y\} &= \mu_Y \\ &= \Pr(Y_1) \times Y_1 + \Pr(Y_2) \times Y_2 + \dots + \Pr(Y_n) \times Y_n + \dots + \Pr(Y_N) \times Y_N \\ &= \sum_n^N \Pr(Y_n) \times Y_n. \end{aligned} \quad (\text{F.1})$$

In order to make reading smoother, the notation $\Pr(Y_n)$ can be replaced by P_n . The expected value of Y is also called the mean of the random variable.

Definition F.4 (Variance). *If Y is a random variable taking N values*

$$Y_1, Y_2, \dots, Y_N$$

with the probability assigned to each value being

$$P_1, P_2, \dots, P_N.$$

and with μ_Y being its expected value (i.e. mean), the variance of Y is denoted σ_Y^2 or $\text{Var}(Y)$ and is computed as:

$$\begin{aligned} \sigma_Y^2 &= \text{Var}(Y) = \mathbf{E}\{(Y - \mu_Y)^2\} \\ &= P_1 \times (Y_1 - \mu_Y)^2 + P_2 \times (Y_2 - \mu_Y)^2 \\ &\quad + \dots + P_N \times (Y_N - \mu_Y)^2 \\ &= \sum_n^N P_n \times (Y_n - \mu_Y)^2. \end{aligned} \quad (\text{F.2})$$

With all these definitions in mind, we can now derive several properties that we use (and overuse) later on when computing expected values of mean squares.

Property 1. *If K is a constant, then:*

$$\mathbf{E}\{KY\} = K\mathbf{E}\{Y\}. \quad (\text{F.3})$$

Proof:

$$\mathbf{E}\{KY\} = \sum P_n KY = K \sum P_n Y = K\mathbf{E}\{Y\}. \quad (\text{F.4})$$

Property 2 (Expectation of sums). *If Y and Z are two random variables then*

$$\mathbf{E}\{Y + Z\} = \mathbf{E}\{Y\} + \mathbf{E}\{Z\} . \quad (\text{F.5})$$

Proof:

$$\mathbf{E}\{Y + Z\} = \sum P_n(Y + Z) = \sum P_n(Y) + \sum P_n(Z) = \mathbf{E}\{Y\} + \mathbf{E}\{Z\} . \quad (\text{F.6})$$

Property 3. *If K is a constant then*

$$\mathbf{E}\{K\} = K. \quad (\text{F.7})$$

Proof:

$$\mathbf{E}\{K\} = \sum P_n K = K \sum P_n = K \quad \left(\text{Remember: } \sum P_n = 1 \right) . \quad (\text{F.8})$$

Property 4. *If K is a constant then*

$$\mathbf{E}\{Y + K\} = \mathbf{E}\{Y\} + K \quad (\text{F.9})$$

(follows from Properties 2 and 3).

Property 5. *The variance of Y can be conveniently computed as*

$$\sigma_Y^2 = \text{Var}(Y) = \mathbf{E}\{Y^2\} - (\mathbf{E}\{Y\})^2 \quad (\text{F.10})$$

Proof: We start with Definition F.2 and develop

$$\begin{aligned} \sigma_Y^2 &= \text{Var}(Y) = \mathbf{E}\left\{(Y - \mathbf{E}\{Y\})^2\right\} \\ &= \mathbf{E}\left\{(Y^2 + \mathbf{E}\{Y\}^2 - 2Y\mathbf{E}\{Y\})\right\} \\ &= \sum P_n \left[Y_n^2 + (\mathbf{E}\{Y\})^2 - 2Y_n \mathbf{E}\{Y\} \right] \\ &= \sum P_n Y_n^2 + (\mathbf{E}\{Y\})^2 + 2\mathbf{E}\{Y\} \times \left(\sum P_n Y_n \right) \\ &= \mathbf{E}\{Y^2\} + (\mathbf{E}\{Y\})^2 - 2(\mathbf{E}\{Y\})^2 \\ &= \mathbf{E}\{Y^2\} - (\mathbf{E}\{Y\})^2 . \end{aligned} \quad (\text{F.11})$$

Property 6. *If K is a constant then*

$$\text{Var}(Y + K) = \text{Var}(Y) . \quad (\text{F.12})$$

Proof:

$$\begin{aligned} \text{Var}(Y + K) &= \mathbf{E}\left\{(Y + K - \mathbf{E}\{Y + K\})^2\right\} \\ &= \mathbf{E}\left\{(Y + K - \mathbf{E}\{Y\} - \mathbf{E}\{K\})^2\right\} \\ &= \mathbf{E}\left\{(Y - \mathbf{E}\{Y\})^2\right\} \\ &= \text{Var}(Y) . \end{aligned} \quad (\text{F.13})$$

Property 7. *If K is a constant then*

$$\text{Var}(Y \times K) = K^2 \text{Var}(Y) . \quad (\text{F.14})$$

Proof:

$$\begin{aligned} \text{Var}(Y \times K) &= \mathbf{E} \left\{ (Y \times K - \mathbf{E}\{Y \times K\})^2 \right\} \\ &= \mathbf{E} \left\{ (Y \times K - \mathbf{E}\{Y\} \times K)^2 \right\} \\ &= \mathbf{E} \left\{ K^2 (Y - \mathbf{E}\{Y\})^2 \right\} \\ &= K^2 \text{Var}(Y) . \end{aligned} \quad (\text{F.15})$$

We need also the definition of the notions of covariance and statistical independence for random variables (we used these notions several times in the text but did not give a precise definition). They now follow.

Definition F.5 (Independence). *If Y and Z are two statistically independent random variables², then*

$$\mathbf{E}\{Y \times Z\} = \mathbf{E}\{Y\} \times \mathbf{E}\{Z\} . \quad (\text{F.16})$$

Definition F.6 (Covariance). *If Y and Z are two statistically independent random variables, their covariance, denoted*

$$\mathbf{E}\{Y \times Z\} - \mathbf{E}\{Y\} \mathbf{E}\{Z\} = \mathbf{E}\{(Y - \mathbf{E}\{Y\})(Z - \mathbf{E}\{Z\})\} . \quad (\text{F.17})$$

Property 8 (Independence and covariance). *Combining Equations F.16 and F.17, we see that when two random variables are statistically independent, their covariance is equal to zero.*

F.3 Expected values: the works for an $\mathcal{S}(\mathcal{A})$ design

F.3.1 A refresher

A mean square is obtained by dividing a sum of squares by its number of degrees of freedom. So, before computing the mean squares, we need to compute the sums of squares corresponding to the different sources of variation. For an $\mathcal{S}(\mathcal{A})$ design, the total sum of squares is partitioned into two orthogonal sums of squares (one sum of squares per source):

- The sum of squares between expresses the effect of the independent variable. It is denoted SS_A .
- The sum of squares within represents the within-group variability. Because it expresses the effect of the subject factor (which is nested in the experimental factor $\mathcal{A}$), this sum of squares is denoted: $SS_{S(\mathcal{A})}$.

² Caution! Don't confuse two notions here: statistically independent variables and independent variables. 'Statistically independent' should be interpreted as being one term only.

In order to find the expected values of the mean squares, it is easier to start by finding the expected value of the sums of squares. Then the expected values of the mean squares are obtained by dividing the sums of squares by their respective degrees of freedom. This can be done because the degrees of freedom are a constant thus allowing the use of Property 1 of expected values.

F.3.2 Another refresher: the score model

For this exercise, we need the score model and also the statistical assumptions (or conditions of validity).

$$Y_{as} = \mu + \alpha_a + e_{s(a)} \quad (\text{F.18})$$

If S denotes the number of subjects per group, and A the number of levels of $\mathcal{A}$, the sum of the scores in the a th condition is expressed from the score model as:

$$Y_{a\cdot} = S\mu + S\alpha_a + e_{\cdot(a)} \quad \left(\text{with } e_{\cdot(a)} = \sum_a e_{s(a)} \right). \quad (\text{F.19})$$

The grand total is expressed as:

$$Y_{\cdot\cdot} = AS\mu + e_{\cdot(\cdot)} \quad \left(\text{with } e_{\cdot(\cdot)} = \sum_s \sum_a e_{s(a)} \right). \quad (\text{F.20})$$

The technical assumptions of the score model correspond to the following conditions.

Condition 1 (Fixed score model).

$$\sum \alpha_a = 0 \quad \text{and} \quad \frac{\sum \alpha_a^2}{A-1} = \vartheta_a^2 \iff \sum \alpha_a^2 = (A-1)\vartheta_a^2. \quad (\text{F.21})$$

Condition 2 (Homogeneity of variance).

$$\mathbf{E}\{e_{s(a)}\} = 0 \quad \text{and} \quad \text{Var}(e_{s(a)}) = \sigma_e^2 \quad \text{for all } a, s. \quad (\text{F.22})$$

Condition 3 (Independence of error).

$$\text{Cov}(e_{s(a)}, e_{s'(a')}) = 0 \quad \text{for all } a, s \neq a', s'. \quad (\text{F.23})$$

Conditions 2 and 3 can be rewritten in a slightly different manner (this will make the following developments easier to derive). From Condition 2, we know that

$$\mathbf{E}\{e_{s(a)}\} = 0 \quad \text{for all } a, s.$$

Then it follows that:

Condition 4 (e is a deviation).

$$\mathbf{E}\{e_{s(a)}^2\} = \text{Var}(e_{s(a)}) = \sigma_e^2 \quad (\text{because } e_{s(a)} = e_{s(a)} - 0). \quad (\text{F.24})$$

Condition 5 (Rectangle error terms are cross-products).

$$\mathbf{E}\{e_{s(a)}e_{s'(a')}\} = \text{Cov}(e_{s(a)}, e_{s'(a')}) \quad \text{for all } a, s \neq a', s'. \quad (\text{F.25})$$

Condition 6 (Independence of error and $\mathcal{A}$).

$$\mathbf{E}\{e_{s(a)}\alpha_a\} = 0 \quad \text{for all } a, s. \quad (\text{F.26})$$

F.3.3 Back to the expected values

The first step to compute the expected values is to rewrite the comprehension formula as computational counterpart:

The sum of squares between is

$$SS_A = \boxed{\mathbf{A}} - \boxed{\mathbf{1}}. \quad (\text{F.27})$$

The sum of squares within is

$$SS_{S(A)} = \boxed{\mathbf{AS}} - \boxed{\mathbf{A}}. \quad (\text{F.28})$$

With

$$\boxed{\mathbf{A}} = \frac{\sum Y_a^2}{S}, \quad \boxed{\mathbf{AS}} = \sum Y_{as}^2 \quad \text{and} \quad \boxed{\mathbf{1}} = \frac{Y_{..}^2}{AS}. \quad (\text{F.29})$$

Using Property 2, we find that we need to derive the expected values of the three 'numbers in the squares'

$$\mathbf{E}\{\boxed{\mathbf{A}}\}, \quad \mathbf{E}\{\boxed{\mathbf{AS}}\} \quad \text{and} \quad \mathbf{E}\{\boxed{\mathbf{1}}\}. \quad (\text{F.30})$$

And then, by addition and subtraction, we will find the expected values of the sums of squares.

F.3.4 Evaluating $\boxed{\mathbf{A}}$

In order to derive $\mathbf{E}\{\boxed{\mathbf{A}}\}$, we first express $\boxed{\mathbf{A}}$ in terms of the score model and develop:

$$\begin{aligned} \boxed{\mathbf{A}} &= \frac{\sum Y_a^2}{S} \\ &= \frac{1}{S} \sum (S\mu + S\alpha_a + e_{\cdot(a)})^2 \\ &= \frac{1}{S} \sum (S^2\mu^2 + S^2\alpha_a^2 + e_{\cdot(a)}^2 + 2S^2\mu\alpha_a + 2S\mu e_{\cdot(a)} + 2S\alpha_a e_{\cdot(a)}) \\ &= AS\mu^2 + S(A-1)\vartheta_a^2 + \frac{1}{S} \sum e_{\cdot(a)}^2 \\ &\quad + 2S\mu \sum \alpha_a + 2\mu \sum e_{\cdot(a)} + 2 \sum \alpha_a e_{\cdot(a)}. \end{aligned} \quad (\text{F.31})$$

Taking into account that $\sum \alpha_a = 0$ (cf. Condition 2), $\boxed{\mathbf{A}}$ becomes:

$$\boxed{\mathbf{A}} = AS\mu^2 + S(A-1)\vartheta_a^2 + \frac{1}{S} \sum e_{\cdot(a)}^2 + 2\mu \sum e_{\cdot(a)} + 2 \sum \alpha_a e_{\cdot(a)}. \quad (\text{F.32})$$

Because $\mathbf{A}$ is a sum of several terms, we need to find the expected value of each term (cf. Property 2). The first four terms are rather straightforward:

$$\mathbf{E}\{AS\mu^2\} = AS\mu^2 \quad (\text{Property 3})$$

$$\mathbf{E}\{S(A-1)\vartheta_a^2\} = S(A-1)\vartheta_a^2 \quad (\text{Property 3})$$

$$\mathbf{E}\left\{2\mu \sum e_{\cdot(a)}\right\} = 2\mu \sum \mathbf{E}\{e_{\cdot(a)}\} = 0 \quad (\text{Property 1, and Condition 2})$$

$$\mathbf{E}\left\{2 \sum \alpha_a e_{\cdot(a)}\right\} = 2 \sum \mathbf{E}\{\alpha_a e_{\cdot(a)}\} = 0 \quad (\text{Property 1, and Condition 6}) .$$

Deriving the expected value for the term $\frac{1}{S} \sum e_{\cdot(a)}^2$ is a bit more complex, so take a very deep breath and here it is:

$$\begin{aligned} \mathbf{E}\left\{\frac{1}{S} \sum e_{\cdot(a)}^2\right\} &= \frac{1}{S} \mathbf{E}\left\{\sum e_{\cdot(a)}^2\right\} \\ &= \frac{1}{S} \sum \mathbf{E}\{e_{\cdot(a)}^2\} \\ &\quad (\text{Properties 3 and 2, and rewriting } e_{\cdot(a)}) \\ &= \frac{1}{S} \sum_a \mathbf{E}\left\{\sum_s e_{\cdot(a)}^2 + 2 \sum_{s < s'} e_{s(a)} e_{s'(a)}\right\} \\ &\quad (\text{Properties of summation}) \\ &= \frac{1}{S} \sum_a \mathbf{E}\left\{\sum_s e_{\cdot(a)}^2\right\} + \frac{2}{S} \sum_a \mathbf{E}\left\{\sum_{s < s'} e_{s(a)} e_{s'(a)}\right\} \\ &\quad (\text{Property 2}) \\ &= \frac{1}{S} \sum_a \sum_s \mathbf{E}\{e_{\cdot(a)}^2\} + \frac{2}{S} \sum_a \sum_{s < s'} \mathbf{E}\{e_{s(a)} e_{s'(a)}\} \\ &\quad (\text{Property 2}) \\ &= \frac{1}{S} \sum_a \sum_s \sigma_e^2 + \frac{2}{S} \sum_a \sum_{s < s'} \text{Cov}(e_{s(a)}, e_{s'(a)}) \\ &\quad (\text{Conditions 2 and 5}) \\ &= \frac{1}{S} \sum_a \sum_s \sigma_e^2 \\ &\quad (\text{Condition 3}) \\ &= \frac{AS\sigma_e^2}{S} \\ &= A\sigma_e^2. \end{aligned} \quad (\text{F.33})$$

Putting all these terms together gives the expected value of $\boxed{\mathbf{A}}$:

$$\mathbf{E}\{\boxed{\mathbf{A}}\} = AS\mu^2 + S(A-1)\vartheta^2 + A\sigma^2. \quad (\text{F.34})$$

(So it was not that bad! Was it?)

F.3.5 Evaluating $\boxed{\mathbf{AS}}$

We start by expressing $\boxed{\mathbf{AS}}$ in terms of the score model:

$$\mathbf{E}\{\boxed{\mathbf{AS}}\} = \sum Y_{as}^2 = \sum (\mu + \alpha_a + e_{s(a)})^2. \quad (\text{F.35})$$

(Note that the sums are for the indices a and s .) Squaring and distributing the $\sum$ sign, Equation F.35 becomes

$$\begin{aligned} \boxed{\mathbf{AS}} &= \sum_{a,s} \left(\mu^2 + \alpha_a^2 + e_{s(a)}^2 + 2\mu\alpha_a + 2\mu e_{s(a)} + 2\alpha_a e_{s(a)} \right) \\ &= \sum_{a,s} \mu^2 + \sum_{a,s} \alpha_a^2 + \sum_{a,s} e_{s(a)}^2 + 2\mu \sum_{a,s} \alpha_a + 2\mu \sum_{a,s} e_{s(a)} + 2 \sum_{a,s} \alpha_a e_{s(a)}. \end{aligned} \quad (\text{F.36})$$

Simplifying and taking into account that $\sum \alpha_a = 0$ and that $\sum \alpha_a^2 = (A-1)\theta_a^2$ (remember: we make these assumptions in the score model for a fixed factor), we obtain:

$$\boxed{\mathbf{AS}} = AS\mu^2 + S(A-1)\vartheta_a^2 + \sum_{a,s} e_{s(a)}^2 + 2\mu \sum_{a,s} e_{s(a)} + 2 \sum_{a,s} \alpha_a e_{s(a)}. \quad (\text{F.37})$$

To find the expected value of $\boxed{\mathbf{AS}}$, we compute the expected value of each of its components (cf. Property 2):

$$\mathbf{E}\{AS\mu^2\} = AS\mu^2 \quad (\text{cf. Property 3})$$

$$\mathbf{E}\{S(A-1)\vartheta_a^2\} = S(A-1)\vartheta_a^2 \quad (\text{cf. Property 3})$$

$$\mathbf{E}\left\{\sum e_{s(a)}^2\right\} = \sum e_{\cdot(a)} e_{s(a)}^2 = AS\sigma_e^2 \quad (\text{cf. Condition 1})$$

$$\mathbf{E}\left\{2\mu \sum e_{s(a)}\right\} = 2\mu \sum \mathbf{E}\{e_{s(a)}\} = 0 \quad (\text{cf. Condition 2})$$

$$\mathbf{E}\left\{2 \sum \alpha_a e_{s(a)}\right\} = 2 \sum \mathbf{E}\{\alpha_a e_{s(a)}\} = 0 \quad (\text{cf. Condition 6})$$

Therefore the expected value of $\boxed{\mathbf{AS}}$ is equal to:

$$\mathbf{E}\{\boxed{\mathbf{AS}}\} = AS\mu^2 + S(A-1)\vartheta_a^2 + AS\sigma_e^2. \quad (\text{F.38})$$

F.3.6 Evaluating **1**

We start by expressing **1** in terms of the score model:

$$\mathbf{1} = \frac{Y_{..}}{AS}. \quad (\text{F.39})$$

Squaring and expanding gives

$$\begin{aligned} \mathbf{1} &= \frac{1}{AS} (AS\mu + e_{\cdot(\cdot)})^2 \\ &= \frac{1}{AS} (A^2S^2\mu^2 + e_{\cdot(\cdot)}^2 + 2AS\mu e_{\cdot(\cdot)}) \\ &= AS\mu^2 + \frac{1}{AS} e_{\cdot(\cdot)}^2 + 2\mu e_{\cdot(\cdot)}. \end{aligned} \quad (\text{F.40})$$

To find the expected value of **1**, we compute the expected value of each of its components (cf. Property 2). Two of the three components of **1** are relatively straightforward:

$$\mathbf{E}\{AS\mu^2\} = AS\mu^2 \quad (\text{cf. Property 3})$$

$$\mathbf{E}\{2\mu e_{\cdot(\cdot)}\} = 2\mu \mathbf{E}\{e_{\cdot(\cdot)}\} = 0 \quad (\text{cf. Condition 2})$$

The third term is a bit trickier, so here it is:

$$\begin{aligned} \mathbf{E}\left\{\frac{1}{AS} e_{\cdot(\cdot)}^2\right\} &= \frac{1}{AS} \mathbf{E}\{e_{\cdot(\cdot)}^2\} \\ &= \frac{1}{AS} \mathbf{E}\left\{\left(\sum e_{s(a)}\right)^2\right\} \\ &= \frac{1}{AS} \mathbf{E}\left\{\sum e_{s(a)}^2 + \sum_{a, s \neq a', s'} e_{s(a)} e_{s'(a')}\right\} \quad (\text{cf. Properties of sums}) \\ &= \frac{1}{AS} \left[\mathbf{E}\left\{\sum e_{s(a)}^2\right\} + \mathbf{E}\left\{\sum_{a, s \neq a', s'} e_{s(a)} e_{s'(a')}\right\} \right] \quad (\text{cf. Property 3}) \\ &= \frac{1}{AS} \left[\mathbf{E}\left\{\sum e_{s(a)}^2\right\} + \sum_{a, s \neq a', s'} \mathbf{E}\{e_{s(a)} e_{s'(a')}\} \right] \quad (\text{cf. Property 3}) \\ &= \frac{1}{AS} \left[\sum_a \sum_s \sigma_e^2 \right] + \left[\sum_{a, s \neq a', s'} \text{Cov}(e_{s(a)}, e_{s'(a')}) \right] \quad (\text{cf. Conditions 2 and 5}) \\ &= \sigma_e^2. \end{aligned} \quad (\text{F.41})$$

Therefore, the expected value of **1** is equal to:

$$\mathbf{E}\{\mathbf{1}\} = AS\mu^2 + \sigma_e^2. \quad (\text{F.42})$$

F.3.7 Expected value of the sums of squares

We first express the sums of squares with numbers in the square and then replace them by their expected values. Let us start with the sum of squares between.

$$\begin{aligned}
 \mathbf{E}\{SS_A\} &= \mathbf{E}\left\{\boxed{A} - \boxed{1}\right\} \\
 &= \mathbf{E}\left\{\boxed{A}\right\} - \mathbf{E}\left\{\boxed{1}\right\} \\
 &= AS\mu^2 + S(A-1)\vartheta_a^2 + A\sigma_a^2 - AS\mu^2 - \sigma_e^2 \\
 &= (A-1)\sigma_e^2 + S(A-1)\vartheta_a^2 \\
 &= (A-1)(\sigma_e^2 + S\vartheta_a^2). \tag{F.43}
 \end{aligned}$$

Likewise, the expected value of the sum of squares within is

$$\begin{aligned}
 \mathbf{E}\{SS_{S(A)}\} &= \mathbf{E}\left\{\boxed{AS} - \boxed{A}\right\} \\
 &= \mathbf{E}\left\{\boxed{AS}\right\} - \mathbf{E}\left\{\boxed{A}\right\} \\
 &= AS\mu^2 + S(A-1)\vartheta_a^2 + AS\sigma_a^2 - AS\mu^2 - S(A-1)\vartheta_a^2 - A\sigma_a^2 \\
 &= A(S-1)\sigma_e^2. \tag{F.44}
 \end{aligned}$$

F.3.8 Expected value of the mean squares

The expected value of the mean squares is obtained by dividing the expected value of the sums of squares by their respective numbers of degrees of freedom. The expected value of M_s between is

$$\begin{aligned}
 \mathbf{E}\{MS_A\} &= \frac{\mathbf{E}\{SS_A\}}{df_A} \\
 &= \frac{\mathbf{E}\{SS_A\}}{A-1} \\
 &= \frac{(A-1)(\sigma_e^2 + S\vartheta_a^2)}{A-1} \\
 &= \sigma_e^2 + S\vartheta_a^2. \tag{F.45}
 \end{aligned}$$

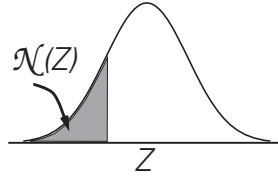
The expected value of M_s within is

$$\begin{aligned}
 \mathbf{E}\{MS_{S(A)}\} &= \frac{\mathbf{E}\{SS_{S(A)}\}}{df_{S(A)}} \\
 &= \frac{\mathbf{E}\{SS_{S(A)}\}}{A(S-1)} \\
 &= \frac{A(S-1)\sigma_e^2}{A(S-1)} \\
 &= \sigma_e^2. \tag{F.46}
 \end{aligned}$$

And finally we have the values given in Chapter 10. (By the way, are you still awake?)

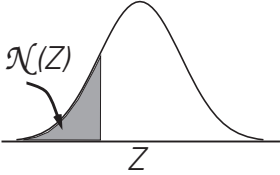
Statistical tables

Table 1	The standardized normal distribution	497
Table 2	Critical values of Fisher's F	499
Table 3	Fisher's Z transform	502
Table 4	Lilliefors test of normality	505
Table 5	Šidák's test	506
Table 6	Bonferroni's test	509
Table 7	Trend analysis: orthogonal polynomials	512
Table 8	Dunnett's test	513
Table 9	F_{range} distribution	514
Table 10	Duncan's test	515

Table 1 The standardized normal distribution $\mathcal{N}(0, 1)$.

Integral of normal distribution transformation: $\mathcal{N}(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^Z \exp\{-\frac{z^2}{2}\} dZ$													
Z	$\mathcal{N}(Z)$	Z	$\mathcal{N}(Z)$	Z	$\mathcal{N}(Z)$	Z	$\mathcal{N}(Z)$	Z	$\mathcal{N}(Z)$	Z	$\mathcal{N}(Z)$	Z	$\mathcal{N}(Z)$
-3.50	0.0002	-3.04	0.0012	-2.58	0.0049	-2.12	0.0170	-1.66	0.0485	-1.20	0.1151	-0.74	0.2296
-3.49	0.0002	-3.03	0.0012	-2.57	0.0051	-2.11	0.0174	-1.65	0.0495	-1.19	0.1170	-0.73	0.2327
-3.48	0.0003	-3.02	0.0013	-2.56	0.0052	-2.10	0.0179	-1.64	0.0505	-1.18	0.1190	-0.72	0.2358
-3.47	0.0003	-3.01	0.0013	-2.55	0.0054	-2.09	0.0183	-1.63	0.0516	-1.17	0.1210	-0.71	0.2389
-3.46	0.0003	-3.00	0.0013	-2.54	0.0055	-2.08	0.0188	-1.62	0.0526	-1.16	0.1230	-0.70	0.2420
-3.45	0.0003	-2.99	0.0014	-2.53	0.0057	-2.07	0.0192	-1.61	0.0537	-1.15	0.1251	-0.69	0.2451
-3.44	0.0003	-2.98	0.0014	-2.52	0.0059	-2.06	0.0197	-1.60	0.0548	-1.14	0.1271	-0.68	0.2483
-3.43	0.0003	-2.97	0.0015	-2.51	0.0060	-2.05	0.0202	-1.59	0.0559	-1.13	0.1292	-0.67	0.2514
-3.42	0.0003	-2.96	0.0015	-2.50	0.0062	-2.04	0.0207	-1.58	0.0571	-1.12	0.1314	-0.66	0.2546
-3.41	0.0003	-2.95	0.0016	-2.49	0.0064	-2.03	0.0212	-1.57	0.0582	-1.11	0.1335	-0.65	0.2578
-3.40	0.0003	-2.94	0.0016	-2.48	0.0066	-2.02	0.0217	-1.56	0.0594	-1.10	0.1357	-0.64	0.2611
-3.39	0.0003	-2.93	0.0017	-2.47	0.0068	-2.01	0.0222	-1.55	0.0606	-1.09	0.1379	-0.63	0.2643
-3.38	0.0004	-2.92	0.0018	-2.46	0.0069	-2.00	0.0228	-1.54	0.0618	-1.08	0.1401	-0.62	0.2676
-3.37	0.0004	-2.91	0.0018	-2.45	0.0071	-1.99	0.0233	-1.53	0.0630	-1.07	0.1423	-0.61	0.2709
-3.36	0.0004	-2.90	0.0019	-2.44	0.0073	-1.98	0.0239	-1.52	0.0643	-1.06	0.1446	-0.60	0.2743
-3.35	0.0004	-2.89	0.0019	-2.43	0.0075	-1.97	0.0244	-1.51	0.0655	-1.05	0.1469	-0.59	0.2776
-3.34	0.0004	-2.88	0.0020	-2.42	0.0078	-1.96	0.0250	-1.50	0.0668	-1.04	0.1492	-0.58	0.2810
-3.33	0.0004	-2.87	0.0021	-2.41	0.0080	-1.95	0.0256	-1.49	0.0681	-1.03	0.1515	-0.57	0.2843
-3.32	0.0005	-2.86	0.0021	-2.40	0.0082	-1.94	0.0262	-1.48	0.0694	-1.02	0.1539	-0.56	0.2877
-3.31	0.0005	-2.85	0.0022	-2.39	0.0084	-1.93	0.0268	-1.47	0.0708	-1.01	0.1562	-0.55	0.2912
-3.30	0.0005	-2.84	0.0023	-2.38	0.0087	-1.92	0.0274	-1.46	0.0721	-1.00	0.1587	-0.54	0.2946
-3.29	0.0005	-2.83	0.0023	-2.37	0.0089	-1.91	0.0281	-1.45	0.0735	-0.99	0.1611	-0.53	0.2981
-3.28	0.0005	-2.82	0.0024	-2.36	0.0091	-1.90	0.0287	-1.44	0.0749	-0.98	0.1635	-0.52	0.3015
-3.27	0.0005	-2.81	0.0025	-2.35	0.0094	-1.89	0.0294	-1.43	0.0764	-0.97	0.1660	-0.51	0.3050
-3.26	0.0006	-2.80	0.0026	-2.34	0.0096	-1.88	0.0301	-1.42	0.0778	-0.96	0.1685	-0.50	0.3085
-3.25	0.0006	-2.79	0.0026	-2.33	0.0099	-1.87	0.0307	-1.41	0.0793	-0.95	0.1711	-0.49	0.3121
-3.24	0.0006	-2.78	0.0027	-2.32	0.0102	-1.86	0.0314	-1.40	0.0808	-0.94	0.1736	-0.48	0.3156
-3.23	0.0006	-2.77	0.0028	-2.31	0.0104	-1.85	0.0322	-1.39	0.0823	-0.93	0.1762	-0.47	0.3192
-3.22	0.0006	-2.76	0.0029	-2.30	0.0107	-1.84	0.0329	-1.38	0.0838	-0.92	0.1788	-0.46	0.3228
-3.21	0.0007	-2.75	0.0030	-2.29	0.0110	-1.83	0.0336	-1.37	0.0853	-0.91	0.1814	-0.45	0.3264
-3.20	0.0007	-2.74	0.0031	-2.28	0.0113	-1.82	0.0344	-1.36	0.0869	-0.90	0.1841	-0.44	0.3300
-3.19	0.0007	-2.73	0.0032	-2.27	0.0116	-1.81	0.0351	-1.35	0.0885	-0.89	0.1867	-0.43	0.3336
-3.18	0.0007	-2.72	0.0033	-2.26	0.0119	-1.80	0.0359	-1.34	0.0901	-0.88	0.1894	-0.42	0.3372
-3.17	0.0008	-2.71	0.0034	-2.25	0.0122	-1.79	0.0367	-1.33	0.0918	-0.87	0.1922	-0.41	0.3409
-3.16	0.0008	-2.70	0.0035	-2.24	0.0125	-1.78	0.0375	-1.32	0.0934	-0.86	0.1949	-0.40	0.3446
-3.15	0.0008	-2.69	0.0036	-2.23	0.0129	-1.77	0.0384	-1.31	0.0951	-0.85	0.1977	-0.39	0.3483
-3.14	0.0008	-2.68	0.0037	-2.22	0.0132	-1.76	0.0392	-1.30	0.0968	-0.84	0.2005	-0.38	0.3520
-3.13	0.0009	-2.67	0.0038	-2.21	0.0136	-1.75	0.0401	-1.29	0.0985	-0.83	0.2033	-0.37	0.3557
-3.12	0.0009	-2.66	0.0039	-2.20	0.0139	-1.74	0.0409	-1.28	0.1003	-0.82	0.2061	-0.36	0.3594
-3.11	0.0009	-2.65	0.0040	-2.19	0.0143	-1.73	0.0418	-1.27	0.1020	-0.81	0.2090	-0.35	0.3632
-3.10	0.0010	-2.64	0.0041	-2.18	0.0146	-1.72	0.0427	-1.26	0.1038	-0.80	0.2119	-0.34	0.3669
-3.09	0.0010	-2.63	0.0043	-2.17	0.0150	-1.71	0.0436	-1.25	0.1056	-0.79	0.2148	-0.33	0.3707
-3.08	0.0010	-2.62	0.0044	-2.16	0.0154	-1.70	0.0446	-1.24	0.1075	-0.78	0.2177	-0.32	0.3745
-3.07	0.0011	-2.61	0.0045	-2.15	0.0158	-1.69	0.0455	-1.23	0.1093	-0.77	0.2206	-0.31	0.3783
-3.06	0.0011	-2.60	0.0047	-2.14	0.0162	-1.68	0.0465	-1.22	0.1112	-0.76	0.2236	-0.30	0.3821
-3.05	0.0011	-2.59	0.0048	-2.13	0.0166	-1.67	0.0475	-1.21	0.1131	-0.75	0.2266	-0.29	0.3859

Table 1 The standardized normal distribution $\mathcal{N}(0, 1)$. (continued)



Integral of normal distribution transformation: $\mathcal{N}(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^Z \exp\{-\frac{z^2}{2}\} dZ$													
Z	$\mathcal{N}(Z)$	Z	$\mathcal{N}(Z)$	Z	$\mathcal{N}(Z)$	Z	$\mathcal{N}(Z)$	Z	$\mathcal{N}(Z)$	Z	$\mathcal{N}(Z)$	Z	$\mathcal{N}(Z)$
0.18	0.5714	0.53	0.7019	0.98	0.8365	1.43	0.9236	1.88	0.9699	2.33	0.9901	2.78	0.9973
		0.54	0.7054	0.99	0.8389	1.44	0.9251	1.89	0.9706	2.34	0.9904	2.79	0.9974
		0.55	0.7088	1.00	0.8413	1.45	0.9265	1.90	0.9713	2.35	0.9906	2.80	0.9974
		0.56	0.7123	1.01	0.8438	1.46	0.9279	1.91	0.9719	2.36	0.9909	2.81	0.9975
		0.57	0.7157	1.02	0.8461	1.47	0.9292	1.92	0.9726	2.37	0.9911	2.82	0.9976
		0.58	0.7190	1.03	0.8485	1.48	0.9306	1.93	0.9732	2.38	0.9913	2.83	0.9977
		0.59	0.7224	1.04	0.8508	1.49	0.9319	1.94	0.9738	2.39	0.9916	2.84	0.9977
		0.60	0.7257	1.05	0.8531	1.50	0.9332	1.95	0.9744	2.40	0.9918	2.85	0.9978
		0.61	0.7291	1.06	0.8554	1.51	0.9345	1.96	0.9750	2.41	0.9920	2.86	0.9979
		0.62	0.7324	1.07	0.8577	1.52	0.9357	1.97	0.9756	2.42	0.9922	2.87	0.9979
		0.63	0.7357	1.08	0.8599	1.53	0.9370	1.98	0.9761	2.43	0.9925	2.88	0.9980
		0.64	0.7389	1.09	0.8621	1.54	0.9382	1.99	0.9767	2.44	0.9927	2.89	0.9981
		0.65	0.7422	1.10	0.8643	1.55	0.9394	2.00	0.9772	2.45	0.9929	2.90	0.9981
		0.66	0.7454	1.11	0.8665	1.56	0.9406	2.01	0.9778	2.46	0.9931	2.91	0.9982
		0.67	0.7486	1.12	0.8686	1.57	0.9418	2.02	0.9783	2.47	0.9932	2.92	0.9982
	0.23	0.5910	0.68	0.7517	1.13	0.8708	1.58	0.9429	2.03	0.9788	2.48	0.9934	2.93
0.24	0.5948	0.69	0.7549	1.14	0.8729	1.59	0.9441	2.04	0.9793	2.49	0.9936	2.94	0.9984
0.25	0.5987	0.70	0.7580	1.15	0.8749	1.60	0.9452	2.05	0.9798	2.50	0.9938	2.95	0.9984
0.26	0.6026	0.71	0.7611	1.16	0.8770	1.61	0.9463	2.06	0.9803	2.51	0.9940	2.96	0.9985
0.27	0.6064	0.72	0.7642	1.17	0.8790	1.62	0.9474	2.07	0.9808	2.52	0.9941	2.97	0.9985
0.28	0.6103	0.73	0.7673	1.18	0.8810	1.63	0.9484	2.08	0.9812	2.53	0.9943	2.98	0.9986
0.29	0.6141	0.74	0.7704	1.19	0.8830	1.64	0.9495	2.09	0.9817	2.54	0.9945	2.99	0.9986
0.30	0.6179	0.75	0.7734	1.20	0.8849	1.65	0.9505	2.10	0.9821	2.55	0.9946	3.00	0.9987
0.31	0.6217	0.76	0.7764	1.21	0.8869	1.66	0.9515	2.11	0.9826	2.56	0.9948	3.01	0.9987
0.32	0.6255	0.77	0.7794	1.22	0.8888	1.67	0.9525	2.12	0.9830	2.57	0.9949	3.02	0.9987
0.33	0.6293	0.78	0.7823	1.23	0.8907	1.68	0.9535	2.13	0.9834	2.58	0.9951	3.03	0.9988
0.34	0.6331	0.79	0.7852	1.24	0.8925	1.69	0.9545	2.14	0.9838	2.59	0.9952	3.04	0.9988
0.35	0.6368	0.80	0.7881	1.25	0.8944	1.70	0.9554	2.15	0.9842	2.60	0.9953	3.05	0.9989
0.36	0.6406	0.81	0.7910	1.26	0.8962	1.71	0.9564	2.16	0.9846	2.61	0.9955	3.06	0.9989
0.37	0.6443	0.82	0.7939	1.27	0.8980	1.72	0.9573	2.17	0.9850	2.62	0.9956	3.07	0.9989
0.38	0.6480	0.83	0.7967	1.28	0.8997	1.73	0.9582	2.18	0.9854	2.63	0.9957	3.08	0.9990
0.39	0.6517	0.84	0.7995	1.29	0.9015	1.74	0.9591	2.19	0.9857	2.64	0.9959	3.09	0.9990
0.40	0.6554	0.85	0.8023	1.30	0.9032	1.75	0.9599	2.20	0.9861	2.65	0.9960	3.10	0.9990
0.41	0.6591	0.86	0.8051	1.31	0.9049	1.76	0.9608	2.21	0.9864	2.66	0.9961	3.11	0.9991
0.42	0.6628	0.87	0.8078	1.32	0.9066	1.77	0.9616	2.22	0.9868	2.67	0.9962	3.12	0.9991
0.43	0.6664	0.88	0.8106	1.33	0.9082	1.78	0.9625	2.23	0.9871	2.68	0.9963	3.13	0.9991
0.44	0.6700	0.89	0.8133	1.34	0.9099	1.79	0.9633	2.24	0.9875	2.69	0.9964	3.14	0.9992
0.45	0.6736	0.90	0.8159	1.35	0.9115	1.80	0.9641	2.25	0.9878	2.70	0.9965	3.15	0.9992
0.46	0.6772	0.91	0.8186	1.36	0.9131	1.81	0.9649	2.26	0.9881	2.71	0.9966	3.16	0.9992
0.47	0.6808	0.92	0.8212	1.37	0.9147	1.82	0.9656	2.27	0.9884	2.72	0.9967	3.17	0.9992
0.48	0.6844	0.93	0.8238	1.38	0.9162	1.83	0.9664	2.28	0.9887	2.73	0.9968	3.18	0.9993
0.49	0.6879	0.94	0.8264	1.39	0.9177	1.84	0.9671	2.29	0.9890	2.74	0.9969	3.19	0.9993
0.50	0.6915	0.95	0.8289	1.40	0.9192	1.85	0.9678	2.30	0.9893	2.75	0.9970	3.20	0.9993
0.51	0.6950	0.96	0.8315	1.41	0.9207	1.86	0.9686	2.31	0.9896	2.76	0.9971	3.21	0.9993
0.52	0.6985	0.97	0.8340	1.42	0.9222	1.87	0.9693	2.32	0.9898	2.77	0.9972	3.22	0.9994

Table 2 Critical values of Fisher's F . $\alpha = .05$ $\alpha = .01$

v_2	v_1																							∞
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	60	75	100	200	500	
1	161 4052	199 4999	215 5403	224 5624	230 5763	233 5858	236 5928	238 5981	240 6022	241 6055	242 6083	243 6106	245 6142	246 6170	248 6208	249 6234	250 6260	251 6286	252 6313	252 6323	253 6334	253 6349	254 6362	
2	18.51 98.50	19.00 99.00	19.16 99.17	19.25 99.25	19.30 99.33	19.33 99.33	19.35 99.36	19.37 99.37	19.38 99.39	19.40 99.40	19.40 99.41	19.41 99.42	19.42 99.43	19.43 99.44	19.45 99.45	19.46 99.46	19.46 99.47	19.47 99.47	19.48 99.48	19.48 99.49	19.49 99.49	19.49 99.50		
3	10.13 34.12	9.55 30.82	9.28 29.46	9.12 28.71	9.01 28.24	8.94 27.91	8.89 27.67	8.85 27.49	8.81 27.35	8.79 27.23	8.76 27.13	8.74 27.05	8.71 26.92	8.69 26.83	8.66 26.69	8.64 26.60	8.62 26.50	8.59 26.41	8.57 26.32	8.56 26.28	8.55 26.24	8.54 26.18		
4	7.71 21.20	6.94 18.00	6.59 16.69	6.39 15.98	6.26 15.52	6.16 15.21	6.09 14.98	6.04 14.80	6.00 14.66	5.96 14.55	5.94 14.45	5.91 14.37	5.87 14.25	5.84 14.15	5.80 14.02	5.77 13.93	5.75 13.84	5.72 13.75	5.69 13.65	5.68 13.61	5.66 13.58	5.65 13.52		
5	6.61 16.26	5.79 13.27	5.41 12.06	5.19 11.39	5.05 10.97	4.95 10.67	4.88 10.46	4.82 10.29	4.77 10.16	4.74 10.05	4.70 9.96	4.68 9.89	4.64 9.77	4.60 9.68	4.56 9.55	4.53 9.47	4.50 9.38	4.46 9.29	4.43 9.20	4.42 9.17	4.41 9.13	4.39 9.04		
6	5.99 13.75	5.14 10.92	4.76 9.78	4.53 9.15	4.39 8.75	4.28 8.47	4.21 8.26	4.15 8.10	4.10 7.98	4.06 7.87	4.03 7.79	4.00 7.72	3.96 7.60	3.92 7.52	3.87 7.40	3.84 7.31	3.81 7.23	3.77 7.14	3.74 7.06	3.73 7.02	3.71 6.99	3.69 6.90		
7	5.59 12.25	4.74 9.95	4.35 8.85	4.12 8.45	3.97 8.15	3.87 7.87	3.79 7.69	3.73 7.54	3.68 7.42	3.64 7.32	3.60 7.24	3.57 7.17	3.53 7.09	3.49 7.01	3.44 6.93	3.41 6.86	3.38 6.79	3.34 6.72	3.30 6.65	3.29 6.62	3.27 6.59	3.25 6.56		
8	5.32 11.26	4.46 8.65	4.07 7.79	3.84 7.01	3.69 6.63	3.58 6.37	3.50 6.18	3.44 6.03	3.39 5.91	3.35 5.81	3.31 5.73	3.28 5.67	3.24 5.56	3.20 5.48	3.15 5.36	3.12 5.28	3.08 5.20	3.04 5.12	3.01 5.03	2.99 5.00	2.97 4.96	2.95 4.91		
9	5.12 10.56	4.26 8.02	3.86 6.99	3.63 6.42	3.48 6.06	3.37 5.80	3.29 5.61	3.23 5.47	3.18 5.35	3.14 5.26	3.10 5.18	3.07 5.11	3.03 5.01	2.99 4.92	2.94 4.81	2.90 4.73	2.86 4.65	2.83 4.57	2.79 4.48	2.77 4.45	2.76 4.41	2.73 4.36		
10	4.96 10.04	4.10 7.56	3.71 6.55	3.48 5.99	3.33 5.64	3.22 5.39	3.14 5.20	3.07 5.06	3.02 4.94	2.98 4.85	2.94 4.77	2.91 4.71	2.86 4.60	2.83 4.52	2.77 4.41	2.74 4.33	2.70 4.25	2.66 4.17	2.62 4.08	2.60 4.05	2.59 4.01	2.56 3.96		
11	4.84 9.65	3.98 7.21	3.59 6.22	3.36 5.67	3.20 5.32	3.09 5.07	3.01 4.89	2.95 4.74	2.90 4.63	2.85 4.54	2.82 4.46	2.79 4.40	2.74 4.29	2.70 4.21	2.65 4.10	2.61 4.02	2.57 3.94	2.53 3.86	2.49 3.78	2.47 3.74	2.46 3.71	2.42 3.66		
12	4.75 9.33	3.89 6.93	3.49 5.95	3.26 5.41	3.11 5.06	3.00 4.82	2.91 4.64	2.85 4.50	2.80 4.39	2.75 4.30	2.72 4.22	2.69 4.16	2.64 4.05	2.60 3.97	2.54 3.86	2.51 3.78	2.47 3.70	2.43 3.62	2.38 3.54	2.37 3.50	2.35 3.47	2.31 3.38		
13	4.67 9.07	3.81 6.70	3.41 5.74	3.18 5.21	3.03 4.86	2.92 4.62	2.83 4.44	2.77 4.30	2.71 4.19	2.67 4.10	2.63 4.02	2.60 3.96	2.55 3.86	2.51 3.78	2.46 3.66	2.42 3.59	2.38 3.51	2.34 3.43	2.30 3.34	2.28 3.31	2.26 3.27	2.23 3.19		
14	4.60 8.86	3.74 6.51	3.34 5.56	3.11 5.04	2.96 4.69	2.85 4.46	2.76 4.28	2.70 4.14	2.65 4.03	2.60 3.94	2.57 3.86	2.53 3.80	2.48 3.70	2.44 3.62	2.39 3.51	2.35 3.43	2.31 3.35	2.27 3.27	2.22 3.18	2.21 3.15	2.19 3.11	2.16 3.06		
15	4.54 8.68	3.68 6.36	3.29 5.42	3.06 4.89	2.90 4.56	2.79 4.32	2.71 4.14	2.64 4.00	2.59 3.89	2.54 3.80	2.51 3.73	2.48 3.67	2.42 3.56	2.38 3.49	2.33 3.37	2.29 3.29	2.25 3.21	2.20 3.13	2.16 3.05	2.14 3.01	2.12 2.98	2.10 2.92		
16	4.49 8.53	3.63 6.23	3.24 5.29	3.01 4.77	2.85 4.44	2.74 4.20	2.66 4.03	2.59 3.89	2.54 3.78	2.49 3.69	2.46 3.62	2.42 3.55	2.37 3.45	2.33 3.37	2.28 3.26	2.24 3.18	2.19 3.10	2.15 3.02	2.11 2.93	2.09 2.90	2.07 2.86	2.04 2.81		

Table 2 Critical values of Fisher's *F* (continued), $\alpha = .05$ $\alpha = .01$

		ν_1																								
ν_2		1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	60	75	100	200	500	∞	
17	4.45 8.40	3.59 6.11	3.20 5.18	2.96 4.67	2.81 4.34	2.70 4.10	2.61 3.93	2.55 3.79	2.49 3.68	2.45 3.59	2.45 3.52	2.41 3.46	2.38 3.46	2.33 3.35	2.29 3.27	2.23 3.16	2.19 3.08	2.15 3.00	2.10 2.92	2.06 2.83	2.04 2.80	2.02 2.76	1.99 2.71	1.97 2.68	1.96 2.65	
18	4.41 8.29	3.55 6.01	3.16 5.09	2.93 4.58	2.77 4.25	2.66 4.01	2.58 3.84	2.51 3.71	2.46 3.60	2.41 3.51	2.41 3.43	2.37 3.37	2.34 3.37	2.29 3.27	2.25 3.19	2.19 3.08	2.15 3.00	2.11 2.92	2.06 2.84	2.02 2.75	2.00 2.71	1.98 2.68	1.95 2.62	1.93 2.59	1.92 2.57	
19	4.38 8.18	3.52 5.93	3.13 5.01	2.90 4.50	2.74 4.17	2.63 3.94	2.54 3.77	2.48 3.63	2.42 3.52	2.38 3.43	2.38 3.43	2.34 3.36	2.31 3.30	2.26 3.19	2.21 3.12	2.16 3.00	2.11 2.92	2.07 2.84	2.03 2.76	1.98 2.67	1.96 2.64	1.94 2.60	1.91 2.55	1.89 2.51	1.88 2.49	
20	4.35 8.10	3.49 5.85	3.10 4.94	2.87 4.43	2.71 4.10	2.60 3.87	2.51 3.70	2.45 3.56	2.39 3.46	2.35 3.37	2.35 3.29	2.31 3.23	2.28 3.23	2.22 3.13	2.18 3.05	2.12 2.94	2.08 2.86	2.04 2.78	1.99 2.69	1.95 2.61	1.93 2.57	1.91 2.54	1.88 2.48	1.86 2.44	1.84 2.42	
21	4.32 8.02	3.47 5.78	3.07 4.87	2.84 4.37	2.68 4.04	2.57 3.81	2.57 3.64	2.49 3.51	2.42 3.40	2.37 3.31	2.32 3.24	2.28 3.17	2.25 3.07	2.20 2.99	2.16 2.99	2.10 2.88	2.05 2.80	2.01 2.72	1.96 2.64	1.92 2.55	1.90 2.51	1.88 2.48	1.84 2.42	1.83 2.38	1.81 2.36	
22	4.30 7.95	3.44 5.72	3.05 4.82	2.82 4.31	2.66 3.99	2.55 3.76	2.55 3.59	2.46 3.45	2.40 3.35	2.34 3.26	2.30 3.18	2.26 3.12	2.23 3.02	2.17 2.94	2.13 2.94	2.07 2.83	2.03 2.75	1.98 2.67	1.94 2.58	1.89 2.50	1.87 2.46	1.85 2.42	1.82 2.36	1.80 2.33	1.78 2.31	
23	4.28 7.88	3.42 5.66	3.03 4.76	2.80 4.26	2.64 3.94	2.53 3.71	2.53 3.54	2.44 3.41	2.37 3.30	2.32 3.21	2.27 3.14	2.24 3.07	2.20 2.96	2.15 2.97	2.11 2.89	2.05 2.78	2.01 2.70	1.96 2.62	1.91 2.54	1.86 2.45	1.84 2.41	1.82 2.37	1.79 2.32	1.77 2.28	1.76 2.26	
24	4.26 7.82	3.40 5.61	3.01 4.72	2.78 4.22	2.62 3.90	2.51 3.67	2.51 3.50	2.42 3.36	2.36 3.26	2.30 3.17	2.25 3.09	2.22 3.03	2.18 2.93	2.13 2.93	2.09 2.85	2.03 2.74	1.98 2.66	1.94 2.58	1.89 2.49	1.84 2.40	1.82 2.37	1.80 2.33	1.77 2.27	1.75 2.24	1.73 2.21	
25	4.24 7.77	3.39 5.57	2.99 4.68	2.76 4.18	2.60 3.85	2.49 3.63	2.49 3.46	2.40 3.32	2.34 3.22	2.24 3.13	2.20 3.06	2.16 2.99	2.12 2.92	2.11 2.89	2.07 2.81	2.01 2.70	1.96 2.62	1.92 2.54	1.87 2.45	1.82 2.36	1.80 2.33	1.78 2.29	1.75 2.23	1.73 2.19	1.71 2.17	
26	4.23 7.72	3.37 5.53	2.98 4.64	2.74 4.14	2.59 3.82	2.47 3.59	2.47 3.42	2.39 3.29	2.32 3.18	2.22 3.09	2.18 3.02	2.15 2.96	2.12 2.90	2.09 2.86	2.05 2.78	1.99 2.66	1.95 2.58	1.90 2.50	1.85 2.42	1.80 2.33	1.78 2.29	1.76 2.25	1.73 2.19	1.71 2.16	1.69 2.13	
27	4.21 7.68	3.35 5.49	2.96 4.60	2.73 4.11	2.57 3.78	2.46 3.56	2.46 3.39	2.37 3.26	2.31 3.15	2.25 3.06	2.20 2.99	2.17 2.93	2.13 2.82	2.08 2.75	2.04 2.75	1.97 2.63	1.93 2.55	1.88 2.47	1.84 2.38	1.79 2.29	1.76 2.26	1.74 2.22	1.71 2.16	1.69 2.12	1.67 2.10	
28	4.20 7.64	3.34 5.45	2.95 4.57	2.71 4.07	2.56 3.75	2.45 3.53	2.45 3.36	2.36 3.23	2.29 3.12	2.24 3.03	2.19 2.96	2.15 2.90	2.12 2.87	2.06 2.79	2.02 2.72	1.96 2.60	1.91 2.52	1.87 2.44	1.82 2.35	1.77 2.26	1.75 2.23	1.73 2.19	1.69 2.13	1.67 2.09	1.65 2.07	
29	4.18 7.60	3.33 5.42	2.93 4.54	2.70 4.04	2.55 3.73	2.43 3.50	2.43 3.33	2.35 3.20	2.28 3.09	2.22 3.00	2.18 2.93	2.14 2.87	2.10 2.82	2.05 2.77	2.01 2.69	1.94 2.57	1.90 2.49	1.85 2.41	1.81 2.33	1.75 2.23	1.73 2.20	1.71 2.16	1.67 2.10	1.65 2.06	1.64 2.04	
30	4.17 7.56	3.32 5.39	2.92 4.51	2.69 4.02	2.53 3.70	2.42 3.47	2.42 3.30	2.37 3.17	2.27 3.07	2.21 2.98	2.16 2.91	2.13 2.84	2.09 2.80	2.04 2.74	1.99 2.66	1.93 2.55	1.89 2.47	1.84 2.39	1.79 2.30	1.74 2.21	1.72 2.17	1.70 2.13	1.66 2.07	1.64 2.03	1.62 2.01	
32	4.15 7.50	3.29 5.34	2.90 4.46	2.67 3.97	2.51 3.65	2.40 3.43	2.40 3.26	2.34 3.13	2.24 3.02	2.19 2.93	2.14 2.86	2.10 2.76	2.07 2.80	2.01 2.70	1.97 2.62	1.91 2.50	1.86 2.42	1.82 2.34	1.77 2.25	1.71 2.16	1.69 2.12	1.67 2.08	1.63 2.02	1.61 1.98	1.60 1.96	
34	4.13 7.44	3.28 5.29	2.88 4.42	2.65 3.93	2.49 3.61	2.38 3.39	2.38 3.22	2.29 3.09	2.23 2.98	2.17 2.89	2.12 2.82	2.08 2.76	2.05 2.70	1.99 2.66	1.95 2.58	1.89 2.46	1.84 2.38	1.80 2.30	1.75 2.21	1.69 2.12	1.67 2.08	1.65 2.04	1.61 1.98	1.59 1.94	1.57 1.91	

36	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.15	2.11	2.07	2.03	1.98	1.93	1.87	1.82	1.78	1.73	1.67	1.65	1.62	1.59	1.56	1.55
	7.40	5.25	4.38	3.89	3.57	3.35	3.18	3.05	2.95	2.86	2.79	2.72	2.62	2.54	2.43	2.35	2.26	2.18	2.08	2.04	2.00	1.94	1.90	1.87
38	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.96	1.92	1.85	1.81	1.76	1.71	1.65	1.63	1.61	1.57	1.54	1.53
	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.92	2.83	2.75	2.69	2.59	2.51	2.40	2.32	2.23	2.14	2.05	2.01	1.97	1.90	1.86	1.84
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.95	1.90	1.84	1.79	1.74	1.69	1.64	1.61	1.59	1.55	1.53	1.51
	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.56	2.48	2.37	2.29	2.20	2.11	2.02	1.98	1.94	1.87	1.83	1.81
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.03	1.99	1.94	1.89	1.83	1.78	1.73	1.68	1.62	1.60	1.57	1.53	1.51	1.49
	7.28	5.15	4.29	3.80	3.49	3.27	3.10	2.97	2.86	2.78	2.70	2.64	2.54	2.46	2.34	2.26	2.18	2.09	1.99	1.95	1.91	1.85	1.80	1.78
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98	1.92	1.88	1.81	1.77	1.72	1.67	1.61	1.59	1.56	1.52	1.49	1.48
	7.25	5.12	4.26	3.78	3.47	3.24	3.08	2.95	2.84	2.75	2.68	2.62	2.52	2.44	2.32	2.24	2.15	2.07	1.97	1.93	1.89	1.82	1.78	1.75
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.15	2.09	2.04	2.00	1.97	1.91	1.87	1.80	1.76	1.71	1.65	1.60	1.57	1.55	1.51	1.48	1.46
	7.22	5.10	4.24	3.76	3.44	3.22	3.06	2.93	2.82	2.73	2.66	2.60	2.50	2.42	2.30	2.22	2.13	2.04	1.95	1.91	1.86	1.80	1.76	1.73
48	4.04	3.19	2.80	2.57	2.41	2.29	2.21	2.14	2.08	2.03	1.99	1.96	1.90	1.86	1.79	1.75	1.70	1.64	1.59	1.56	1.54	1.49	1.47	1.45
	7.19	5.08	4.22	3.74	3.43	3.20	3.04	2.91	2.80	2.71	2.64	2.58	2.48	2.40	2.28	2.20	2.12	2.02	1.93	1.89	1.84	1.78	1.73	1.71
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.99	1.95	1.89	1.85	1.78	1.74	1.69	1.63	1.58	1.55	1.52	1.48	1.46	1.44
	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.63	2.56	2.46	2.38	2.27	2.18	2.10	2.01	1.91	1.87	1.82	1.76	1.71	1.68
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.86	1.82	1.75	1.70	1.65	1.59	1.53	1.51	1.48	1.44	1.41	1.39
	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.39	2.31	2.20	2.12	2.03	1.94	1.84	1.79	1.75	1.68	1.63	1.60
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.03	1.98	1.94	1.90	1.85	1.80	1.73	1.69	1.63	1.58	1.52	1.49	1.46	1.42	1.39	1.37
	7.04	4.95	4.10	3.62	3.31	3.09	2.93	2.80	2.69	2.61	2.53	2.47	2.37	2.29	2.17	2.09	2.00	1.91	1.81	1.77	1.72	1.65	1.60	1.57
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.93	1.89	1.84	1.79	1.72	1.67	1.62	1.57	1.50	1.48	1.45	1.40	1.37	1.35
	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67	2.59	2.51	2.45	2.35	2.27	2.15	2.07	1.98	1.89	1.78	1.74	1.70	1.62	1.57	1.54
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.91	1.88	1.82	1.77	1.70	1.65	1.60	1.54	1.48	1.45	1.43	1.38	1.35	1.33
	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55	2.48	2.42	2.31	2.23	2.12	2.03	1.94	1.85	1.75	1.70	1.65	1.58	1.53	1.50
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.89	1.85	1.79	1.75	1.68	1.63	1.57	1.52	1.45	1.42	1.39	1.34	1.31	1.28
	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.43	2.37	2.27	2.19	2.07	1.98	1.89	1.80	1.69	1.65	1.60	1.52	1.47	1.43
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.96	1.91	1.87	1.83	1.77	1.73	1.66	1.60	1.55	1.49	1.42	1.40	1.36	1.31	1.27	1.25
	6.84	4.78	3.94	3.47	3.17	2.95	2.79	2.66	2.55	2.47	2.39	2.33	2.23	2.15	2.03	1.94	1.85	1.76	1.65	1.60	1.55	1.47	1.41	1.37
150	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82	1.76	1.71	1.64	1.59	1.54	1.48	1.41	1.38	1.34	1.29	1.25	1.22
	6.81	4.75	3.91	3.45	3.14	2.92	2.76	2.63	2.53	2.44	2.37	2.31	2.20	2.12	2.00	1.92	1.83	1.73	1.62	1.57	1.52	1.43	1.38	1.33
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.84	1.80	1.74	1.69	1.62	1.57	1.52	1.46	1.39	1.35	1.32	1.26	1.22	1.19
	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.34	2.27	2.17	2.09	1.97	1.89	1.79	1.69	1.58	1.53	1.48	1.39	1.33	1.28
400	3.86	3.02	2.63	2.39	2.24	2.12	2.03	1.96	1.90	1.85	1.81	1.78	1.72	1.67	1.60	1.54	1.49	1.42	1.35	1.32	1.28	1.22	1.17	1.13
	6.70	4.66	3.83	3.37	3.06	2.85	2.68	2.56	2.45	2.37	2.29	2.23	2.13	2.05	1.92	1.84	1.75	1.64	1.53	1.48	1.42	1.32	1.25	1.19
1000	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.80	1.76	1.70	1.65	1.58	1.53	1.47	1.41	1.33	1.30	1.26	1.19	1.13	1.08
	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34	2.27	2.20	2.10	2.02	1.90	1.81	1.72	1.61	1.50	1.44	1.38	1.28	1.19	1.11
5000	3.84	3.00	2.61	2.37	2.22	2.10	2.01	1.94	1.88	1.83	1.79	1.75	1.69	1.65	1.57	1.52	1.46	1.40	1.32	1.29	1.25	1.17	1.11	1.00
	6.64	4.61	3.79	3.32	3.02	2.81	2.64	2.51	2.41	2.32	2.25	2.19	2.09	2.00	1.88	1.79	1.70	1.60	1.48	1.42	1.36	1.25	1.15	1.00

Table 3 Fisher's Z transform.

$Z = \frac{1}{2} [\ln(1 + r) - \ln(1 - r)]$										$r = \frac{\exp \{2 \times Z\} - 1}{\exp \{2 \times Z\} + 1}$									
<i>r</i>		<i>Z</i>		<i>r</i>		<i>Z</i>		<i>r</i>		<i>Z</i>		<i>r</i>		<i>Z</i>		<i>r</i>		<i>Z</i>	
0.001	0.0010	0.043	0.0430	0.086	0.0862	0.129	0.1297	0.172	0.1737	0.215	0.2184	0.258	0.2640	0.301	0.3106				
0.002	0.0020	0.044	0.0440	0.087	0.0872	0.130	0.1307	0.173	0.1748	0.216	0.2195	0.259	0.2650	0.302	0.3117				
0.003	0.0030	0.045	0.0450	0.088	0.0882	0.131	0.1318	0.174	0.1758	0.217	0.2205	0.260	0.2661	0.303	0.3128				
0.004	0.0040	0.046	0.0460	0.089	0.0892	0.132	0.1328	0.175	0.1768	0.218	0.2216	0.261	0.2672	0.304	0.3139				
0.005	0.0050	0.047	0.0470	0.090	0.0902	0.133	0.1338	0.176	0.1779	0.219	0.2226	0.262	0.2683	0.305	0.3150				
0.006	0.0060	0.048	0.0480	0.091	0.0913	0.134	0.1348	0.177	0.1789	0.220	0.2237	0.263	0.2693	0.306	0.3161				
0.007	0.0070	0.049	0.0490	0.092	0.0923	0.135	0.1358	0.178	0.1799	0.221	0.2247	0.264	0.2704	0.307	0.3172				
0.008	0.0080	0.050	0.0500	0.093	0.0933	0.136	0.1368	0.179	0.1809	0.222	0.2258	0.265	0.2715	0.308	0.3183				
0.009	0.0090	0.051	0.0510	0.094	0.0943	0.137	0.1379	0.180	0.1820	0.223	0.2268	0.266	0.2726	0.309	0.3194				
0.010	0.0100	0.052	0.0520	0.095	0.0953	0.138	0.1389	0.181	0.1830	0.224	0.2279	0.267	0.2736	0.310	0.3205				
0.011	0.0110	0.053	0.0530	0.096	0.0963	0.139	0.1399	0.182	0.1841	0.225	0.2289	0.268	0.2747	0.311	0.3217				
0.012	0.0120	0.054	0.0541	0.097	0.0973	0.140	0.1409	0.183	0.1851	0.226	0.2300	0.269	0.2758	0.312	0.3228				
0.013	0.0130	0.055	0.0551	0.098	0.0983	0.141	0.1419	0.184	0.1861	0.227	0.2310	0.270	0.2769	0.313	0.3239				
0.014	0.0140	0.056	0.0561	0.099	0.0993	0.142	0.1430	0.185	0.1872	0.228	0.2321	0.271	0.2779	0.314	0.3250				
0.015	0.0150	0.057	0.0571	0.100	0.1003	0.143	0.1440	0.186	0.1882	0.229	0.2331	0.272	0.2790	0.315	0.3261				
0.016	0.0160	0.058	0.0581	0.101	0.1013	0.144	0.1450	0.187	0.1892	0.230	0.2342	0.273	0.2801	0.316	0.3272				
0.017	0.0170	0.059	0.0591	0.102	0.1024	0.145	0.1460	0.188	0.1903	0.231	0.2352	0.274	0.2812	0.317	0.3283				
0.018	0.0180	0.060	0.0601	0.103	0.1034	0.146	0.1471	0.189	0.1913	0.232	0.2363	0.275	0.2823	0.318	0.3294				
0.019	0.0190	0.061	0.0611	0.104	0.1044	0.147	0.1481	0.190	0.1923	0.233	0.2374	0.276	0.2833	0.319	0.3305				
0.020	0.0200	0.062	0.0621	0.105	0.1054	0.148	0.1491	0.191	0.1934	0.234	0.2384	0.277	0.2844	0.320	0.3316				
0.021	0.0210	0.063	0.0631	0.106	0.1064	0.149	0.1501	0.192	0.1944	0.235	0.2395	0.278	0.2855	0.321	0.3328				
0.022	0.0220	0.064	0.0641	0.107	0.1074	0.150	0.1511	0.193	0.1955	0.236	0.2405	0.279	0.2866	0.322	0.3339				
0.023	0.0230	0.065	0.0651	0.108	0.1084	0.151	0.1522	0.194	0.1965	0.237	0.2416	0.280	0.2877	0.323	0.3350				
0.024	0.0240	0.066	0.0661	0.109	0.1094	0.152	0.1532	0.195	0.1975	0.238	0.2427	0.281	0.2888	0.324	0.3361				
0.025	0.0250	0.067	0.0671	0.110	0.1104	0.153	0.1542	0.196	0.1986	0.239	0.2437	0.282	0.2899	0.325	0.3372				
0.026	0.0260	0.068	0.0681	0.111	0.1115	0.154	0.1552	0.197	0.1996	0.240	0.2448	0.283	0.2909	0.326	0.3383				
0.027	0.0270	0.069	0.0691	0.112	0.1125	0.155	0.1563	0.198	0.2007	0.241	0.2458	0.284	0.2920	0.327	0.3395				
0.028	0.0280	0.070	0.0701	0.113	0.1135	0.156	0.1573	0.199	0.2017	0.242	0.2469	0.285	0.2931	0.328	0.3406				
0.029	0.0290	0.071	0.0711	0.114	0.1145	0.157	0.1583	0.200	0.2027	0.243	0.2480	0.286	0.2942	0.329	0.3417				
0.030	0.0300	0.072	0.0721	0.115	0.1155	0.158	0.1593	0.201	0.2038	0.244	0.2490	0.287	0.2953	0.330	0.3428				
0.031	0.0310	0.073	0.0731	0.116	0.1165	0.159	0.1604	0.202	0.2048	0.245	0.2501	0.288	0.2964	0.331	0.3440				
0.032	0.0320	0.074	0.0741	0.117	0.1175	0.160	0.1614	0.203	0.2059	0.246	0.2512	0.289	0.2975	0.332	0.3451				
0.033	0.0330	0.075	0.0751	0.118	0.1186	0.161	0.1624	0.204	0.2069	0.247	0.2522	0.290	0.2986	0.333	0.3462				
0.034	0.0340	0.076	0.0761	0.119	0.1196	0.162	0.1634	0.205	0.2079	0.248	0.2533	0.291	0.2997	0.334	0.3473				
0.035	0.0350	0.077	0.0772	0.120	0.1206	0.163	0.1645	0.206	0.2090	0.249	0.2543	0.292	0.3008	0.335	0.3484				
0.036	0.0360	0.078	0.0782	0.121	0.1216	0.164	0.1655	0.207	0.2100	0.250	0.2554	0.293	0.3018	0.336	0.3496				
0.037	0.0370	0.079	0.0792	0.122	0.1226	0.165	0.1665	0.208	0.2111	0.251	0.2565	0.294	0.3029	0.337	0.3507				
0.038	0.0380	0.080	0.0802	0.123	0.1236	0.166	0.1676	0.209	0.2121	0.252	0.2575	0.295	0.3040	0.338	0.3518				
0.039	0.0390	0.081	0.0812	0.124	0.1246	0.167	0.1686	0.210	0.2132	0.253	0.2586	0.296	0.3051	0.339	0.3530				
0.040	0.0400	0.082	0.0822	0.125	0.1257	0.168	0.1696	0.211	0.2142	0.254	0.2597	0.297	0.3062	0.340	0.3541				
0.041	0.0410	0.083	0.0832	0.126	0.1267	0.169	0.1706	0.212	0.2153	0.255	0.2608	0.298	0.3073	0.341	0.3552				
0.042	0.0420	0.084	0.0842	0.127	0.1277	0.170	0.1717	0.213	0.2163	0.256	0.2618	0.299	0.3084	0.342	0.3564				
0.344	0.3586	0.085	0.0852	0.128	0.1287	0.171	0.1727	0.214	0.2174	0.257	0.2629	0.300	0.3095	0.343	0.3575				

Table 3 Fisher's Z transform (*continued*).

$Z = \frac{1}{2} [\ln(1 + r) - \ln(1 - r)]$												$r = \frac{\exp [2 \times Z] - 1}{\exp [2 \times Z] + 1}$			
r	Z	r	Z	r	Z	r	Z	r	Z	r	Z	r	Z	r	Z
0.345	0.3598	0.387	0.4083	0.430	0.4599	0.473	0.5139	0.516	0.5709	0.559	0.6314	0.602	0.6963	0.645	0.7667
0.346	0.3609	0.388	0.4094	0.431	0.4611	0.474	0.5152	0.517	0.5722	0.560	0.6328	0.603	0.6978	0.646	0.7684
0.347	0.3620	0.389	0.4106	0.432	0.4624	0.475	0.5165	0.518	0.5736	0.561	0.6343	0.604	0.6994	0.647	0.7701
0.348	0.3632	0.390	0.4118	0.433	0.4636	0.476	0.5178	0.519	0.5750	0.562	0.6358	0.605	0.7010	0.648	0.7718
0.349	0.3643	0.391	0.4130	0.434	0.4648	0.477	0.5191	0.520	0.5763	0.563	0.6372	0.606	0.7026	0.649	0.7736
0.350	0.3654	0.392	0.4142	0.435	0.4660	0.478	0.5204	0.521	0.5777	0.564	0.6387	0.607	0.7042	0.650	0.7753
0.351	0.3666	0.393	0.4153	0.436	0.4673	0.479	0.5217	0.522	0.5791	0.565	0.6401	0.608	0.7057	0.651	0.7770
0.352	0.3677	0.394	0.4165	0.437	0.4685	0.480	0.5230	0.523	0.5805	0.566	0.6416	0.609	0.7073	0.652	0.7788
0.353	0.3689	0.395	0.4177	0.438	0.4698	0.481	0.5243	0.524	0.5818	0.567	0.6431	0.610	0.7089	0.653	0.7805
0.354	0.3700	0.396	0.4189	0.439	0.4710	0.482	0.5256	0.525	0.5832	0.568	0.6446	0.611	0.7105	0.654	0.7823
0.355	0.3712	0.397	0.4201	0.440	0.4722	0.483	0.5269	0.526	0.5846	0.569	0.6460	0.612	0.7121	0.655	0.7840
0.356	0.3723	0.398	0.4213	0.441	0.4735	0.484	0.5282	0.527	0.5860	0.570	0.6475	0.613	0.7137	0.656	0.7858
0.357	0.3734	0.399	0.4225	0.442	0.4747	0.485	0.5295	0.528	0.5874	0.571	0.6490	0.614	0.7153	0.657	0.7875
0.358	0.3746	0.400	0.4236	0.443	0.4760	0.486	0.5308	0.529	0.5888	0.572	0.6505	0.615	0.7169	0.658	0.7893
0.359	0.3757	0.401	0.4248	0.444	0.4772	0.487	0.5321	0.530	0.5901	0.573	0.6520	0.616	0.7185	0.659	0.7910
0.360	0.3769	0.402	0.4260	0.445	0.4784	0.488	0.5334	0.531	0.5915	0.574	0.6535	0.617	0.7201	0.660	0.7928
0.361	0.3780	0.403	0.4272	0.446	0.4797	0.489	0.5347	0.532	0.5929	0.575	0.6550	0.618	0.7218	0.661	0.7946
0.362	0.3792	0.404	0.4284	0.447	0.4809	0.490	0.5361	0.533	0.5943	0.576	0.6565	0.619	0.7234	0.662	0.7964
0.363	0.3803	0.405	0.4296	0.448	0.4822	0.491	0.5374	0.534	0.5957	0.577	0.6580	0.620	0.7250	0.663	0.7981
0.364	0.3815	0.406	0.4308	0.449	0.4834	0.492	0.5387	0.535	0.5971	0.578	0.6595	0.621	0.7266	0.664	0.7999
0.365	0.3826	0.407	0.4320	0.450	0.4847	0.493	0.5400	0.536	0.5985	0.579	0.6610	0.622	0.7283	0.665	0.8017
0.366	0.3838	0.408	0.4332	0.451	0.4860	0.494	0.5413	0.537	0.5999	0.580	0.6625	0.623	0.7299	0.666	0.8035
0.367	0.3850	0.409	0.4344	0.452	0.4872	0.495	0.5427	0.538	0.6013	0.581	0.6640	0.624	0.7315	0.667	0.8053
0.368	0.3861	0.410	0.4356	0.453	0.4885	0.496	0.5440	0.539	0.6027	0.582	0.6655	0.625	0.7332	0.668	0.8071
0.369	0.3873	0.411	0.4368	0.454	0.4897	0.497	0.5453	0.540	0.6042	0.583	0.6670	0.626	0.7348	0.669	0.8089
0.370	0.3884	0.412	0.4380	0.455	0.4910	0.498	0.5466	0.541	0.6056	0.584	0.6685	0.627	0.7365	0.670	0.8107
0.371	0.3896	0.413	0.4392	0.456	0.4922	0.499	0.5480	0.542	0.6070	0.585	0.6700	0.628	0.7381	0.671	0.8126
0.372	0.3907	0.414	0.4404	0.457	0.4935	0.500	0.5493	0.543	0.6084	0.586	0.6716	0.629	0.7398	0.672	0.8144
0.373	0.3919	0.415	0.4416	0.458	0.4948	0.501	0.5506	0.544	0.6098	0.587	0.6731	0.630	0.7414	0.673	0.8162
0.374	0.3931	0.416	0.4428	0.459	0.4960	0.502	0.5520	0.545	0.6112	0.588	0.6746	0.631	0.7431	0.674	0.8180
0.375	0.3942	0.417	0.4441	0.460	0.4973	0.503	0.5533	0.546	0.6127	0.589	0.6761	0.632	0.7447	0.675	0.8199
0.376	0.3954	0.418	0.4453	0.461	0.4986	0.504	0.5547	0.547	0.6141	0.590	0.6777	0.633	0.7464	0.676	0.8217
0.377	0.3966	0.419	0.4465	0.462	0.4999	0.505	0.5560	0.548	0.6155	0.591	0.6792	0.634	0.7481	0.677	0.8236
0.378	0.3977	0.420	0.4477	0.463	0.5011	0.506	0.5573	0.549	0.6169	0.592	0.6807	0.635	0.7498	0.678	0.8254
0.379	0.3989	0.421	0.4489	0.464	0.5024	0.507	0.5587	0.550	0.6184	0.593	0.6823	0.636	0.7514	0.679	0.8273
0.380	0.4001	0.422	0.4501	0.465	0.5037	0.508	0.5600	0.551	0.6198	0.594	0.6838	0.637	0.7531	0.680	0.8291
0.381	0.4012	0.423	0.4513	0.466	0.5049	0.509	0.5614	0.552	0.6213	0.595	0.6854	0.638	0.7548	0.681	0.8310
0.382	0.4024	0.424	0.4526	0.467	0.5062	0.510	0.5627	0.553	0.6227	0.596	0.6869	0.639	0.7565	0.682	0.8328
0.383	0.4036	0.425	0.4538	0.468	0.5075	0.511	0.5641	0.554	0.6241	0.597	0.6885	0.640	0.7582	0.683	0.8347
0.384	0.4047	0.426	0.4550	0.469	0.5088	0.512	0.5654	0.555	0.6256	0.598	0.6900	0.641	0.7599	0.684	0.8366
0.385	0.4059	0.427	0.4562	0.470	0.5101	0.513	0.5668	0.556	0.6270	0.599	0.6916	0.642	0.7616	0.685	0.8385
0.386	0.4071	0.428	0.4574	0.471	0.5114	0.514	0.5682	0.557	0.6285	0.600	0.6931	0.643	0.7633	0.686	0.8404
0.688	0.8441	0.429	0.4587	0.472	0.5126	0.515	0.5695	0.558	0.6299	0.601	0.6947	0.644	0.7650	0.687	0.8423

Table 3 Fisher's Z transform (continued).

$Z = \frac{1}{2} [\ln(1+r) - \ln(1-r)]$										$r = \frac{\exp\{2 \times Z\} - 1}{\exp\{2 \times Z\} + 1}$									
r	Z	r	Z	r	Z	r	Z	r	Z	r	Z	r	Z	r	Z	r	Z	r	Z
0.689	0.8460	0.731	0.9309	0.774	1.0302	0.817	1.1477	0.860	1.2933	0.903	1.4882	0.946	1.7923	0.989	2.5987				
0.690	0.8480	0.732	0.9330	0.775	1.0327	0.818	1.1507	0.861	1.2972	0.904	1.4937	0.947	1.8019	0.990	2.6467				
0.691	0.8499	0.733	0.9352	0.776	1.0352	0.819	1.1538	0.862	1.3011	0.905	1.4992	0.948	1.8117	0.991	2.6996				
0.692	0.8518	0.734	0.9373	0.777	1.0378	0.820	1.1568	0.863	1.3050	0.906	1.5047	0.949	1.8216	0.992	2.7587				
0.693	0.8537	0.735	0.9395	0.778	1.0403	0.821	1.1599	0.864	1.3089	0.907	1.5103	0.950	1.8318	0.993	2.8257				
0.694	0.8556	0.736	0.9417	0.779	1.0428	0.822	1.1630	0.865	1.3129	0.908	1.5160	0.951	1.8421	0.994	2.9031				
0.695	0.8576	0.737	0.9439	0.780	1.0454	0.823	1.1660	0.866	1.3169	0.909	1.5217	0.952	1.8527	0.995	2.9945				
0.696	0.8595	0.738	0.9461	0.781	1.0479	0.824	1.1692	0.867	1.3209	0.910	1.5275	0.953	1.8635	0.996	3.1063				
0.697	0.8614	0.739	0.9483	0.782	1.0505	0.825	1.1723	0.868	1.3249	0.911	1.5334	0.954	1.8745	0.997	3.2504				
0.698	0.8634	0.740	0.9505	0.783	1.0531	0.826	1.1754	0.869	1.3290	0.912	1.5393	0.955	1.8857	0.998	3.4534				
0.699	0.8653	0.741	0.9527	0.784	1.0557	0.827	1.1786	0.870	1.3331	0.913	1.5453	0.956	1.8972	0.999	3.8002				
0.700	0.8673	0.742	0.9549	0.785	1.0583	0.828	1.1817	0.871	1.3372	0.914	1.5513	0.957	1.9090						
0.701	0.8693	0.743	0.9571	0.786	1.0609	0.829	1.1849	0.872	1.3414	0.915	1.5574	0.958	1.9210						
0.702	0.8712	0.744	0.9594	0.787	1.0635	0.830	1.1881	0.873	1.3456	0.916	1.5636	0.959	1.9333						
0.703	0.8732	0.745	0.9616	0.788	1.0661	0.831	1.1914	0.874	1.3498	0.917	1.5698	0.960	1.9459						
0.704	0.8752	0.746	0.9639	0.789	1.0688	0.832	1.1946	0.875	1.3540	0.918	1.5762	0.961	1.9588						
0.705	0.8772	0.747	0.9661	0.790	1.0714	0.833	1.1979	0.876	1.3583	0.919	1.5826	0.962	1.9721						
0.706	0.8792	0.748	0.9684	0.791	1.0741	0.834	1.2011	0.877	1.3626	0.920	1.5890	0.963	1.9857						
0.707	0.8812	0.749	0.9707	0.792	1.0768	0.835	1.2044	0.878	1.3670	0.921	1.5956	0.964	1.9996						
0.708	0.8832	0.750	0.9730	0.793	1.0795	0.836	1.2077	0.879	1.3714	0.922	1.6022	0.965	2.0139						
0.709	0.8852	0.751	0.9752	0.794	1.0822	0.837	1.2111	0.880	1.3758	0.923	1.6089	0.966	2.0287						
0.710	0.8872	0.752	0.9775	0.795	1.0849	0.838	1.2144	0.881	1.3802	0.924	1.6157	0.967	2.0439						
0.711	0.8892	0.753	0.9798	0.796	1.0876	0.839	1.2178	0.882	1.3847	0.925	1.6226	0.968	2.0595						
0.712	0.8912	0.754	0.9822	0.797	1.0903	0.840	1.2212	0.883	1.3892	0.926	1.6296	0.969	2.0756						
0.713	0.8933	0.755	0.9845	0.798	1.0931	0.841	1.2246	0.884	1.3938	0.927	1.6366	0.970	2.0923						
0.714	0.8953	0.756	0.9868	0.799	1.0958	0.842	1.2280	0.885	1.3984	0.928	1.6438	0.971	2.1095						
0.715	0.8973	0.757	0.9892	0.800	1.0986	0.843	1.2315	0.886	1.4030	0.929	1.6510	0.972	2.1273						
0.716	0.8994	0.758	0.9915	0.801	1.1014	0.844	1.2349	0.887	1.4077	0.930	1.6584	0.973	2.1457						
0.717	0.9014	0.759	0.9939	0.802	1.1042	0.845	1.2384	0.888	1.4124	0.931	1.6658	0.974	2.1649						
0.718	0.9035	0.760	0.9962	0.803	1.1070	0.846	1.2419	0.889	1.4171	0.932	1.6734	0.975	2.1847						
0.719	0.9056	0.761	0.9986	0.804	1.1098	0.847	1.2454	0.890	1.4219	0.933	1.6811	0.976	2.2054						
0.720	0.9076	0.762	1.0010	0.805	1.1127	0.848	1.2490	0.891	1.4268	0.934	1.6888	0.977	2.2269						
0.721	0.9097	0.763	1.0034	0.806	1.1155	0.849	1.2526	0.892	1.4316	0.935	1.6967	0.978	2.2494						
0.722	0.9118	0.764	1.0058	0.807	1.1184	0.850	1.2562	0.893	1.4365	0.936	1.7047	0.979	2.2729						
0.723	0.9139	0.765	1.0082	0.808	1.1212	0.851	1.2598	0.894	1.4415	0.937	1.7129	0.980	2.2976						
0.724	0.9160	0.766	1.0106	0.809	1.1241	0.852	1.2634	0.895	1.4465	0.938	1.7211	0.981	2.3235						
0.725	0.9181	0.767	1.0130	0.810	1.1270	0.853	1.2671	0.896	1.4516	0.939	1.7295	0.982	2.3507						
0.726	0.9202	0.768	1.0154	0.811	1.1299	0.854	1.2707	0.897	1.4566	0.940	1.7380	0.983	2.3796						
0.727	0.9223	0.769	1.0179	0.812	1.1329	0.855	1.2745	0.898	1.4618	0.941	1.7467	0.984	2.4101						
0.728	0.9245	0.770	1.0203	0.813	1.1358	0.856	1.2782	0.899	1.4670	0.942	1.7555	0.985	2.4427						
0.729	0.9266	0.771	1.0228	0.814	1.1388	0.857	1.2819	0.900	1.4722	0.943	1.7645	0.986	2.4774						
0.730	0.9287	0.772	1.0253	0.815	1.1417	0.858	1.2857	0.901	1.4775	0.944	1.7736	0.987	2.5147						
		0.773	1.0277	0.816	1.1447	0.859	1.2895	0.902	1.4828	0.945	1.7828	0.988	2.5550						

Table 4 Critical values for the Lilliefors test of normality (table obtained with 100,000 samples for each sample size). The intersection of a given row and column shows the critical value L_{critical} for the sample size labeling the row and the alpha level labeling the column. For $N > 50$ the critical value can be found by using $f_N = \frac{.83 + N}{\sqrt{N}} - .01$.

N	$\alpha = .20$	$\alpha = .15$	$\alpha = .10$	$\alpha = .05$	$\alpha = .01$
4	.3027	.3216	.3456	.3754	.4129
5	.2893	.3027	.3188	.3427	.3959
6	.2694	.2816	.2982	.3245	.3728
7	.2521	.2641	.2802	.3041	.3504
8	.2387	.2502	.2649	.2875	.3331
9	.2273	.2382	.2522	.2744	.3162
10	.2171	.2273	.2410	.2616	.3037
11	.2080	.2179	.2306	.2506	.2905
12	.2004	.2101	.2228	.2426	.2812
13	.1932	.2025	.2147	.2337	.2714
14	.1869	.1959	.2077	.2257	.2627
15	.1811	.1899	.2016	.2196	.2545
16	.1758	.1843	.1956	.2128	.2477
17	.1711	.1794	.1902	.2071	.2408
18	.1666	.1747	.1852	.2018	.2345
19	.1624	.1700	.1803	.1965	.2285
20	.1589	.1666	.1764	.1920	.2226
21	.1553	.1629	.1726	.1881	.2190
22	.1517	.1592	.1690	.1840	.2141
23	.1484	.1555	.1650	.1798	.2090
24	.1458	.1527	.1619	.1766	.2053
25	.1429	.1498	.1589	.1726	.2010
26	.1406	.1472	.1562	.1699	.1985
27	.1381	.1448	.1533	.1665	.1941
28	.1358	.1423	.1509	.1641	.1911
29	.1334	.1398	.1483	.1614	.1886
30	.1315	.1378	.1460	.1590	.1848
31	.1291	.1353	.1432	.1559	.1820
32	.1274	.1336	.1415	.1542	.1798
33	.1254	.1314	.1392	.1518	.1770
34	.1236	.1295	.1373	.1497	.1747
35	.1220	.1278	.1356	.1478	.1720
36	.1203	.1260	.1336	.1454	.1695
37	.1188	.1245	.1320	.1436	.1677
38	.1174	.1230	.1303	.1421	.1653
39	.1159	.1214	.1288	.1402	.1634
40	.1147	.1204	.1275	.1386	.1616
41	.1131	.1186	.1258	.1373	.1599
42	.1119	.1172	.1244	.1353	.1573
43	.1106	.1159	.1228	.1339	.1556
44	.1095	.1148	.1216	.1322	.1542
45	.1083	.1134	.1204	.1309	.1525
46	.1071	.1123	.1189	.1293	.1512
47	.1062	.1113	.1180	.1282	.1499
48	.1047	.1098	.1165	.1269	.1476
49	.1040	.1089	.1153	.1256	.1463
50	.1030	.1079	.1142	.1246	.1457
> 50	$\frac{0.741}{f_N}$	$\frac{0.775}{f_N}$	$\frac{0.819}{f_N}$	$\frac{0.895}{f_N}$	$\frac{1.035}{f_N}$

Table 5 Šidák's test. Critical values of Fisher's *F*. Corrected for multiple comparisons with Šidák's formula: $\alpha[PC] = 1 - (1 - \alpha(PF))^{\frac{1}{C}}$.
To be used for *a priori* contrasts ($v_1 = 1$).

		C = number of planned comparisons																							
		$\alpha[PC]$ (corrected for C comparisons)																							
		Critical values of F																							
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	25	30	40	50
.05	.0500	.0253	.0170	.0127	.0102	.0085	.0073	.0064	.0057	.0051	.0047	.0043	.0039	.0037	.0034	.0032	.0030	.0028	.0027	.0026	.0020	.0017	.0013	.0010	
.01	.0100	.0050	.0033	.0025	.0020	.0017	.0014	.0013	.0011	.0010	.0009	.0008	.0008	.0007	.0007	.0006	.0006	.0006	.0005	.0005	.0004	.0003	.0003	.0002	
v_2																									
4	7.71	12.12	15.53	18.41	20.95	23.25	25.36	27.34	29.19	30.94	32.60	34.20	35.72	37.19	38.61	39.98	41.31	42.60	43.86	45.08	50.79	55.95	65.10	73.17	
5	21.20	31.29	39.04	45.58	51.34	56.56	61.35	65.81	70.00	73.96	77.73	81.33	84.79	88.11	91.32	94.42	97.43	100.35	103.19	105.96	118.85	130.51	151.21	169.45	
6	6.61	9.94	12.38	14.37	16.09	17.62	19.00	20.26	21.44	22.54	23.57	24.56	25.49	26.38	27.23	28.06	28.85	29.61	30.35	31.06	34.35	37.27	42.32	46.67	
7	16.26	22.76	27.49	31.35	34.67	37.61	40.27	42.71	44.97	47.09	49.08	50.97	52.76	54.48	56.12	57.70	59.22	60.69	62.11	63.48	69.80	75.41	85.12	93.46	
8	5.99	8.75	10.72	12.29	13.62	14.78	15.82	16.77	17.64	18.45	19.21	19.93	20.60	21.24	21.86	22.44	23.00	23.54	24.06	24.57	26.86	28.86	32.28	35.17	
9	13.75	18.62	22.04	24.77	27.08	29.09	30.90	32.54	34.05	35.45	36.76	37.99	39.16	40.26	41.32	42.33	43.30	44.23	45.13	46.00	49.95	53.41	59.30	64.28	
10	5.59	8.02	9.71	11.03	12.14	13.10	13.96	14.73	15.44	16.09	16.70	17.27	17.81	18.32	18.80	19.26	19.70	20.12	20.53	20.92	22.70	24.23	26.83	28.99	
11	12.25	16.22	18.94	21.08	22.87	24.41	25.78	27.02	28.15	29.20	30.17	31.09	31.95	32.76	33.53	34.27	34.98	35.65	36.30	36.93	39.76	42.22	46.36	49.81	
12	5.32	7.53	9.03	10.20	11.17	12.00	12.74	13.40	14.01	14.56	15.08	15.56	16.02	16.45	16.85	17.24	17.61	17.96	18.30	18.62	20.09	21.35	23.46	25.21	
13	11.26	14.67	16.97	18.76	20.23	21.50	22.62	23.62	24.54	25.38	26.16	26.89	27.57	28.22	28.83	29.42	29.97	30.51	31.02	31.51	33.72	35.62	38.81	41.44	
14	5.12	7.17	8.54	9.60	10.48	11.23	11.89	12.48	13.01	13.50	13.96	14.38	14.78	15.16	15.51	15.85	16.17	16.48	16.77	17.05	18.32	19.40	21.21	22.69	
15	10.56	13.60	15.62	17.17	18.44	19.53	20.48	21.34	22.11	22.83	23.48	24.10	24.67	25.22	25.73	26.21	26.68	27.12	27.55	27.95	29.78	31.34	33.94	36.07	
16	4.96	6.90	8.18	9.16	9.97	10.66	11.26	11.80	12.28	12.73	13.14	13.52	13.88	14.22	14.54	14.84	15.13	15.40	15.66	15.92	17.04	18.00	19.59	20.89	
17	10.04	12.82	14.63	16.02	17.15	18.11	18.96	19.71	20.39	21.01	21.59	22.12	22.62	23.09	23.54	23.96	24.36	24.74	25.11	25.46	27.03	28.36	30.57	32.38	
18	4.84	6.69	7.90	8.82	9.58	10.22	10.78	11.27	11.72	12.14	12.51	12.87	13.20	13.51	13.80	14.07	14.34	14.59	14.83	15.06	16.08	16.95	18.39	19.55	
19	9.65	12.22	13.89	15.15	16.18	17.05	17.81	18.49	19.10	19.66	20.18	20.66	21.10	21.52	21.92	22.29	22.65	22.99	23.31	23.62	25.01	26.19	28.12	29.70	
20	4.75	6.52	7.67	8.55	9.26	9.87	10.39	10.86	11.28	11.67	12.02	12.35	12.66	12.95	13.22	13.48	13.72	13.95	14.18	14.39	15.34	16.14	17.45	18.52	
21	9.33	11.74	13.30	14.47	15.42	16.23	16.93	17.55	18.11	18.62	19.09	19.53	19.93	20.31	20.67	21.01	21.34	21.64	21.94	22.22	23.47	24.53	26.26	27.67	
22	4.67	6.38	7.49	8.33	9.01	9.59	10.09	10.53	10.93	11.29	11.63	11.94	12.23	12.50	12.75	13.00	13.23	13.44	13.65	13.85	14.74	15.49	16.71	17.70	
23	9.07	11.36	12.83	13.93	14.82	15.57	16.22	16.80	17.32	17.79	18.23	18.63	19.01	19.36	19.69	20.00	20.30	20.58	20.85	21.11	22.26	23.23	24.81	26.09	
24	4.60	6.26	7.34	8.15	8.80	9.35	9.83	10.25	10.64	10.98	11.30	11.60	11.87	12.13	12.37	12.60	12.82	13.03	13.22	13.41	14.25	14.96	16.11	17.04	
25	8.86	11.05	12.45	13.49	14.33	15.04	15.65	16.19	16.68	17.12	17.53	17.91	18.26	18.59	18.90	19.19	19.47	19.73	19.98	20.22	21.29	22.18	23.65	24.82	
26	4.54	6.17	7.21	7.99	8.62	9.16	9.62	10.02	10.39	10.72	11.03	11.31	11.58	11.82	12.06	12.27	12.48	12.68	12.87	13.05	13.84	14.52	15.61	16.49	
27	8.68	10.79	12.12	13.12	13.92	14.59	15.17	15.69	16.15	16.57	16.95	17.31	17.64	17.95	18.24	18.52	18.78	19.02	19.26	19.49	20.48	21.32	22.69	23.79	
28	4.49	6.08	7.10	7.86	8.47	8.99	9.43	9.83	10.18	10.50	10.80	11.07	11.33	11.56	11.79	12.00	12.20	12.38	12.57	12.74	13.50	14.14	15.19	16.02	
29	8.53	10.57	11.85	12.81	13.57	14.22	14.77	15.26	15.70	16.10	16.47	16.81	17.12	17.42	17.69	17.95	18.20	18.43	18.66	18.87	19.82	20.61	21.90	22.93	
30	4.45	6.01	7.01	7.75	8.34	8.85	9.28	9.66	10.00	10.32	10.60	10.86	11.11	11.34	11.55	11.76	11.95	12.13	12.31	12.47	13.21	13.83	14.83	15.63	
31	8.40	10.38	11.62	12.54	13.28	13.90	14.43	14.90	15.32	15.71	16.06	16.38	16.68	16.96	17.22	17.47	17.71	17.93	18.15	18.35	19.25	20.00	21.23	22.20	

18	4.41	5.95	6.92	7.65	8.23	8.72	9.14	9.51	9.85	10.15	10.43	10.68	10.92	11.15	11.35	11.55	11.74	11.91	12.08	12.24	12.36	12.46	12.54	12.60	12.65	12.69	12.72	12.75	12.78	12.80	12.82	12.84	12.86	12.88	12.90	12.92	12.94	12.96	12.98	13.00	13.02	13.04	13.06	13.08	13.10	13.12	13.14	13.16	13.18	13.20	13.22	13.24	13.26	13.28	13.30	13.32	13.34	13.36	13.38	13.40	13.42	13.44	13.46	13.48	13.50	13.52	13.54	13.56	13.58	13.60	13.62	13.64	13.66	13.68	13.70	13.72	13.74	13.76	13.78	13.80	13.82	13.84	13.86	13.88	13.90	13.92	13.94	13.96	13.98	14.00	14.02	14.04	14.06	14.08	14.10	14.12	14.14	14.16	14.18	14.20	14.22	14.24	14.26	14.28	14.30	14.32	14.34	14.36	14.38	14.40	14.42	14.44	14.46	14.48	14.50	14.52	14.54	14.56	14.58	14.60	14.62	14.64	14.66	14.68	14.70	14.72	14.74	14.76	14.78	14.80	14.82	14.84	14.86	14.88	14.90	14.92	14.94	14.96	14.98	15.00	15.02	15.04	15.06	15.08	15.10	15.12	15.14	15.16	15.18	15.20	15.22	15.24	15.26	15.28	15.30	15.32	15.34	15.36	15.38	15.40	15.42	15.44	15.46	15.48	15.50	15.52	15.54	15.56	15.58	15.60	15.62	15.64	15.66	15.68	15.70	15.72	15.74	15.76	15.78	15.80	15.82	15.84	15.86	15.88	15.90	15.92	15.94	15.96	15.98	16.00	16.02	16.04	16.06	16.08	16.10	16.12	16.14	16.16	16.18	16.20	16.22	16.24	16.26	16.28	16.30	16.32	16.34	16.36	16.38	16.40	16.42	16.44	16.46	16.48	16.50	16.52	16.54	16.56	16.58	16.60	16.62	16.64	16.66	16.68	16.70	16.72	16.74	16.76	16.78	16.80	16.82	16.84	16.86	16.88	16.90	16.92	16.94	16.96	16.98	17.00	17.02	17.04	17.06	17.08	17.10	17.12	17.14	17.16	17.18	17.20	17.22	17.24	17.26	17.28	17.30	17.32	17.34	17.36	17.38	17.40	17.42	17.44	17.46	17.48	17.50	17.52	17.54	17.56	17.58	17.60	17.62	17.64	17.66	17.68	17.70	17.72	17.74	17.76	17.78	17.80	17.82	17.84	17.86	17.88	17.90	17.92	17.94	17.96	17.98	18.00	18.02	18.04	18.06	18.08	18.10	18.12	18.14	18.16	18.18	18.20	18.22	18.24	18.26	18.28	18.30	18.32	18.34	18.36	18.38	18.40	18.42	18.44	18.46	18.48	18.50	18.52	18.54	18.56	18.58	18.60	18.62	18.64	18.66	18.68	18.70	18.72	18.74	18.76	18.78	18.80	18.82	18.84	18.86	18.88	18.90	18.92	18.94	18.96	18.98	19.00	19.02	19.04	19.06	19.08	19.10	19.12	19.14	19.16	19.18	19.20	19.22	19.24	19.26	19.28	19.30	19.32	19.34	19.36	19.38	19.40	19.42	19.44	19.46	19.48	19.50	19.52	19.54	19.56	19.58	19.60	19.62	19.64	19.66	19.68	19.70	19.72	19.74	19.76	19.78	19.80	19.82	19.84	19.86	19.88	19.90	19.92	19.94	19.96	19.98	20.00
19	4.38	5.89	6.85	7.56	8.13	8.61	9.02	9.39	9.71	10.01	10.28	10.53	10.76	10.98	11.18	11.37	11.55	11.72	11.89	12.04	12.17	12.29	12.40	12.50	12.59	12.67	12.75	12.82	12.89	12.95	13.01	13.07	13.13	13.18	13.23	13.28	13.32	13.36	13.40	13.44	13.48	13.52	13.56	13.60	13.64	13.68	13.72	13.76	13.80	13.84	13.88	13.92	13.96	14.00	14.04	14.08	14.12	14.16	14.20	14.24	14.28	14.32	14.36	14.40	14.44	14.48	14.52	14.56	14.60	14.64	14.68	14.72	14.76	14.80	14.84	14.88	14.92	14.96	15.00	15.04	15.08	15.12	15.16	15.20	15.24	15.28	15.32	15.36	15.40	15.44	15.48	15.52	15.56	15.60	15.64	15.68	15.72	15.76	15.80	15.84	15.88	15.92	15.96	16.00	16.04	16.08	16.12	16.16	16.20	16.24	16.28	16.32	16.36	16.40	16.44	16.48	16.52	16.56	16.60	16.64	16.68	16.72	16.76	16.80	16.84	16.88	16.92	16.96	17.00	17.04	17.08	17.12	17.16	17.20	17.24	17.28	17.32	17.36	17.40	17.44	17.48	17.52	17.56	17.60	17.64	17.68	17.72	17.76	17.80	17.84	17.88	17.92	17.96	18.00	18.04	18.08	18.12	18.16	18.20	18.24	18.28	18.32	18.36	18.40	18.44	18.48	18.52	18.56	18.60	18.64	18.68	18.72	18.76	18.80	18.84	18.88	18.92	18.96	19.00	19.04	19.08	19.12	19.16	19.20	19.24	19.28	19.32	19.36	19.40	19.44	19.48	19.52	19.56	19.60	19.64	19.68	19.72	19.76	19.80	19.84	19.88	19.92	19.96	20.00																																																																																																																																																																																										
20	4.35	5.84	6.78	7.48	8.04	8.51	8.92	9.27	9.59	9.88	10.14	10.39	10.61	10.83	11.02	11.21	11.39	11.56	11.72	11.87	12.00	12.12	12.24	12.35	12.45	12.55	12.64	12.72	12.80	12.87	12.94	13.00	13.06	13.12	13.17	13.22	13.27	13.32	13.37	13.42	13.47	13.52	13.57	13.62	13.67	13.72	13.77	13.82	13.87	13.92	13.97	14.02	14.07	14.12	14.17	14.22	14.27	14.32	14.37	14.42	14.47	14.52	14.57	14.62	14.67	14.72	14.77	14.82	14.87	14.92	14.97	15.02	15.07	15.12	15.17	15.22	15.27	15.32	15.37	15.42	15.47	15.52	15.57	15.62	15.67	15.72	15.77	15.82	15.87	15.92	15.97	16.02	16.07	16.12	16.17	16.22	16.27	16.32	16.37	16.42	16.47	16.52	16.57	16.62	16.67	16.72	16.77	16.82	16.87	16.92	16.97	17.02	17.07	17.12	17.17	17.22	17.27	17.32	17.37	17.42	17.47	17.52	17.57	17.62	17.67	17.72	17.77	17.82	17.87	17.92	17.97	18.02	18.07	18.12	18.17	18.22	18.27	18.32	18.37	18.42	18.47	18.52	18.57	18.62	18.67	18.72	18.77	18.82	18.87	18.92	18.97	19.02	19.07	19.12	19.17	19.22	19.27	19.32	19.37	19.42	19.47	19.52	19.57	19.62	19.67	19.72	19.77	19.82	19.87	19.92	19.97	20.00																																																																																																																																																																																																																										
21	4.32	5.80	6.73	7.41	7.97	8.43	8.82	9.17	9.48	9.77	10.03	10.26	10.49	10.69	10.89	11.07	11.24	11.41	11.56	11.71	11.85	11.99	12.12	12.24	12.35	12.45	12.54	12.62	12.70	12.77	12.84	12.90	12.96	13.02	13.07	13.12	13.17	13.22	13.27	13.32	13.37	13.42	13.47	13.52	13.57	13.62	13.67	13.72	13.77	13.82	13.87	13.92	13.97	14.02	14.07	14.12	14.17	14.22	14.27	14.32	14.37	14.42	14.47	14.52	14.57	14.62	14.67	14.72	14.77	14.82	14.87	14.92	14.97	15.02	15.07	15.12	15.17	15.22	15.27	15.32	15.37	15.42	15.47	15.52	15.57	15.62	15.67	15.72	15.77	15.82	15.87	15.92	15.97	16.02	16.07	16.12	16.17	16.22	16.27	16.32	16.37	16.42	16.47	16.52	16.57	16.62	16.67	16.72	16.77	16.82	16.87	16.92	16.97	17.02	17.07	17.12	17.17	17.22	17.27	17.32	17.37	17.42	17.47	17.52	17.57	17.62	17.67	17.72	17.77	17.82	17.87	17.92	17.97	18.02	18.07	18.12	18.17	18.22	18.27	18.32	18.37	18.42	18.47	18.52	18.57	18.62	18.67	18.72	18.77	18.82	18.87	18.92	18.97	19.02	19.07	19.12	19.17	19.22	19.27	19.32	19.37	19.42	19.47	19.52	19.57	19.62	19.67	19.72	19.77	19.82	19.87	19.92	19.97	20.00																																																																																																																																																																																																																								
22	4.30	5.76	6.67	7.35	7.89	8.35	8.74	9.08	9.39	9.67	9.92	10.15	10.37	10.57	10.76	10.94	11.11	11.27	11.43	11.57	11.71	11.85	11.99	12.12	12.24	12.35	12.45	12.54	12.62	12.70	12.77	12.84	12.90	12.96	13.02	13.07	13.12	13.17	13.22	13.27	13.32	13.37	13.42	13.47	13.52	13.57	13.62	13.67	13.72	13.77	13.82	13.87	13.92	13.97	14.02	14.07	14.12	14.17	14.22	14.27	14.32	14.37	14.42	14.47	14.52	14.57	14.62	14.67	14.72	14.77	14.82	14.87	14.92	14.97	15.02	15.07	15.12	15.17	15.22	15.27	15.32	15.37	15.42	15.47	15.52	15.57	15.62	15.67	15.72	15.77	15.82	15.87	15.92	15.97	16.02	16.07	16.12	16.17	16.22	16.27	16.32	16.37	16.42	16.47	16.52	16.57	16.62	16.67	16.72	16.77	16.82	16.87	16.92	16.97	17.02	17.07	17.12	17.17	17.22	17.27	17.32	17.37	17.42	17.47	17.52	17.57	17.62	17.67	17.72	17.77	17.82	17.87	17.92	17.97	18.02	18.07	18.12	18.17	18.22	18.27	18.32	18.37	18.42	18.47	18.52	18.57	18.62	18.67	18.72	18.77	18.82	18.87	18.92	18.97	19.02	19.07	19.12	19.17	19.22	19.27	19.32	19.37	19.42	19.47	19.52	19.57	19.62	19.67	19.72	19.77	19.82	19.87	19.92	19.97	20.00																																																																																																																																																																																																																							
23	4.28	5.72	6.63	7.30	7.83	8.28	8.66	9.00	9.30	9.57	9.82	10.05	10.27	10.47	10.65	10.83	11.00	11.15	11.30	11.45	11.59	11.73	11.87	12.00	12.12	12.24	12.35	12.45	12.54	12.62	12.70	12.77	12.84	12.90	12.96	13.02	13.07	13.12	13.17	13.22	13.27	13.32	13.37	13.42	13.47	13.52	13.57	13.62	13.67	13.72	13.77	13.82	13.87	13.92	13.97	14.02	14.07	14.12	14.17	14.22	14.27	14.32	14.37	14.42	14.47	14.52	14.57	14.62	14.67	14.72	14.77	14.82	14.87	14.92	14.97	15.02	15.07	15.12	15.17	15.22	15.27	15.32	15.37	15.42	15.47	15.52	15.57	15.62	15.67	15.72	15.77	15.82	15.87	15.92	15.97	16.02	16.07	16.12	16.17	16.22	16.27	16.32	16.37	16.42	16.47	16.52	16.57	16.62	16.67	16.72	16.77	16.82	16.87																																																																																																																																																																																																																																																																																					

Table 5 Šidák's test. Critical values of Fisher's F . Corrected for multiple comparisons with Šidák's formula: $\alpha[PC] = 1 - (1 - \alpha[PF])^{\frac{1}{C}}$. To be used for *a priori* contrasts ($\nu_1 = 1$) (continued).

			C = number of planned comparisons																					
			$\alpha[PC]$ (corrected for C comparisons)																					
			Critical values of F																					
ν_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	25	30	40	50
.05	.0500	.0253	.0170	.0127	.0102	.0085	.0073	.0064	.0057	.0051	.0047	.0043	.0039	.0037	.0034	.0032	.0030	.0028	.0027	.0026	.0020	.0017	.0013	.0010
.01	.0100	.0050	.0033	.0025	.0020	.0017	.0014	.0013	.0011	.0010	.0009	.0008	.0008	.0007	.0007	.0006	.0006	.0006	.0005	.0005	.0004	.0003	.0003	.0002
40	4.08	5.40	6.21	6.80	7.27	7.66	7.99	8.28	8.54	8.78	8.99	9.18	9.37	9.53	9.69	9.84	9.98	10.11	10.24	10.35	10.88	11.31	12.00	12.55
42	4.31	5.82	6.74	7.40	7.93	8.36	8.73	9.05	9.31	9.53	9.73	9.92	10.11	10.29	10.47	10.65	10.83	11.01	11.19	11.37	12.00	12.43	13.12	13.67
44	4.54	6.18	7.16	7.88	8.46	8.94	9.36	9.73	10.05	10.31	10.53	10.73	10.92	11.11	11.29	11.47	11.65	11.83	12.01	12.19	12.92	13.35	14.04	14.59
46	4.77	6.56	7.60	8.38	8.99	9.52	10.00	10.42	10.79	11.07	11.34	11.61	11.88	12.15	12.42	12.69	12.96	13.23	13.50	13.77	14.60	15.03	15.72	16.27
48	5.00	6.93	8.04	8.88	9.54	10.14	10.71	11.23	11.75	12.22	12.69	13.16	13.63	14.10	14.57	15.04	15.51	15.98	16.45	16.92	17.85	18.28	19.07	19.62
50	5.23	7.30	8.48	9.38	10.11	10.76	11.39	12.00	12.59	13.16	13.73	14.30	14.87	15.44	16.01	16.58	17.15	17.72	18.29	18.86	19.89	20.32	21.11	21.66
60	6.00	8.26	9.61	10.64	11.45	12.15	12.82	13.48	14.13	14.78	15.43	16.08	16.73	17.38	18.03	18.68	19.33	19.98	20.63	21.28	22.41	22.84	23.63	24.18
65	6.54	8.95	10.46	11.63	12.51	13.18	13.83	14.48	15.13	15.78	16.43	17.08	17.73	18.38	19.03	19.68	20.33	20.98	21.63	22.28	23.51	23.94	24.73	25.28
70	7.01	9.58	11.25	12.56	13.51	14.16	14.81	15.46	16.11	16.76	17.41	18.06	18.71	19.36	20.01	20.66	21.31	21.96	22.61	23.26	24.59	25.02	25.81	26.36
80	7.96	10.78	12.63	14.10	15.21	15.85	16.50	17.15	17.80	18.45	19.10	19.75	20.40	21.05	21.70	22.35	23.00	23.65	24.30	24.95	26.39	26.82	27.61	28.16
100	9.34	12.44	14.54	16.25	17.51	18.15	18.80	19.45	20.10	20.75	21.40	22.05	22.70	23.35	24.00	24.65	25.30	25.95	26.60	27.25	28.79	29.22	30.01	30.56
125	10.64	14.04	16.36	18.30	19.71	20.35	21.00	21.65	22.30	22.95	23.60	24.25	24.90	25.55	26.20	26.85	27.50	28.15	28.80	29.45	31.09	31.52	32.31	32.86
150	11.81	15.48	17.94	20.13	21.69	22.33	23.00	23.67	24.34	25.01	25.68	26.35	27.02	27.69	28.36	29.03	29.70	30.37	31.04	31.71	33.35	33.78	34.57	35.12
200	13.76	17.80	20.42	22.84	24.56	25.20	25.97	26.64	27.31	27.98	28.65	29.32	30.00	30.67	31.34	32.01	32.68	33.35	34.02	34.69	36.33	36.76	37.55	38.10
400	17.54	22.94	25.76	28.44	30.36	31.00	31.87	32.54	33.21	33.88	34.55	35.22	35.90	36.57	37.24	37.91	38.58	39.25	39.92	40.59	42.23	42.66	43.45	43.90
1000	23.68	30.64	33.81	36.78	38.89	39.53	40.40	41.07	41.74	42.41	43.08	43.75	44.42	45.09	45.76	46.43	47.10	47.77	48.44	49.11	50.75	51.18	51.97	52.52
5000	30.82	39.44	43.01	46.29	48.61	49.25	50.12	50.79	51.46	52.13	52.80	53.47	54.14	54.81	55.48	56.15	56.82	57.49	58.16	58.83	60.47	60.90	61.69	62.24

Table 6 Bonferroni's test. Critical values of Fisher's F . Corrected for multiple comparisons with Bonferroni's formula:
 $\alpha[PC] = \frac{\alpha[PF]}{C}$. To be used for *a priori* contrasts ($\nu_1 = 1$). $\alpha = .05$ $\alpha = .01$

		C = number of planned comparisons																							
		$\alpha[PC]$ (corrected for C comparisons)																							
		Critical values of F																							
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	25	30	40	50
.05		.0500	.0250	.0167	.0125	.0100	.0083	.0071	.0063	.0056	.0050	.0045	.0042	.0038	.0036	.0033	.0031	.0029	.0028	.0026	.0025	.0020	.0017	.0013	.0010
.01		.0100	.0050	.0033	.0025	.0020	.0017	.0014	.0013	.0011	.0010	.0009	.0008	.0008	.0007	.0007	.0006	.0006	.0006	.0005	.0005	.0004	.0003	.0003	.0002
ν_2																									
4	7.71	12.22	15.69	18.62	21.20	23.53	25.68	27.68	29.56	31.33	33.02	34.64	36.19	37.68	39.11	40.50	41.85	43.16	44.43	45.67	51.45	56.68	65.96	74.14	
	21.20	31.33	39.11	45.67	51.45	56.68	61.49	65.96	70.16	74.14	77.92	81.53	84.99	88.33	91.54	94.65	97.67	100.60	103.45	106.22	119.15	130.84	151.59	169.88	
5	6.61	10.01	12.49	14.52	16.26	17.80	19.20	20.48	21.67	22.78	23.83	24.83	25.77	26.67	27.54	28.37	29.16	29.94	30.68	31.41	34.73	37.68	42.79	47.18	
	16.26	22.78	27.54	31.41	34.73	37.68	40.34	42.79	45.06	47.18	49.18	51.07	52.87	54.59	56.23	57.82	59.34	60.81	62.23	63.61	69.95	75.56	85.29	93.66	
6	5.99	8.81	10.81	12.40	13.75	14.92	15.98	16.93	17.82	18.63	19.40	20.12	20.80	21.45	22.07	22.66	23.23	23.77	24.30	24.81	27.12	29.14	32.59	35.51	
	13.75	18.63	22.07	24.81	27.12	29.14	30.95	32.59	34.10	35.51	36.82	38.06	39.23	40.34	41.39	42.41	43.38	44.31	45.21	46.08	50.04	53.50	59.41	64.40	
7	5.59	8.07	9.78	11.12	12.25	13.22	14.08	14.86	15.58	16.24	16.85	17.42	17.97	18.48	18.97	19.43	19.88	20.31	20.72	21.11	22.90	24.45	27.06	29.25	
	12.25	16.24	18.97	21.11	22.90	24.45	25.82	27.06	28.20	29.25	30.22	31.14	32.00	32.81	33.59	34.33	35.03	35.71	36.36	36.99	39.83	42.28	46.43	49.89	
8	5.32	7.57	9.09	10.28	11.26	12.10	12.85	13.52	14.13	14.69	15.21	15.70	16.15	16.58	16.99	17.38	17.75	18.11	18.45	18.78	20.26	21.53	23.66	25.41	
	11.26	14.69	16.99	18.78	20.26	21.53	22.65	23.66	24.57	25.41	26.20	26.93	27.61	28.26	28.88	29.46	30.02	30.55	31.06	31.56	33.77	35.68	38.86	41.50	
9	5.12	7.21	8.60	9.68	10.56	11.32	11.98	12.58	13.12	13.61	14.07	14.50	14.90	15.28	15.64	15.98	16.30	16.61	16.90	17.19	18.46	19.55	21.37	22.86	
	10.56	13.61	15.64	17.19	18.46	19.55	20.51	21.37	22.14	22.86	23.52	24.13	24.71	25.25	25.76	26.25	26.71	27.16	27.58	27.99	29.82	31.38	33.99	36.12	
10	4.96	6.94	8.24	9.23	10.04	10.74	11.34	11.89	12.38	12.83	13.24	13.63	13.99	14.33	14.65	14.95	15.24	15.52	15.78	16.04	17.17	18.14	19.73	21.04	
	10.04	12.83	14.65	16.04	17.17	18.14	18.98	19.73	20.42	21.04	21.62	22.15	22.65	23.12	23.57	23.99	24.39	24.77	25.14	25.49	27.06	28.40	30.61	32.42	
11	4.84	6.72	7.95	8.89	9.65	10.29	10.86	11.36	11.81	12.23	12.61	12.96	13.29	13.61	13.90	14.18	14.44	14.70	14.94	15.17	16.20	17.07	18.51	19.69	
	9.65	12.23	13.90	15.17	16.20	17.07	17.84	18.51	19.13	19.69	20.20	20.68	21.13	21.55	21.94	22.32	22.68	23.02	23.34	23.65	25.04	26.22	28.16	29.73	
12	4.75	6.55	7.73	8.61	9.33	9.94	10.47	10.94	11.37	11.75	12.11	12.44	12.75	13.04	13.31	13.57	13.82	14.05	14.28	14.49	15.44	16.25	17.57	18.64	
	9.33	11.75	13.31	14.49	15.44	16.25	16.95	17.57	18.13	18.64	19.11	19.55	19.96	20.34	20.70	21.04	21.36	21.67	21.96	22.24	23.50	24.56	26.29	27.70	
13	4.67	6.41	7.54	8.39	9.07	9.65	10.16	10.60	11.01	11.37	11.71	12.02	12.31	12.59	12.84	13.09	13.32	13.54	13.75	13.95	14.84	15.59	16.82	17.82	
	9.07	11.37	12.84	13.95	14.84	15.59	16.24	16.82	17.34	17.82	18.25	18.65	19.03	19.38	19.71	20.03	20.33	20.61	20.88	21.14	22.28	23.25	24.84	26.12	
14	4.60	6.30	7.39	8.20	8.86	9.42	9.90	10.33	10.71	11.06	11.38	11.68	11.95	12.21	12.46	12.69	12.91	13.11	13.31	13.50	14.34	15.05	16.21	17.14	
	8.86	11.06	12.46	13.50	14.34	15.05	15.67	16.21	16.70	17.14	17.55	17.93	18.28	18.61	18.92	19.21	19.49	19.75	20.00	20.24	21.31	22.21	23.67	24.85	
15	4.54	6.20	7.26	8.05	8.68	9.22	9.68	10.09	10.46	10.80	11.11	11.39	11.66	11.90	12.14	12.36	12.56	12.76	12.95	13.13	13.93	14.61	15.71	16.59	
	8.68	10.80	12.14	13.13	13.93	14.61	15.19	15.71	16.17	16.59	16.97	17.33	17.66	17.97	18.26	18.54	18.80	19.04	19.28	19.51	20.51	21.35	22.72	23.81	
16	4.49	6.12	7.15	7.91	8.53	9.05	9.50	9.90	10.25	10.58	10.87	11.15	11.40	11.64	11.86	12.07	12.28	12.47	12.65	12.82	13.59	14.23	15.28	16.12	
	8.53	10.58	11.86	12.82	13.59	14.23	14.79	15.28	15.72	16.12	16.49	16.83	17.14	17.43	17.71	17.97	18.22	18.45	18.68	18.89	19.84	20.63	21.92	22.95	
17	4.45	6.04	7.05	7.80	8.40	8.90	9.34	9.73	10.07	10.38	10.67	10.94	11.18	11.41	11.63	11.83	12.03	12.21	12.39	12.55	13.29	13.91	14.92	15.72	
	8.40	10.38	11.63	12.55	13.29	13.91	14.45	14.92	15.34	15.72	16.07	16.40	16.70	16.98	17.24	17.49	17.73	17.95	18.16	18.37	19.27	20.02	21.25	22.23	
18	4.41	5.98	6.97	7.70	8.29	8.78	9.20	9.58	9.91	10.22	10.50	10.75	10.99	11.22	11.43	11.62	11.81	11.99	12.16	12.32	13.04	13.63	14.60	15.38	
	8.29	10.22	11.43	12.32	13.04	13.63	14.15	14.60	15.01	15.38	15.72	16.03	16.32	16.59	16.84	17.08	17.31	17.52	17.72	17.92	18.78	19.50	20.67	21.60	
19	4.38	5.92	6.89	7.61	8.18	8.67	9.08	9.45	9.78	10.07	10.34	10.59	10.83	11.05	11.25	11.44	11.62	11.80	11.96	12.12	12.81	13.39	14.33	15.08	
	8.18	10.07	11.25	12.12	12.81	13.39	13.89	14.33	14.72	15.08	15.41	15.71	15.99	16.25	16.49	16.72	16.94	17.15	17.34	17.53	18.36	19.05	20.17	21.07	
20	4.35	5.87	6.83	7.53	8.10	8.57	8.97	9.33	9.65	9.94	10.21	10.45	10.68	10.89	11.09	11.28	11.46	11.63	11.79	11.94	12.62	13.18	14.09	14.82	
	8.10	9.94	11.09	11.94	12.62	13.18	13.66	14.09	14.47	14.82	15.13	15.43	15.70	15.95	16.18	16.41	16.62	16.82	17.01	17.19	17.99	18.66	19.74	20.60	

Table 6 Bonferroni's test. Critical values of Fisher's F . Corrected for multiple comparisons with Bonferroni's formula: $\alpha[PC] = \frac{\alpha[PE]}{C}$. To be used for $a priori$ contrasts ($\nu_1 = 1$). $\alpha = .05$ $\alpha = .01$

		C = number of planned comparisons																											
		$\alpha[PC]$ (corrected for C comparisons)																											
		Critical values of F																											
ν_2		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	25	30	40	50				
.05	.0500	.0250	.0167	.0125	.0100	.0083	.0071	.0063	.0056	.0050	.0045	.0042	.0038	.0036	.0033	.0031	.0029	.0028	.0026	.0025	.0025	.0020	.0017	.0013	.0010				
.01	.0100	.0050	.0033	.0025	.0020	.0017	.0014	.0013	.0011	.0010	.0009	.0008	.0008	.0007	.0007	.0006	.0006	.0006	.0005	.0005	.0005	.0004	.0003	.0003	.0002				
21	4.32	5.83	6.77	7.46	8.02	8.48	8.88	9.23	9.55	9.83	10.09	10.33	10.55	10.76	10.95	11.14	11.31	11.48	11.63	11.78	12.44	12.99	13.88	14.59					
22	8.02	9.83	10.95	11.78	12.44	12.99	13.46	13.88	14.25	14.59	14.89	15.18	15.44	15.68	15.91	16.13	16.33	16.53	16.71	16.89	17.66	18.31	19.36	20.19					
23	4.30	5.79	6.71	7.40	7.95	8.40	8.79	9.14	9.45	9.73	9.98	10.22	10.44	10.64	10.83	11.01	11.18	11.34	11.49	11.64	12.28	12.82	13.69	14.38					
24	7.95	9.73	10.83	11.64	12.28	12.82	13.28	13.69	14.05	14.38	14.68	14.96	15.21	15.45	15.67	15.88	16.08	16.27	16.45	16.62	17.37	18.00	19.02	19.82					
25	4.28	5.75	6.67	7.34	7.88	8.33	8.72	9.06	9.36	9.63	9.89	10.12	10.33	10.53	10.72	10.89	11.06	11.22	11.37	11.51	12.14	12.67	13.52	14.20					
26	7.88	9.63	10.72	11.51	12.14	12.67	13.12	13.52	13.87	14.20	14.49	14.76	15.01	15.24	15.46	15.66	15.86	16.04	16.22	16.38	17.12	17.73	18.71	19.49					
27	4.26	5.72	6.62	7.29	7.82	8.27	8.65	8.98	9.28	9.55	9.80	10.03	10.24	10.43	10.62	10.79	10.95	11.11	11.26	11.40	12.02	12.53	13.37	14.03					
28	7.82	9.55	10.62	11.40	12.02	12.53	12.98	13.37	13.71	14.03	14.31	14.58	14.82	15.05	15.26	15.46	15.65	15.83	16.00	16.17	16.88	17.48	18.44	19.20					
29	4.24	5.69	6.58	7.24	7.77	8.21	8.58	8.91	9.21	9.48	9.72	9.94	10.15	10.34	10.53	10.70	10.86	11.01	11.16	11.29	11.90	12.41	13.23	13.88					
30	7.77	9.48	10.53	11.29	11.90	12.41	12.85	13.23	13.57	13.88	14.16	14.42	14.65	14.88	15.09	15.28	15.47	15.65	15.81	15.97	16.67	17.26	18.19	18.94					
31	4.23	5.66	6.55	7.20	7.72	8.15	8.53	8.85	9.14	9.41	9.65	9.87	10.07	10.26	10.44	10.61	10.77	10.92	11.06	11.20	11.80	12.30	13.10	13.74					
32	7.72	9.41	10.44	11.20	11.80	12.30	12.73	13.10	13.44	13.74	14.01	14.27	14.50	14.72	14.93	15.12	15.30	15.47	15.64	15.79	16.48	17.05	17.97	18.70					
33	4.21	5.63	6.52	7.16	7.68	8.10	8.47	8.79	9.08	9.34	9.58	9.80	10.00	10.19	10.36	10.53	10.69	10.84	10.98	11.11	11.70	12.20	12.99	13.61					
34	7.68	9.34	10.36	11.11	11.70	12.20	12.62	12.99	13.32	13.61	13.88	14.13	14.36	14.58	14.78	14.97	15.15	15.32	15.48	15.63	16.31	16.87	17.77	18.48					
35	4.20	5.61	6.48	7.13	7.64	8.06	8.42	8.74	9.03	9.28	9.52	9.73	9.93	10.12	10.29	10.46	10.61	10.76	10.90	11.03	11.62	12.10	12.88	13.50					
36	7.64	9.28	10.29	11.03	11.62	12.10	12.52	12.88	13.20	13.50	13.76	14.01	14.24	14.45	14.65	14.83	15.01	15.18	15.33	15.48	16.15	16.70	17.58	18.28					
37	4.18	5.59	6.46	7.09	7.60	8.02	8.38	8.69	8.98	9.23	9.46	9.67	9.87	10.06	10.23	10.39	10.54	10.69	10.83	10.96	11.53	12.01	12.78	13.39					
38	7.60	9.23	10.23	10.96	11.53	12.01	12.42	12.78	13.10	13.39	13.65	13.90	14.12	14.33	14.52	14.71	14.88	15.04	15.20	15.35	16.00	16.54	17.41	18.10					
39	4.17	5.57	6.43	7.06	7.56	7.98	8.34	8.65	8.93	9.18	9.41	9.62	9.82	10.00	10.17	10.33	10.48	10.62	10.76	10.89	11.46	11.93	12.69	13.29					
40	7.56	9.18	10.17	10.89	11.46	11.93	12.34	12.69	13.01	13.29	13.55	13.79	14.01	14.22	14.41	14.59	14.78	14.92	15.08	15.22	15.87	16.40	17.25	17.93					
41	4.16	5.55	6.41	7.03	7.53	7.94	8.30	8.61	8.88	9.13	9.36	9.57	9.76	9.94	10.11	10.27	10.42	10.56	10.70	10.83	11.39	11.86	12.61	13.20					
42	7.53	9.13	10.11	10.83	11.39	11.86	12.26	12.61	12.92	13.20	13.46	13.69	13.91	14.11	14.30	14.48	14.65	14.81	14.96	15.11	15.74	16.26	17.11	17.77					
43	4.15	5.53	6.38	7.01	7.50	7.91	8.26	8.57	8.84	9.09	9.32	9.52	9.71	9.89	10.06	10.22	10.37	10.51	10.64	10.77	11.33	11.79	12.53	13.12					
44	7.50	9.09	10.06	10.77	11.33	11.79	12.18	12.53	12.84	13.12	13.37	13.60	13.82	14.02	14.21	14.38	14.55	14.71	14.86	15.00	15.62	16.14	16.97	17.63					
45	4.14	5.51	6.36	6.98	7.47	7.88	8.23	8.53	8.80	9.05	9.27	9.48	9.67	9.85	10.01	10.17	10.31	10.45	10.59	10.71	11.26	11.72	12.46	13.04					
46	7.47	9.05	10.01	10.71	11.26	11.72	12.12	12.46	12.76	13.04	13.29	13.52	13.73	13.93	14.12	14.29	14.45	14.61	14.76	14.90	15.52	16.03	16.85	17.49					
47	4.13	5.50	6.34	6.96	7.44	7.85	8.19	8.50	8.77	9.01	9.23	9.44	9.63	9.80	9.97	10.12	10.27	10.40	10.54	10.66	11.21	11.66	12.39	12.97					
48	7.44	9.01	9.97	10.66	11.21	11.66	12.05	12.39	12.69	12.97	13.21	13.44	13.65	13.85	14.03	14.20	14.37	14.52	14.67	14.80	15.41	15.92	16.73	17.37					

35	4.12	5.48	6.32	6.93	7.42	7.82	8.17	8.47	8.73	8.98	9.20	9.40	9.59	9.76	9.92	10.08	10.22	10.36	10.49	10.61	11.16	11.61	12.33	12.90
	7.42	8.98	9.92	10.61	11.16	11.61	11.99	12.33	12.63	12.90	13.14	13.37	13.58	13.77	13.95	14.12	14.28	14.43	14.58	14.72	15.32	15.82	16.62	17.25
36	4.11	5.47	6.31	6.91	7.40	7.80	8.14	8.44	8.70	8.94	9.16	9.36	9.55	9.72	9.88	10.04	10.18	10.32	10.44	10.57	11.11	11.55	12.27	12.83
	7.40	8.94	9.88	10.57	11.11	11.55	11.93	12.27	12.56	12.83	13.08	13.30	13.51	13.70	13.88	14.05	14.20	14.36	14.50	14.63	15.23	15.73	16.52	17.14
37	4.11	5.46	6.29	6.89	7.37	7.77	8.11	8.41	8.67	8.91	9.13	9.33	9.51	9.69	9.85	10.00	10.14	10.27	10.40	10.52	11.06	11.50	12.21	12.77
	7.37	8.91	9.85	10.52	11.06	11.50	11.88	12.21	12.51	12.77	13.01	13.23	13.44	13.63	13.81	13.97	14.13	14.28	14.42	14.56	15.15	15.64	16.42	17.04
38	4.10	5.45	6.27	6.88	7.35	7.75	8.09	8.38	8.64	8.88	9.10	9.30	9.48	9.65	9.81	9.96	10.10	10.24	10.36	10.48	11.02	11.46	12.16	12.71
	7.35	8.88	9.81	10.48	11.02	11.46	11.83	12.16	12.45	12.71	12.95	13.17	13.38	13.56	13.74	13.91	14.06	14.21	14.35	14.48	15.07	15.55	16.33	16.94
39	4.09	5.43	6.26	6.86	7.33	7.73	8.06	8.36	8.62	8.85	9.07	9.27	9.45	9.62	9.78	9.93	10.07	10.20	10.33	10.45	10.97	11.41	12.11	12.66
	7.33	8.85	9.78	10.45	10.97	11.41	11.78	12.11	12.40	12.66	12.90	13.12	13.32	13.50	13.68	13.84	14.00	14.15	14.28	14.42	15.00	15.48	16.24	16.85
40	4.08	5.42	6.24	6.84	7.31	7.71	8.04	8.33	8.59	8.83	9.04	9.24	9.42	9.59	9.75	9.89	10.03	10.17	10.29	10.41	10.94	11.37	12.06	12.61
	7.31	8.83	9.75	10.41	10.94	11.37	11.74	12.06	12.35	12.61	12.84	13.06	13.26	13.45	13.62	13.78	13.94	14.08	14.22	14.35	14.93	15.40	16.16	16.76
42	4.07	5.40	6.22	6.81	7.28	7.67	8.00	8.29	8.55	8.78	8.99	9.18	9.36	9.53	9.69	9.83	9.97	10.10	10.23	10.35	10.86	11.29	11.98	12.52
	7.28	8.78	9.69	10.35	10.86	11.29	11.66	11.98	12.26	12.52	12.75	12.96	13.16	13.34	13.51	13.67	13.82	13.97	14.10	14.23	14.80	15.27	16.02	16.60
44	4.06	5.39	6.19	6.78	7.25	7.63	7.96	8.25	8.51	8.74	8.94	9.14	9.31	9.48	9.63	9.78	9.92	10.05	10.17	10.29	10.80	11.22	11.90	12.43
	7.25	8.74	9.63	10.29	10.80	11.22	11.58	11.90	12.18	12.43	12.66	12.87	13.07	13.25	13.42	13.57	13.72	13.86	14.00	14.13	14.69	15.15	15.88	16.46
46	4.05	5.37	6.17	6.76	7.22	7.60	7.93	8.21	8.47	8.70	8.90	9.09	9.27	9.43	9.59	9.73	9.87	9.99	10.12	10.23	10.74	11.16	11.83	12.35
	7.22	8.70	9.59	10.23	10.74	11.16	11.52	11.83	12.11	12.35	12.58	12.79	12.98	13.16	13.33	13.48	13.63	13.77	13.90	14.03	14.58	15.04	15.76	16.33
48	4.04	5.35	6.15	6.74	7.19	7.57	7.90	8.18	8.43	8.66	8.87	9.05	9.23	9.39	9.54	9.69	9.82	9.95	10.07	10.18	10.69	11.10	11.76	12.29
	7.19	8.66	9.54	10.18	10.69	11.10	11.46	11.76	12.04	12.29	12.51	12.72	12.91	13.08	13.25	13.40	13.55	13.69	13.82	13.94	14.49	14.94	15.65	16.22
50	4.03	5.34	6.14	6.71	7.17	7.55	7.87	8.15	8.40	8.63	8.83	9.02	9.19	9.35	9.50	9.65	9.78	9.90	10.02	10.14	10.64	11.05	11.71	12.22
	7.17	8.63	9.50	10.14	10.64	11.05	11.40	11.71	11.98	12.22	12.44	12.65	12.84	13.01	13.17	13.33	13.47	13.61	13.74	13.86	14.40	14.85	15.56	16.11
60	4.00	5.29	6.07	6.63	7.08	7.44	7.76	8.03	8.28	8.49	8.69	8.88	9.04	9.20	9.35	9.48	9.61	9.74	9.85	9.96	10.44	10.84	11.48	11.97
	7.08	8.49	9.35	9.96	10.44	10.84	11.18	11.48	11.74	11.97	12.19	12.38	12.56	12.73	12.89	13.04	13.17	13.31	13.43	13.55	14.06	14.49	15.17	15.70
65	3.99	5.26	6.04	6.60	7.04	7.41	7.72	7.99	8.23	8.44	8.64	8.82	8.99	9.14	9.29	9.42	9.55	9.67	9.79	9.89	10.37	10.76	11.39	11.88
	7.04	8.44	9.29	9.89	10.37	10.76	11.10	11.39	11.65	11.88	12.09	12.28	12.46	12.63	12.78	12.93	13.06	13.19	13.31	13.43	13.94	14.36	15.02	15.54
70	3.98	5.25	6.02	6.57	7.01	7.37	7.68	7.95	8.19	8.40	8.60	8.78	8.94	9.09	9.24	9.37	9.50	9.62	9.73	9.84	10.31	10.70	11.32	11.80
	7.01	8.40	9.24	9.84	10.31	10.70	11.03	11.32	11.57	11.80	12.01	12.20	12.37	12.54	12.69	12.83	12.97	13.09	13.21	13.33	13.83	14.24	14.90	15.41
80	3.96	5.22	5.98	6.53	6.96	7.32	7.62	7.89	8.12	8.33	8.53	8.70	8.86	9.02	9.16	9.29	9.41	9.53	9.64	9.75	10.21	10.59	11.20	11.67
	6.96	8.33	9.16	9.75	10.21	10.59	10.91	11.20	11.45	11.67	11.88	12.06	12.23	12.39	12.54	12.68	12.81	12.94	13.06	13.17	13.66	14.06	14.70	15.20
100	3.94	5.18	5.93	6.47	6.90	7.25	7.54	7.80	8.03	8.24	8.43	8.60	8.76	8.91	9.04	9.17	9.29	9.41	9.52	9.62	10.07	10.44	11.03	11.50
	6.90	8.24	9.04	9.62	10.07	10.44	10.76	11.03	11.28	11.50	11.69	11.87	12.04	12.20	12.34	12.48	12.60	12.73	12.84	12.95	13.42	13.81	14.43	14.91
125	3.92	5.15	5.89	6.42	6.84	7.19	7.48	7.74	7.96	8.17	8.35	8.52	8.68	8.82	8.96	9.08	9.20	9.31	9.42	9.52	9.96	10.33	10.91	11.36
	6.84	8.17	8.96	9.52	9.96	10.33	10.64	10.91	11.14	11.36	11.55	11.73	11.89	12.04	12.18	12.32	12.44	12.56	12.67	12.78	13.24	13.62	14.22	14.69
150	3.90	5.13	5.86	6.39	6.81	7.15	7.44	7.69	7.92	8.12	8.30	8.47	8.62	8.76	8.90	9.02	9.14	9.25	9.36	9.46	9.89	10.25	10.82	11.27
	6.81	8.12	8.90	9.46	9.89	10.25	10.56	10.82	11.06	11.27	11.46	11.63	11.79	11.94	12.08	12.21	12.33	12.45	12.56	12.66	13.12	13.49	14.08	14.54
200	3.89	5.10	5.83	6.35	6.76	7.10	7.39	7.64	7.86	8.06	8.24	8.40	8.55	8.69	8.83	8.95	9.07	9.17	9.28	9.38	9.81	10.16	10.72	11.15
	6.76	8.06	8.83	9.38	9.81	10.16	10.46	10.72	10.95	11.15	11.34	11.51	11.67	11.82	11.95	12.08	12.20	12.31	12.42	12.52	12.97	13.33	13.91	14.36
400	3.86	5.06	5.78	6.30	6.70	7.03	7.31	7.56	7.77	7.97	8.14	8.31	8.45	8.59	8.72	8.84	8.95	9.06	9.16	9.26	9.68	10.02	10.57	10.99
	6.70	7.97	8.72	9.26	9.68	10.02	10.31	10.57	10.79	10.99	11.17	11.34	11.49	11.63	11.76	11.89	12.00	12.11	12.22	12.32	12.75	13.10	13.66	14.09
1000	3.85	5.04	5.75	6.26	6.66	6.99	7.27	7.51	7.72	7.91	8.09	8.25	8.39	8.53	8.66	8.78	8.89	8.99	9.09	9.19	9.60	9.94	10.47	10.89
	6.66	7.91	8.66	9.19	9.60	9.94	10.23	10.47	10.69	10.89	11.07	11.23	11.38	11.52	11.65	11.77	11.89	11.99	12.09	12.19	12.62	12.96	13.51	13.93
5000	3.84	5.03	5.73	6.24	6.64	6.97	7.24	7.48	7.70	7.89	8.06	8.22	8.36	8.50	8.62	8.74	8.85	8.96	9.06	9.15	9.56	9.90	10.43	10.84
	6.64	7.89	8.62	9.15	9.56	9.90	10.18	10.43	10.64	10.84	11.02	11.18	11.33	11.47	11.59	11.71	11.83	11.93	12.04	12.13	12.55	12.89	13.43	13.85

Table 7 Trend analysis: orthogonal polynomials. The table gives the coefficient of the orthogonal polynomial for up to a degree fourth.

		Number of groups = A										$\sum_a C_a^2$
A	Trend	1	2	3	4	5	6	7	8	9	10	
3	Linear	−1	0	1								2
	Quadratic	1	−2	1								6
4	Linear	−3	−1	1	3							20
	Quadratic	1	1	−1	1							4
	Cubic	−1	3	−3	1							20
5	Linear	−2	−1	0	1	2						10
	Quadratic	2	−1	−2	−1	2						14
	Cubic	−1	2	0	−2	1						10
	Quintic	1	−4	6	−4	1						70
6	Linear	−5	−3	−1	1	3	5					70
	Quadratic	5	−1	−4	−4	−1	5					84
	Cubic	−5	7	4	−4	−7	5					180
	Quintic	1	−3	2	2	−3	1					28
7	Linear	−3	−2	−1	0	1	2	3				28
	Quadratic	5	0	−3	−4	−3	0	5				84
	Cubic	−1	1	1	1	−1	−1	1				6
	Quintic	3	−7	1	6	1	−7	3				154
8	Linear	−7	−5	−3	−1	1	3	5	7			168
	Quadratic	7	1	−3	−5	−5	−3	1	7			168
	Cubic	−7	5	7	3	−3	−7	−5	7			264
	Quintic	7	= 13	−3	9	9	−3	−13	7			616
9	Linear	−4	−3	−2	−1	0	1	2	3	4		60
	Quadratic	28	7	−8	−17	−20	−17	−8	7	28		2772
	Cubic	−14	7	13	9	0	−9	−13	−7	14		990
	Quintic	14	−21	−11	9	18	9	−11	−21	14		2002
10	Linear	−9	−7	−5	−3	−1	1	3	5	7	9	330
	Quadratic	6	2	−1	−3	−4	−4	−3	−1	2	6	132
	Cubic	−42	14	35	31	12	−12	−31	−35	−14	42	8580
	Quintic	18	−22	−17	3	18	18	3	−17	−22	18	780

Table 8 Dunnett's test. Critical values of $F_{\text{Dunnett, critical}}$ for $\alpha = .05$ $\alpha = .01$.

		$R = \text{range (number of groups = A)}$								
ν_2	$\alpha[PF]$	2	3	4	5	6	7	8	9	10
6	.05	6.00	8.18	9.61	10.63	11.49	12.18	12.74	13.25	13.78
	.01	13.76	17.72	20.34	22.18	23.72	25.00	26.01	27.04	27.88
7	.05	5.57	7.56	8.82	9.73	10.50	11.09	11.63	12.04	12.46
	.01	12.25	15.60	17.72	19.27	20.52	21.53	22.47	23.23	23.91
8	.05	5.34	7.13	8.29	9.12	9.80	10.37	10.82	11.22	11.63
	.01	11.29	14.21	16.00	17.39	18.40	19.36	20.07	20.79	21.34
9	.05	5.11	6.81	7.90	8.70	9.30	9.86	10.24	10.63	11.02
	.01	10.56	13.18	14.82	16.08	16.97	17.81	18.49	19.10	19.62
10	.05	4.97	6.60	7.62	8.35	8.94	9.42	9.86	10.18	10.50
	.01	10.05	12.46	13.99	15.05	15.92	16.40	17.31	17.81	18.32
11	.05	4.84	6.40	7.40	8.07	8.64	9.12	9.49	9.86	10.18
	.01	9.67	11.90	13.32	14.36	15.13	15.84	16.40	16.89	17.31
12	.05	4.75	6.25	7.18	7.90	8.41	8.88	9.24	9.55	9.86
	.01	9.30	11.49	12.82	13.76	14.52	15.13	15.68	16.16	16.56
13	.05	4.76	6.15	7.02	7.73	8.24	8.64	9.00	9.36	9.61
	.01	9.06	11.09	12.39	13.32	13.99	14.59	15.13	15.52	15.92
14	.05	4.58	6.05	6.92	7.56	8.07	8.47	8.82	9.12	9.42
	.01	8.88	10.82	12.04	12.89	13.62	14.14	14.67	15.05	15.44
15	.05	5.54	5.95	6.81	7.45	7.95	8.35	8.70	9.00	9.24
	.01	8.70	10.56	11.76	12.60	13.25	13.76	14.29	14.67	15.05
16	.05	4.49	5.86	6.71	7.34	7.84	8.24	8.53	8.82	9.12
	.01	8.53	10.37	11.49	12.32	12.96	13.47	13.91	14.29	14.67
17	.05	4.45	5.81	6.66	7.24	7.73	8.12	8.41	8.70	9.00
	.01	8.41	10.18	11.29	12.04	12.67	13.18	13.62	13.99	14.36
18	.05	4.41	5.76	6.55	7.18	7.62	8.01	8.35	8.64	8.88
	.01	8.29	10.05	11.09	11.83	12.46	12.96	13.40	13.76	14.06
19	.05	4.37	5.71	6.50	7.08	7.56	7.90	8.24	8.53	8.76
	.01	8.18	9.92	10.96	11.70	12.25	12.74	13.18	13.54	13.84
20	.05	4.37	5.66	6.45	7.02	7.45	7.84	8.18	8.41	8.70
	.01	8.12	9.80	10.82	11.56	12.11	12.60	12.96	13.32	13.62
24	.05	4.24	5.52	6.30	6.81	7.29	7.62	7.90	8.18	8.41
	.01	7.84	9.42	10.37	11.02	11.56	12.04	12.39	12.74	13.03
30	.05	4.16	5.38	6.10	6.66	7.08	7.40	7.67	7.95	8.18
	.01	7.56	9.06	9.92	10.56	11.09	11.49	11.83	12.18	12.39
40	.05	4.08	5.24	5.95	6.45	6.86	7.18	7.45	7.67	7.90
	.01	7.29	8.70	9.55	10.18	10.63	11.02	11.36	11.63	11.83
60	.05	4.00	5.15	5.81	6.30	6.66	6.97	7.24	7.45	7.67
	.01	7.08	8.41	9.18	9.73	10.18	10.56	10.82	11.09	11.36
120	.05	3.92	5.02	5.66	6.10	6.50	6.76	7.02	7.24	7.45
	.01	6.86	8.1	8.82	9.36	9.73	10.11	10.37	10.63	10.82
∞	.05	3.84	4.88	5.52	5.95	6.30	6.60	6.81	7.02	7.24
	.01	6.66	7.78	8.53	9.00	9.36	9.67	9.92	10.18	10.37

Table 9 *F*_{range} **distribution.** (from studentized range *q*. $F = \frac{q^2}{2}$)
Table of the critical values of *F*_{range}. $\alpha = .05$ $\alpha = .01$

<i>R</i> = range (number of groups = <i>A</i>)														
ν_2	2	3	4	5	6	7	8	9	10	12	14	16	18	20
6	5.99	9.42	12.01	14.05	15.85	17.41	18.73	19.97	21.06	23.05	24.71	26.21	27.60	28.80
	13.73	20.03	24.71	28.58	31.76	34.61	37.07	39.34	41.40	41.40	44.94	50.80	53.25	55.55
7	5.58	8.65	10.95	12.80	14.36	15.74	16.94	18.00	18.97	20.67	22.18	23.46	24.64	25.70
	12.25	17.52	21.39	24.57	27.16	29.49	31.52	32.81	35.03	37.93	40.50	42.69	44.75	46.56
8	5.31	8.16	10.26	11.96	13.36	14.58	15.68	16.65	17.52	19.10	20.42	21.58	22.65	23.60
	11.28	15.90	19.22	21.91	24.22	26.21	27.90	29.49	30.89	33.46	35.62	37.50	39.16	40.77
9	5.12	7.80	9.72	11.33	12.60	13.73	14.74	15.62	16.47	17.88	19.16	20.22	21.19	22.04
	10.58	14.74	17.76	20.16	22.18	23.87	25.42	26.86	28.05	30.26	32.24	33.87	35.36	36.72
10	4.96	7.53	9.37	10.81	12.05	13.11	14.05	14.91	15.68	16.99	18.18	19.16	20.10	20.93
	10.04	13.89	16.65	18.85	20.67	22.24	23.60	24.85	25.99	28.05	29.72	31.28	32.64	33.87
11	4.84	7.30	9.07	10.44	11.62	12.65	13.52	14.31	15.07	16.30	17.41	18.36	19.22	20.03
	9.64	13.26	15.79	17.82	19.53	21.00	22.24	23.39	24.43	26.28	27.90	29.26	30.50	31.60
12	4.74	7.11	8.82	10.17	11.28	12.25	13.11	13.89	14.53	15.79	16.82	17.70	18.54	19.28
	9.33	12.75	15.13	17.05	18.60	19.97	21.19	22.24	23.19	24.92	26.43	27.68	28.80	29.88
13	4.68	6.96	8.61	9.90	11.00	11.91	12.75	13.47	14.15	15.29	16.30	17.17	18.00	18.67
	9.07	12.30	14.58	16.42	17.88	19.16	20.29	21.32	22.24	23.81	25.20	26.43	27.53	28.50
14	4.59	6.85	8.45	9.72	10.76	11.66	12.45	13.16	13.78	14.91	15.90	16.76	17.52	18.18
	8.86	11.96	14.15	15.85	17.29	18.48	19.59	20.54	21.39	22.92	24.22	25.42	26.43	27.38
15	4.53	6.73	8.32	9.55	10.53	11.42	12.20	12.90	13.52	14.58	15.51	16.36	17.11	17.76
	8.69	11.71	13.78	15.46	16.82	17.94	18.97	19.91	20.74	22.18	23.46	24.50	25.49	26.35
16	4.50	6.66	8.20	9.37	10.40	11.23	12.01	12.65	13.26	14.31	15.24	16.02	16.76	17.41
	8.53	11.47	13.47	15.07	16.36	17.52	18.48	19.34	20.16	21.52	22.71	23.81	24.71	25.56
17	4.44	6.59	8.08	9.24	10.22	11.05	11.81	12.45	13.06	14.10	14.96	15.74	16.42	17.05
	8.40	11.23	13.21	14.74	16.02	17.11	18.06	18.91	19.66	21.00	22.18	23.19	24.08	24.85
18	4.41	6.52	8.00	9.16	10.08	10.90	11.62	12.30	12.85	13.89	14.74	15.51	16.19	16.76
	8.28	11.05	12.95	14.47	15.68	16.76	17.64	18.48	19.22	20.54	21.65	22.65	23.46	24.29
19	4.38	6.44	7.92	9.03	9.99	10.81	11.47	12.10	12.70	13.68	14.53	15.29	15.96	16.53
	8.20	10.90	12.75	14.20	15.40	16.42	17.35	18.12	18.85	20.10	21.19	22.11	22.98	23.74
20	4.35	6.41	7.84	8.95	9.90	10.67	11.38	12.01	12.55	13.52	14.36	15.07	15.74	16.30
	8.08	10.76	12.60	13.99	15.18	16.19	17.05	17.82	18.54	19.78	20.80	21.71	22.51	23.26
24	4.26	6.23	7.60	8.69	9.55	10.31	10.95	11.57	12.10	13.00	13.78	14.47	14.80	15.62
	7.84	10.35	12.05	13.36	14.42	15.35	16.19	16.88	17.52	18.67	19.59	20.42	21.19	21.85
30	4.18	6.09	7.41	8.40	9.24	9.95	10.58	11.14	11.62	12.50	13.26	13.89	14.47	15.02
	7.57	9.90	11.52	12.75	13.73	14.58	15.35	15.96	16.59	17.58	18.48	19.22	19.91	20.54
40	4.09	5.92	7.18	8.16	8.95	9.64	10.22	10.72	11.19	12.01	12.70	13.31	13.89	14.36
	7.30	9.55	11.05	12.15	13.06	13.83	14.53	15.13	15.68	16.59	17.41	18.12	18.73	19.28
60	4.00	5.78	6.99	7.92	8.65	9.29	9.86	10.35	10.81	11.57	12.20	12.80	13.26	13.73
	7.07	9.16	10.53	11.62	12.45	13.16	13.78	14.36	14.85	15.68	16.42	17.05	17.58	18.12
120	3.92	5.64	6.77	7.68	8.40	8.99	9.50	9.99	10.40	11.09	11.71	12.25	12.70	13.16
	6.85	8.82	10.13	11.09	11.86	12.55	13.11	13.57	14.05	14.80	15.46	16.02	16.53	16.99
∞	3.84	5.48	6.59	7.45	8.12	8.69	9.20	9.64	9.99	10.67	11.23	11.76	12.15	12.55
	6.62	8.49	9.68	10.58	11.33	11.91	12.45	12.90	13.31	13.99	14.58	15.07	15.51	15.96

Table 10 Duncan's test. Critical values of Fisher's F . Corrected for multiple pairwise comparisons with Duncan's formula: $\alpha[PC] = 1 - (1 - \alpha[PF])^{\frac{1}{R-1}}$ (with R being the range) $\alpha = .05$ $\alpha = .01$

		R = range (i.e. number of means)																								
		$\alpha[PC]$ (corrected for the range)																								
$\alpha[PF] = .05$ $\alpha[PF] = .01$	ν_2	Critical values of F																								
		4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	26	31	41	51			
	4	7.71 21.20	12.12 31.29	15.53 39.04	18.41 45.58	20.95 51.34	23.25 56.56	25.36 61.35	27.34 65.81	29.19 70.00	30.94 73.96	32.60 77.73	34.20 81.33	35.72 84.79	37.19 88.11	38.61 91.32	39.98 94.42	41.31 97.43	42.60 100.35	43.86 103.19	45.08 105.96	50.79 118.85	55.95 130.51	65.10 151.21	73.17 169.45	
	5	6.61 16.26	9.94 22.76	12.38 27.49	14.37 31.35	16.09 34.67	17.62 37.61	19.00 40.27	20.26 42.71	21.44 44.97	22.54 47.09	23.57 49.08	24.56 50.97	25.49 52.76	26.38 54.48	27.23 56.12	28.06 57.70	28.85 59.22	29.61 60.69	30.35 62.11	31.06 63.48	34.35 69.80	37.27 75.41	42.32 85.12	46.67 93.46	
	6	5.99 13.75	8.75 18.62	10.72 22.04	12.29 24.77	13.62 27.08	14.78 29.09	15.82 30.90	16.77 32.54	17.64 34.05	18.45 35.45	19.21 36.76	19.93 37.99	20.60 39.16	21.24 40.26	21.86 41.32	22.44 42.33	23.00 43.30	23.54 44.23	24.06 45.13	24.57 46.00	26.86 49.95	28.86 53.41	32.28 59.30	35.17 64.28	
	7	5.59 12.25	8.02 16.22	9.71 18.94	11.03 21.08	12.14 22.87	13.10 24.41	13.96 25.78	14.73 27.02	15.44 28.15	16.09 29.20	16.70 30.17	17.27 31.09	17.81 31.95	18.32 32.76	18.80 33.53	19.26 34.27	19.70 34.98	20.12 35.65	20.53 36.30	20.92 36.93	22.70 39.76	24.23 42.22	26.83 46.36	28.99 49.81	
	8	5.32 11.26	7.53 14.67	9.03 16.97	10.20 18.76	11.17 20.23	12.00 21.50	12.74 22.62	13.40 23.62	14.01 24.54	14.56 25.38	15.08 26.16	15.56 26.89	16.02 27.57	16.45 28.22	16.85 28.83	17.24 29.42	17.61 29.97	17.96 30.51	18.30 31.02	18.62 31.51	20.09 33.72	21.35 35.62	23.46 38.81	25.21 41.44	
	9	5.12 10.56	7.17 13.60	8.54 15.62	9.60 17.17	10.48 18.44	11.23 19.53	11.89 20.48	12.48 21.34	13.01 22.11	13.50 22.83	13.96 23.48	14.38 24.10	14.78 24.67	15.16 25.22	15.51 25.73	15.85 26.21	16.17 26.68	16.48 27.12	16.77 27.55	17.05 27.95	18.32 29.78	19.40 31.34	21.21 33.94	22.69 36.07	
	10	4.96 10.04	6.90 12.82	8.18 14.63	9.16 16.02	9.97 17.15	10.66 18.11	11.26 17.05	11.80 17.81	12.28 18.49	12.73 19.10	13.14 19.66	13.52 20.18	13.88 20.66	14.22 21.10	14.54 21.52	14.84 21.92	15.13 22.29	15.40 22.65	15.66 22.99	15.92 23.31	17.04 25.01	18.00 26.19	19.59 28.12	20.89 29.70	
	11	4.84 9.65	6.69 12.22	7.90 13.89	8.82 15.15	9.58 16.18	10.22 17.05	10.78 17.81	11.27 18.49	11.72 19.10	12.14 19.66	12.51 20.18	12.87 20.66	13.20 21.10	13.51 21.52	13.80 21.92	14.07 22.29	14.34 22.65	14.59 22.99	14.83 23.31	15.06 23.62	16.08 25.01	16.95 26.19	18.39 28.12	19.55 29.70	
	12	4.75 9.33	6.52 11.74	7.67 13.30	8.55 14.47	9.26 15.42	9.87 16.23	10.39 16.93	10.86 17.55	11.28 18.11	11.67 18.62	12.02 19.09	12.35 19.53	12.66 19.93	12.95 20.31	13.22 20.67	13.48 21.01	13.72 21.34	13.95 21.64	14.18 21.94	14.39 22.22	15.34 23.47	16.14 24.53	17.45 26.26	18.52 27.67	
	13	4.67 9.07	6.38 11.36	7.49 12.83	8.33 13.93	9.01 14.82	9.59 15.57	10.09 16.22	10.53 16.80	10.93 17.32	11.29 17.79	11.63 18.23	11.94 18.63	12.23 19.01	12.50 19.36	12.75 19.69	13.00 20.00	13.23 20.30	13.44 20.58	13.65 20.85	13.85 21.11	14.74 22.26	15.49 23.23	16.71 24.81	17.70 26.09	
	14	4.60 8.86	6.26 11.05	7.34 12.45	8.15 13.49	8.80 14.33	9.35 15.04	9.83 15.65	10.25 16.19	10.64 16.68	10.98 17.12	11.30 17.53	11.60 17.91	11.87 18.26	12.13 18.59	12.37 18.90	12.60 19.19	12.82 19.47	13.03 19.73	13.22 19.98	13.41 20.22	14.25 21.29	14.96 22.18	16.11 23.65	17.04 24.82	
	15	4.54 8.68	6.17 10.79	7.21 12.12	7.99 13.12	8.62 13.92	9.16 14.59	9.62 15.17	10.02 15.69	10.39 16.15	10.72 16.57	11.03 16.95	11.31 17.31	11.58 17.64	11.82 17.95	12.06 18.24	12.27 18.52	12.48 18.78	12.68 19.02	12.87 19.26	13.05 19.49	13.84 20.48	14.52 21.32	15.61 22.69	16.49 23.79	
	16	4.49 8.53	6.08 10.57	7.10 11.85	7.86 12.81	8.47 13.57	8.99 14.22	9.43 14.77	9.83 15.26	10.18 15.70	10.50 16.10	10.80 16.47	11.07 16.81	11.33 17.12	11.56 17.42	11.79 17.69	12.00 17.95	12.20 18.20	12.38 18.43	12.57 18.66	12.74 18.87	13.50 19.82	14.14 20.61	15.19 21.90	16.02 22.93	
	17	4.45 8.40	6.01 10.38	7.01 11.62	7.75 12.54	8.34 13.28	8.85 13.90	9.28 14.43	9.66 14.90	10.00 15.32	10.32 15.71	10.60 16.06	10.86 16.38	11.11 16.68	11.34 16.96	11.55 17.22	11.76 17.47	11.95 17.71	12.13 17.93	12.31 18.15	12.47 18.35	13.21 19.25	13.83 20.00	14.83 21.23	15.63 22.20	
	18	4.41 8.29	5.95 10.21	6.92 11.42	7.65 12.31	8.23 13.02	8.72 13.62	9.14 14.14	9.51 14.59	9.85 15.00	10.15 15.36	10.43 15.70	10.68 16.01	10.92 16.30	11.15 16.57	11.35 16.82	11.55 17.06	11.74 17.29	11.91 17.50	12.08 17.71	12.24 17.90	12.96 18.76	13.55 19.48	14.52 20.65	15.29 21.58	
	19	4.38 8.18	5.89 10.07	6.85 11.24	7.56 12.11	8.13 12.80	8.61 13.38	9.02 13.88	9.39 14.32	9.71 14.71	10.01 15.07	10.28 15.39	10.53 15.69	10.76 15.97	10.98 16.23	11.18 16.47	11.37 16.70	11.55 16.92	11.72 17.13	11.89 17.33	12.04 17.51	12.74 18.34	13.31 19.03	14.25 20.15	15.00 21.05	
	20	4.35 8.10	5.84 10.04	6.78 11.08	7.48 11.93	8.04 12.60	8.51 13.17	8.92 13.65	9.27 14.08	9.59 14.46	9.88 14.80	10.14 15.12	10.39 15.41	10.61 15.68	10.83 15.93	11.02 16.17	11.21 16.39	11.39 16.60	11.56 16.80	11.72 16.99	11.87 17.17	12.54 17.97	13.10 18.64	14.01 19.72	14.74 20.58	

Table 10 Duncan's test. Critical values of Fisher's *F*. Corrected for multiple pairwise comparisons with Duncan's formula: $\alpha[PC] = 1 - (1 - \alpha[PF])^{\frac{1}{r-1}}$ (with *R* being the range) (continued). $\alpha = .05$ $\alpha = .01$

		R = range (i.e. number of means)																									
		$\alpha[PC]$ (corrected for the range)																									
		Critical values of F																									
ν_2		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	26	31	41	51		
$\alpha[PF] = .05$ $\alpha[PF] = .01$	0.0500	0.0253	0.0170	0.0127	0.0102	0.0085	0.0073	0.0064	0.0057	0.0051	0.0047	0.0043	0.0039	0.0037	0.0034	0.0032	0.0030	0.0028	0.0027	0.0026	0.0020	0.0017	0.0013	0.0011	0.0010		
	0.0100	0.0050	0.0033	0.0025	0.0020	0.0017	0.0014	0.0013	0.0011	0.0010	0.0009	0.0008	0.0008	0.0007	0.0007	0.0007	0.0006	0.0006	0.0006	0.0005	0.0004	0.0003	0.0003	0.0003	0.0002		
21	4.32	5.80	6.73	7.41	7.97	8.43	8.82	9.17	9.48	9.77	10.03	10.26	10.49	10.69	10.89	11.07	11.24	11.41	11.56	11.71	12.37	12.92	13.80	14.51			
22	8.02	9.82	10.94	11.77	12.43	12.98	13.45	13.87	14.24	14.57	14.88	15.16	15.42	15.67	15.90	16.11	16.32	16.51	16.70	16.87	17.65	18.29	19.34	20.17			
	4.30	5.76	6.67	7.35	7.89	8.35	8.74	9.08	9.39	9.67	9.92	10.15	10.37	10.57	10.76	10.94	11.11	11.27	11.43	11.57	12.21	12.75	13.61	14.30			
23	7.95	9.72	10.82	11.63	12.27	12.81	13.27	13.68	14.04	14.37	14.67	14.94	15.20	15.43	15.66	15.87	16.07	16.26	16.43	16.61	17.36	17.99	19.00	19.80			
	4.28	5.72	6.63	7.30	7.83	8.28	8.66	9.00	9.30	9.57	9.82	10.05	10.27	10.47	10.65	10.83	11.00	11.15	11.30	11.45	12.07	12.60	13.45	14.12			
24	7.88	9.63	10.71	11.50	12.13	12.66	13.11	13.51	13.86	14.18	14.47	14.74	14.99	15.22	15.44	15.65	15.84	16.02	16.20	16.37	17.10	17.71	18.70	19.48			
	4.26	5.69	6.58	7.25	7.77	8.21	8.59	8.93	9.22	9.49	9.74	9.96	10.17	10.37	10.55	10.73	10.89	11.04	11.19	11.33	11.95	12.46	13.29	13.95			
25	7.82	9.54	10.61	11.39	12.01	12.52	12.97	13.35	13.70	14.01	14.30	14.56	14.81	15.04	15.25	15.45	15.64	15.82	15.99	16.15	16.87	17.46	18.42	19.18			
	4.24	5.66	6.55	7.20	7.72	8.16	8.53	8.86	9.15	9.42	9.66	9.88	10.09	10.28	10.46	10.63	10.79	10.95	11.09	11.23	11.84	12.34	13.16	13.80			
26	7.77	9.47	10.52	11.28	11.89	12.40	12.83	13.22	13.56	13.86	14.14	14.40	14.64	14.86	15.07	15.27	15.45	15.63	15.80	15.96	16.66	17.24	18.18	18.92			
	4.23	5.63	6.51	7.16	7.67	8.10	8.47	8.80	9.09	9.35	9.59	9.81	10.01	10.20	10.38	10.55	10.71	10.86	11.00	11.13	11.73	12.23	13.03	13.67			
27	7.72	9.40	10.43	11.19	11.79	12.29	12.71	13.09	13.42	13.73	14.00	14.25	14.49	14.71	14.91	15.11	15.29	15.46	15.62	15.78	16.47	17.04	17.95	18.68			
	4.21	5.61	6.48	7.12	7.63	8.05	8.42	8.74	9.03	9.29	9.52	9.74	9.94	10.13	10.30	10.47	10.63	10.77	10.91	11.05	11.64	12.13	12.92	13.54			
28	7.68	9.34	10.36	11.10	11.69	12.18	12.61	12.97	13.30	13.60	13.87	14.12	14.35	14.57	14.77	14.96	15.14	15.30	15.47	15.62	16.29	16.85	17.75	18.46			
	4.20	5.58	6.45	7.08	7.59	8.01	8.37	8.69	8.97	9.23	9.46	9.68	9.87	10.06	10.23	10.40	10.55	10.70	10.84	10.97	11.55	12.03	12.81	13.43			
29	7.64	9.28	10.29	11.02	11.60	12.09	12.50	12.87	13.19	13.48	13.75	14.00	14.22	14.44	14.63	14.82	15.00	15.16	15.32	15.47	16.13	16.68	17.57	18.26			
	4.18	5.56	6.42	7.05	7.55	7.97	8.33	8.64	8.92	9.17	9.40	9.62	9.81	10.00	10.17	10.33	10.48	10.63	10.76	10.90	11.47	11.95	12.72	13.32			
30	7.60	9.22	10.22	10.95	11.52	12.00	12.41	12.77	13.09	13.38	13.64	13.88	14.11	14.32	14.51	14.69	14.87	15.03	15.19	15.33	15.99	16.53	17.39	18.08			
	4.17	5.54	6.39	7.02	7.52	7.93	8.29	8.60	8.87	9.12	9.35	9.56	9.76	9.94	10.11	10.27	10.42	10.56	10.70	10.83	11.40	11.87	12.63	13.23			
31	7.56	9.17	10.16	10.88	11.45	11.92	12.33	12.68	13.00	13.28	13.54	13.78	14.00	14.20	14.40	14.58	14.75	14.91	15.06	15.21	15.85	16.38	17.24	17.91			
	4.16	5.52	6.37	6.99	7.48	7.89	8.25	8.56	8.83	9.08	9.31	9.51	9.71	9.89	10.05	10.21	10.36	10.50	10.64	10.77	11.33	11.79	12.54	13.13			
32	7.53	9.13	10.10	10.82	11.38	11.85	12.25	12.60	12.91	13.19	13.45	13.68	13.90	14.10	14.29	14.47	14.64	14.80	14.95	15.09	15.73	16.25	17.09	17.76			
	4.15	5.50	6.35	6.96	7.45	7.86	8.21	8.52	8.79	9.04	9.26	9.47	9.66	9.84	10.00	10.16	10.31	10.45	10.58	10.71	11.26	11.73	12.47	13.05			
33	7.50	9.08	10.05	10.76	11.32	11.78	12.17	12.52	12.83	13.11	13.36	13.59	13.81	14.01	14.19	14.37	14.54	14.69	14.84	14.99	15.61	16.13	16.96	17.61			
	4.14	5.49	6.33	6.94	7.43	7.83	8.18	8.48	8.75	9.00	9.22	9.42	9.61	9.79	9.95	10.11	10.26	10.40	10.53	10.65	11.20	11.66	12.39	12.97			
34	7.47	9.04	10.00	10.70	11.25	11.71	12.10	12.45	12.75	13.03	13.28	13.51	13.72	13.92	14.10	14.28	14.44	14.60	14.74	14.88	15.50	16.01	16.83	17.48			
	4.13	5.47	6.31	6.92	7.40	7.80	8.15	8.45	8.72	8.96	9.18	9.38	9.57	9.75	9.91	10.06	10.21	10.35	10.48	10.60	11.15	11.60	12.33	12.90			
35	7.44	9.01	9.96	10.65	11.20	11.65	12.04	12.38	12.68	12.95	13.20	13.43	13.64	13.84	14.02	14.19	14.35	14.51	14.65	14.79	15.40	15.91	16.71	17.35			

35	4.12	5.46	6.29	9.92	8.97	7.42	7.37	7.77	8.12	8.42	8.68	8.92	9.14	9.34	9.53	9.70	9.87	10.02	10.16	10.30	10.43	10.55	11.10	11.54	12.26	12.83
36	4.11	5.45	6.27	9.88	8.94	7.40	7.35	7.75	8.09	8.39	8.65	8.89	9.11	9.31	9.49	9.67	9.83	9.98	10.12	10.26	10.39	10.51	11.05	11.49	12.21	12.77
37	4.11	5.43	6.25	9.84	8.91	7.37	7.33	7.72	8.06	8.36	8.62	8.86	9.08	9.27	9.46	9.63	9.79	9.94	10.08	10.22	10.35	10.47	11.00	11.44	12.15	12.71
38	4.10	5.42	6.24	9.84	8.91	7.37	7.31	7.70	8.04	8.33	8.59	8.83	9.05	9.24	9.43	9.60	9.76	9.91	10.05	10.18	10.31	10.43	10.96	11.40	12.10	12.65
39	7.35	8.88	9.80	10.48	11.01	11.45	11.82	12.15	12.44	12.70	12.94	13.16	13.36	13.55	13.73	13.89	14.05	14.20	14.34	14.47	15.06	15.54	16.32	16.93		
40	4.09	5.41	6.22	9.84	8.91	7.37	7.29	7.68	8.01	8.31	8.57	8.80	9.02	9.21	9.39	9.56	9.72	9.87	10.01	10.14	10.27	10.39	10.92	11.35	12.05	12.60
41	7.33	8.85	9.77	10.44	10.97	11.40	11.77	12.10	12.39	12.65	12.89	13.10	13.30	13.49	13.67	13.83	13.99	14.13	14.27	14.40	14.98	15.46	16.23	16.84		
42	4.08	5.40	6.21	9.80	8.87	7.35	7.27	7.66	7.99	8.28	8.54	8.78	8.99	9.18	9.37	9.53	9.69	9.84	9.98	10.11	10.24	10.35	10.88	11.31	12.00	12.55
43	7.31	8.82	9.74	10.40	10.93	11.36	11.73	12.05	12.34	12.60	12.83	13.05	13.25	13.43	13.61	13.77	13.93	14.07	14.21	14.34	14.92	15.39	16.15	16.75		
44	4.07	5.38	6.18	9.77	8.84	7.32	7.24	7.62	7.95	8.24	8.50	8.73	8.94	9.13	9.31	9.48	9.63	9.78	9.92	10.05	10.17	10.29	10.81	11.23	11.92	12.45
45	7.28	8.77	9.68	10.34	10.87	11.28	11.65	11.97	12.25	12.50	12.74	12.95	13.15	13.33	13.50	13.66	13.81	13.96	14.09	14.22	14.79	15.26	16.00	16.59		
46	4.06	5.36	6.16	9.74	8.81	7.29	7.21	7.59	7.92	8.20	8.46	8.68	8.89	9.09	9.26	9.43	9.58	9.73	9.86	9.99	10.11	10.23	10.74	11.16	11.84	12.37
47	7.25	8.73	9.63	10.28	10.79	11.21	11.57	11.89	12.17	12.42	12.65	12.86	13.05	13.23	13.40	13.56	13.71	13.85	13.99	14.11	14.67	15.13	15.87	16.45		
48	4.05	5.34	6.14	9.72	8.79	7.27	7.18	7.56	7.88	8.17	8.42	8.65	8.85	9.04	9.22	9.38	9.53	9.68	9.81	9.94	10.06	10.18	10.68	11.10	11.77	12.30
49	7.22	8.69	9.58	10.22	10.73	11.15	11.51	11.82	12.10	12.34	12.57	12.78	12.97	13.15	13.32	13.47	13.62	13.76	13.89	14.02	14.57	15.02	15.75	16.32		
50	4.04	5.33	6.12	9.70	8.77	7.25	7.15	7.53	7.85	8.13	8.38	8.61	8.81	9.00	9.18	9.34	9.49	9.63	9.77	9.89	10.01	10.13	10.63	11.04	11.71	12.23
51	7.19	8.65	9.54	10.17	10.68	11.09	11.45	11.75	12.03	12.27	12.50	12.70	12.89	13.07	13.24	13.39	13.54	13.67	13.81	13.93	14.48	14.93	15.64	16.21		
52	4.03	5.32	6.10	9.68	8.75	7.23	7.13	7.50	7.82	8.10	8.35	8.58	8.78	8.97	9.14	9.30	9.45	9.59	9.73	9.85	9.97	10.08	10.58	10.99	11.65	12.16
53	7.17	8.62	9.50	10.13	10.63	11.04	11.39	11.70	11.97	12.21	12.43	12.64	12.83	13.00	13.16	13.32	13.46	13.60	13.73	13.85	14.39	14.83	15.54	16.10		
54	4.00	5.26	6.03	9.65	8.72	7.20	7.10	7.47	7.79	8.07	8.32	8.54	8.74	8.93	9.10	9.26	9.41	9.56	9.68	9.80	9.91	10.39	10.79	11.42	11.92	
55	7.08	8.49	9.34	9.95	10.44	10.83	11.17	11.47	11.73	11.96	12.18	12.37	12.55	12.72	12.88	13.03	13.16	13.29	13.42	13.54	14.05	14.48	15.16	15.69		
56	3.99	5.24	6.01	9.60	8.67	7.15	7.06	7.43	7.75	8.03	8.28	8.50	8.70	8.89	9.06	9.22	9.37	9.50	9.62	9.73	9.84	10.32	10.71	11.33	11.82	
57	7.04	8.44	9.28	9.89	10.36	10.75	11.09	11.38	11.64	11.87	12.08	12.27	12.45	12.62	12.77	12.92	13.05	13.18	13.30	13.42	13.93	14.34	15.01	15.53		
58	3.98	5.22	5.98	9.23	8.30	6.97	7.33	7.64	7.90	8.14	8.36	8.55	8.73	8.89	9.05	9.19	9.32	9.45	9.57	9.68	9.79	10.26	10.64	11.26	11.74	
59	7.01	8.40	9.23	9.83	10.30	10.69	11.02	11.31	11.56	11.79	12.00	12.19	12.36	12.53	12.68	12.82	12.96	13.08	13.20	13.32	13.82	14.23	14.89	15.40		
60	3.96	5.19	5.95	9.23	8.30	6.92	7.28	7.58	7.84	8.08	8.29	8.48	8.65	8.82	8.97	9.11	9.24	9.36	9.48	9.59	9.70	10.16	10.54	11.14	11.62	
61	6.96	8.33	9.15	9.74	10.20	10.58	10.91	11.19	11.44	11.66	11.87	12.05	12.22	12.38	12.53	12.67	12.80	12.93	13.05	13.16	13.65	14.05	14.69	15.19		
62	3.94	5.16	5.90	9.20	8.27	6.86	7.20	7.50	7.76	7.99	8.20	8.38	8.55	8.71	8.86	9.00	9.13	9.25	9.36	9.47	9.57	10.02	10.39	10.98	11.44	
63	6.90	8.24	9.04	9.61	10.06	10.44	10.75	11.03	11.27	11.49	11.68	11.87	12.03	12.19	12.33	12.47	12.59	12.72	12.83	12.94	13.41	13.80	14.42	14.90		
64	3.92	5.12	5.86	9.16	8.23	6.80	7.15	7.44	7.69	7.92	8.12	8.31	8.47	8.63	8.77	8.91	9.04	9.15	9.27	9.37	9.47	9.92	10.28	10.86	11.31	
65	6.84	8.16	8.95	9.51	9.96	10.32	10.63	10.90	11.13	11.35	11.54	11.72	11.88	12.03	12.17	12.31	12.43	12.55	12.66	12.77	13.23	13.61	14.21	14.68		
66	3.90	5.10	5.83	9.16	8.23	6.77	7.11	7.40	7.65	7.87	8.07	8.26	8.42	8.58	8.72	8.85	8.98	9.09	9.20	9.31	9.41	9.85	10.20	10.77	11.22	
67	6.81	8.11	8.89	9.45	9.89	10.24	10.55	10.81	11.05	11.26	11.45	11.62	11.78	11.93	12.07	12.20	12.32	12.44	12.55	12.65	13.11	13.48	14.07	14.53		
68	3.89	5.08	5.80	9.12	8.19	6.73	7.06	7.35	7.60	7.82	8.01	8.19	8.36	8.51	8.65	8.78	8.90	9.02	9.13	9.23	9.33	9.76	10.11	11.11	11.60	
69	6.76	8.05	8.82	9.37	9.80	10.15	10.45	10.71	10.94	11.15	11.33	11.50	11.66	11.81	11.94	12.07	12.19	12.30	12.41	12.51	12.96	13.32	13.90	14.35		
70	3.86	5.04	5.75	9.14	8.21	6.66	6.99	7.27	7.52	7.73	7.93	8.10	8.26	8.41	8.55	8.68	8.80	8.91	9.02	9.12	9.21	9.63	9.97	10.52	10.94	
71	6.70	7.96	8.71	9.25	9.67	10.01	10.30	10.56	10.78	10.98	11.16	11.33	11.48	11.62	11.76	11.88	12.00	12.10	12.21	12.31	12.74	13.09	13.65	14.08		
72	3.85	5.02	5.72	9.14	8.21	6.62	6.95	7.23	7.47	7.68	7.87	8.05	8.20	8.35	8.49	8.61	8.73	8.84	8.95	9.05	9.14	9.55	9.89	10.43	10.85	
73	6.66	7.91	8.65	9.18	9.59	9.93	10.22	10.47	10.69	10.88	11.06	11.23	11.38	11.51	11.64	11.77	11.88	11.99	12.09	12.19	12.61	12.95	13.50	13.92		
74	3.84	5.00	5.71	9.14	8.21	6.60	6.93	7.20	7.44	7.65	7.84	8.02	8.17	8.32	8.45	8.58	8.70	8.81	8.91	9.01	9.11	9.51	9.85	10.38	10.79	
75	6.64	7.88	8.62	9.14	9.55	9.89	10.17	10.42	10.64	10.83	11.01	11.17	11.32	11.46	11.59	11.71	11.82	11.93	12.03	12.12	12.54	12.88	13.42	13.84		

References

- ABDI H. (1987). *Introduction au traitement des données expérimentales*. Grenoble: Presses Universitaires de Grenoble.
- ABDI H. (2003). Partial regression coefficients. In M. Lewis-Beck, A. Bryman, T. Futing (Eds): *Encyclopedia for research methods for the social sciences*, pp. 978–982. Thousand Oaks (CA): Sage.
- ABDI H. (2007). O'Brien test for homogeneity of variance. In N.J. Salkind (Ed.): *Encyclopedia of measurement and statistics*. Thousand Oaks (CA): Sage. pp. 701–704.
- ABDI H., MOLIN P. (2007). Lilliefors test of normality In N.J. Salkind (Ed.): *Encyclopedia of measurement and statistics*. Thousand Oaks (CA): Sage. pp. 540–544.
- ADORNO T.W., FRENKELBRUNSWIK E., LEVINSON D.I., SANFORD R.N. (1950). *The authoritarian personality*. New York: Harper.
- ALGINA J., KESELMAN H.J. (1997). Detecting repeated measures effects with univariate and multivariate statistics. *Psychological Methods*, **2**, 208–218.
- ANDERSON J.R. (1980). *Cognitive psychology*. New York: Freeman.
- APPELRAUM M.I., MCCALL R.B. (1983). *Design and analysis in developmental Psychology*. In Mussen P.H. (Ed.): *Handbook of child psychology*. New York: Wiley.
- ARMOR D.J. (1972). School and family effects on black and white achievement. In F. Mosteller, D.P. Moynihan (Eds.): *On equality of educational opportunity*. New York: Vintage Books.
- ATKINSON R.C., HOLMGREN J.E., JUOLA J.F. (1969). Processing time as influenced by the number of elements in a visual display. *Perception and Psychophysics*, **6**, 321–327.
- BADDELEY A.D. (1966). Short term memory for word sequence as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, **18**, 362–365.
- BADDELEY A.D. (1975). Theory of amnesia. In A. Kennedy, A. Wilkes (Eds.): *Studies in long term memory*. London: Wiley.
- BADDELEY A.D. (1976). *The psychology of memory*. New York: Basic Books.
- BADDELEY A.D. (1990). *Human memory: theory and practice*. Boston: Allyn and Bacon.
- BADDELEY A.D. (1994). *Your memory: A user's guide*. London: Carlton Books Ltd.
- BAHRICK H.P. (1984). Semantic memory in permastore: fifty years of memory for spanish learned in school. *Journal of Experimental Psychology: General*, **113**, 1–19.
- BAKEN D. (1966). The test of significance in psychological research. *Psychological Bulletin*, **66**, 423–427.
- BARNETT V., LEWIS T. (1994). *Outliers in statistical data* (3rd edition). Chichester: Wiley.
- BARTEN A.P. (1962). Note on unbiased estimation of the squared multiple correlation coefficient. *Statistica Neerlandica*, **16**, 151–163.
- BARTRAM D.J. (1973). The effects of familiarity and practice on naming picture of objects. *Memory and Cognition*, **1**, 101–105.
- BECHHOFFER R.E., DUNNETT C.W. (1982). Multiple comparisons for orthogonal contrasts: examples and tables. *Technometrics*, **24**, 213–222.

- BENNET C.A., FRANKLIN N.L. (1954) *Statistical analysis in chemistry and the chemical industry*. New York: Wiley.
- BEVAN M.F., DENTON J.Q., MYERS J.L. (1974). The robustness of the F test to violations of continuity and form of treatment population. *British Journal of Mathematical and Statistical Psychology*, **27**, 199–204.
- BINDER A. (1963). Further considerations on testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychological Bulletin*, **70**, 107–115.
- BLAKE M.J.F. (1967). Time of the day effects on performance in a range of tasks. *Psychonomic Science*, **9**, 349–350.
- BOIK R.J. (1975). Interactions in the analysis of variance: a procedure for interpretation and a Monte-Carlo comparison of univariate and multivariate methods for repeated measure designs. Ph. D. Thesis, Baylor University.
- BOIK R.J. (1979). Interactions, partial interactions and interaction contrasts in the analysis of variance. *Psychological Bulletin*, **86**, 1084–1089.
- BOIK R.J. (1981). A priori tests in repeated measures design: effect of non-sphericity. *Psychometrika*, **46**, 241–255.
- BOWER G.H. (1972). Mental imagery and associative learning. In L.W. Gregg (Ed.): *Cognition in learning and memory*. New York: Wiley.
- BOX G.E.P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems. II: Effects of inequality of variance and correlation between errors in the two-way classification. *Annals of Mathematical Statistics*, **25**, 484–498.
- BOX G.E.P., COX D.R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, **26**, 211–252.
- BRADY J.V. (1958). Ulcers in executive monkeys. *Scientific American*, **199**, 95–100.
- BRANSFORD J.D. (1979). *Human cognition*. Belmont: Wadsworth.
- BRANSFORD J.D., BARCLAY J.R., FRANK J.J. (1972). Sentence memory: a constructive versus interpretative approach. *Cognitive Psychology*, **3**, 193–209.
- BRANSFORD J.D., JOHNSON M.K. (1972). Contextual prerequisites for understanding: some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, **11**, 717–726.
- BROOKS L.R. (1967). The suppression of visualization in reading. *Quarterly Journal of Experimental Psychology*, **19**, 289–299.
- BROOKS L.R. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology*, **22**, 349–368.
- BROWN, H., PRESCOTT, R. (1999). *Applied mixed models in medicine*. London: Wiley.
- BROWN, R. W., BLACK, A. H., HOROWITZ, A. E. (1955). Phonetic symbolism in natural languages. *Journal of Abnormal and Social Psychology*, **50**, 388–393.
- BRUNSWICK E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.
- BUDESCU D.V., APPELBAUM M.I. (1981). Variance stabilizing transformations and the power of the F test. *Journal of Educational Statistics*, **6**, 55–74.
- CAMPBELL D.T., STANLEY J.C. (1966). *Experimental and quasi-experimental design research*. Skokie: Rand McNally.

- CAMPBELL S.K. (1974). *Flaws and fallacies in statistical thinking*. Englewood Cliffs (NJ): Prentice-Hall.
- CARROL R.M, NORDHOLM L.A. (1975). Sampling characteristic of Kelley η^2 and Hays ω^2 . *Educational and Psychological Measurement*, **35**, 541–554.
- CHANQUOY, L. (2005). *Statistiques appliquées à la psychologie et aux sciences humaines et sociales*. Paris: Hachette.
- CHASTAING M. (1958). Le symbolisme des voyelles. *Journal de Psychologie Normale et Pathologique*, **55**, 403–23 et 461–481.
- CHASTAING M. (1986). Comment jouer avec des enfants à associer des mots. *Journal de Psychologie Normale et Pathologique*, **1–2**, 42–63.
- CHATEAU J. (1972). *Le malaise de la psychologie*. Paris: Flammarion.
- CHERULNIK P. D. (1983). *Behavioral research*. Cambridge: Harper & Row.
- CLARK H.H. (1973). The language as a Fixed-effect Fallacy: a critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, **12**, 335–359.
- CLARK H.H. (1976). Reply to Wike and Church. *Journal of Verbal Learning and Verbal Behavior*, **15**, 257–261.
- COHEN J. (1965). Some statistical issues in psychological research. In B.B. Wolman (Ed.), *Handbook of clinical psychology*. New York: McGraw-Hill.
- COHEN J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- COHEN J. (1973). Eta squared and partial eta squared in fixed factor ANOVA design. *Educational and Psychological Measurement*, **33**, 107–112.
- COHEN J. (1977). *Statistical power analysis methods in behavioral research* (revised edition). New York: Academic Press.
- COHEN J. (1982). A new look. In G. Keren (Ed.): *Statistical and methodological issues in psychology and social sciences research*. Hillsdale (NJ): Erlbaum.
- COHEN J., COHEN P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale (NJ): Erlbaum.
- COLEMAN J., CAMPBELL E., HOBSON C., PARTLAND J., MOOD A., WEINFELD F., YORK R. (1966). *Equality of educational opportunity*. Washington (DC): US office of Education.
- COLLINS A.M., QUILLIAN M.R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, **8**, 240–248.
- COLLINS A.M., QUILLIAN M.R. (1970). Does category size affect categorization time? *Journal of Verbal Learning and Verbal Behavior*, **9**, 432–438.
- CONOVER W.J. (1971). *Practical nonparametric statistics*. New York: Wiley.
- CONRAD E., MAUD T. (1981). *Introduction to experimental psychology*. New York: Wiley.
- CONRAD R. (1972). The developmental role of vocalizing in short term memory. *Journal of Verbal Learning and Verbal Behavior*, **11**, 521–533.
- COOK T.D., CAMPBELL D.T. (1979). *Quasi experimentation: design and analysis issues for field settings*. Chicago: Rand McNally.
- CORDIER F. (1985). Formal and locative categories: are there typical instances. *Psychologica Belgica*, **25**, 115–125.

- CORNFIELD J., TUKEY J.W. (1956). Average values of mean squares in factorials. *Annals of Mathematical Statistics*, **27**, 907–949.
- COWLES M., DAVIS C. (1987). The subject matter of psychology volunteers. *British Journal of Psychology*, **26**, 97–102.
- COZBY P.C. (1977). *Methods in behavioral research*. Palo Alto (CA): Mayfield.
- CRAIK F.I., TULVING E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, **104**, 268–294.
- CROWDER M.J., HAND D.J. (1990). *Analysis of repeated measures*. London: Chapman & Hall.
- CRUMP S.L. (1946). The estimation of variance components in analysis of variance. *Biometrics*, **6**, 7–11.
- DAGNELIE P. (1968). A propos de l'emploi du test de Kolmogorov-Smirnov comme test de normalité. *Biométrie et Praximétrie*, **9**, 3–13.
- DARLINGTON R.B. (1990). *Regression and linear models*. New York: McGraw-Hill.
- DAVENPORT J.M., WEBSTER J.T. (1973). A comparison of some approximate *F* tests. *Technometrics*, **15**, 779–789.
- DE GROOT A.D. (1965). *Thought and choice in chess*. New York: Basic Books.
- DEVLIN K. (2008). *The unfinished game: Pascal, Fermat, and the seventeenth-century letter that made the world modern*. New York: Basic Books.
- DOOLING D.J., DANKS J.H. (1975). Going beyond tests of significance. Is psychology ready? *Bulletin of the Psychonomic Society*, **5**, 15–17.
- DRAPER N.R., SMITH H. (1981). *Applied regression analysis*. New York: Wiley.
- DUNCAN D.B. (1955). Multiple range and multiple *F* tests. *Biometrics*, **11**, 1–42.
- DUNCAN D.B. (1958). Multiple range tests for correlated and heteroscedastic means. *Biometrics*, **13**, 164–76.
- DUNN O.J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, **56**, 52–64.
- DUNN O.J. (1974). On multiple tests and confidence intervals. *Communications in Statistics*, **3**, 101–103.
- DUNNETT C.W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, **50**, 1096–1121.
- DUNNETT C.W. (1964). New tables for multiple comparisons with a control. *Biometrics*, **20**, 482–91.
- DUNNETT C.W. (1980). Pairwise multiple comparisons in the homogeneous variance, unequal sample size case. *Journal of the American Statistical Association*, **75**, 789–795.
- DUQUENNE V. (1986). What can lattices do for experimental design? *Mathematical Social Sciences*, **11**, 243–281.
- DUTTON D.G., ARON A.P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology*, **30**, 510–517.
- DUTOIT S., VAN DER LAAN M.J. (2008). *Multiple testing procedures with applications in genomics*. New York: Springer.
- EBBINGHAUS H. (1885). *Über das gedächtnis*. Leipzig: Dunker.

- EDINGTON E.S. (1974). A new tabulation of statistical procedures used in APA journals. *American Psychologist*, **29**, 351–363.
- EDWARDS A.L. (1964). *Expected values of discrete random variables and elementary statistics*. New York: Wiley.
- EDWARDS A.L. (1985). *Experimental design in psychological research*. Cambridge: Harper & Row.
- EDWARDS W. (1965). Tactical note on the relation between scientific and statistical hypothesis. *Psychological Bulletin*, **63**, 400–402.
- EFRON B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, **7**, 1–26.
- EFRON B. (1981). Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika*, **68**, 589–599.
- EFRON B., TIBSHIRANI R.J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- EICH J.E. (1980). The cue-dependent nature of state dependent retrieval. *Memory and Cognition*, **8**, 157–173.
- EISENHART C.M., HASTAY M.W., WALLIS W. (1947). *Techniques of statistical analysis*. New York: McGraw–Hill.
- ESTES W.K. (1991) *Statistical models in behavioral research*. Hillsdale: Erlbaum.
- FAYOL M., ABDI H. (1986). Ponctuation et connecteurs: étude expérimentale de leur fonctionnement dans des paires de propositions. *Actes de la Table ronde “les agencements discursifs et leur système de representation”*.
- FIELD A.P. (1998). A bluffer’s guide to sphericity. *Newsletter of the Mathematical Statistical and Computing Section of British Psychological Society*, **6**, 12–22.
- FIELD A.P. (2005). *Discovering statistics using SPSS*. Thousand Oaks (CA): Sage.
- FISHER R.A. (1935). *The design of experiments*. Edinburgh: Olivier & Boyd.
- FLEISHMAN A.L. (1980). Confidence intervals for correlation ratio. *Educational and Psychological Measurement*, **40**, 659–670.
- FLEISS J.L. (1969). Estimating the magnitude of experimental effect. *Psychological Bulletin*, **72**, 273–276.
- FOSTER K.I., DICKINSON R.G. (1976). More on the language as fixed effect fallacy: Monte Carlo estimates of error rates for F_1 , F_2 , F' and min F . *Journal of Verbal Learning and Verbal Behavior*, **15**, 135–142.
- FOWLER R.L. (1985). Point estimates and confidence intervals in measures of association. *Psychological Bulletin*, **98**, 160–165.
- FRIEDMAN H. (1968). Magnitude of experimental effect and a table for its rapid estimation. *Psychological Bulletin*, **70**, 245–251.
- GAITO J. (1965). Unequal intervals and unequal n in trend analyses. *Psychological Bulletin*, **63**, 125–127.
- GALTON (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland*, **15**, 246–263.
- GALWEY N.W. (2006). *Introduction to mixed modelling*. London: Wiley.
- GAMES P.A. (1971). Multiple comparisons of means. *American Educational Research Journal*, **8**, 531–565.

- GAMES P.A. (1977). An improved t table for simultaneous control on g contrasts. *Journal of the American Statistical Association*, **72**, 531–534.
- GAMES P.A., HOWELL J.F. (1976). Pairwise multiple comparison procedures with unequal n 's and/or variances: a Monte-Carlo study. *Journal of Educational Statistics*, **1**, 113–125.
- GAMES P.A., LUCAS P.A. (1966). Power and the analysis of variance of independent groups on normal and normally transformed data. *Educational and Psychological Measurement*, **16**, 311–327.
- GAMES P.A., KESELMAN H.J., CLINCH J.J. (1979). Tests for homogeneity of variance in factorial design. *Psychological Bulletin*, **86**, 978–984.
- GAMES P.A., KESELMAN H.J., ROGAN J.C. (1983). A review of simultaneous pairwise multiple comparisons. *Statistica Neerlandica*, **37**, 53–58.
- GAYLOR D.W., HOPPER F.N.N. (1969). Estimating the degrees of freedom for linear combinations of mean squares by Satterthwaite's formula. *Technometrics*, **11**, 691–706.
- GLASS G.V., HAKSTIAN R. (1969). Measures of association in comparative experiments: their development and interpretation. *American Educational Research Journal*, **6**, 403–424.
- GLASS G.V., STANLEY J.C. (1970). *Statistical methods in education and psychology*. Englewood Cliffs (NJ): Prentice-Hall.
- GODDEN A., BADDELEY A. (1980). When does context influence recognition memory? *British Journal of Psychology*, **71**, 99–104.
- GRANT D.A. (1962). Testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychological Review*, **69**, 54–61.
- GREENHOUSE S.W., GEISSER S. (1959). On methods in the analysis of profile data. *Psychometrika*, **24**, 95–112.
- GRIEVE A.P. (1984). Tests of sphericity of normal distributions and the analysis of repeated measures designs. *Psychometrika*, **49**, 257–267.
- HALE G.H. (1977). The use of ANOVA in developmental research. *Child development*, **48**, 1101–1106.
- HARTER H.L. (1957). Error rates and sample sizes for range tests in multiple comparisons. *Biometrics*, **13**, 511–536.
- HARTER H.L. (1970). Multiple comparison procedures for interactions. *American Statistician*, **24**, 30–32.
- HAYS W. (1963). *Statistics for psychologists*. New York: Holt.
- HAYS W.L. (1981). *Statistics*. New York: Holt.
- HEDGES L.V. (1980). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, **92**, 490–499.
- HERZBERG P. A. (1969) The parameters of cross-validation. *Psychometrika Monograph Supplement*, **34**.
- HERTZOG C., ROVINE M. (1985). Repeated-measure analysis of variance in developmental research. *Child Development*, **56**, 787–800.
- HOCHBERG Y., TAMHANE A. C. (1987). *Multiple comparison procedures*. New York: Wiley.
- HOEL P. (1971). *Introduction to mathematical statistics*. New York: Wiley.

- HONECK R.P., KIBLER C.T., SUGAR J. (1983). *Experimental design and analysis*. Lanham: U.P.A.
- HSU T., FELDT L.S. (1969). The effect of limitations on the number of criterion score values on the significance level of the F test. *American Education Research Journal*, **6**, 515–527.
- HUDSON J.D., KRUTCHHOFF R.C. (1968). A Monte-Carlo investigation of the size and power of tests employing Satterthwaite's formula. *Biometrika*, **55**, 431–433.
- HULME, C., THOMSON, N., MUIR, C., LAWRENCE, A. (1984). Speech rate and the development of short term memory span. *Journal of Experimental Psychology*, **38**, 241–253.
- HUNTER I.M.L. (1964). *Memory*. London: Penguin.
- HUYNH H., FELDT L.S. (1970). Conditions under which mean square ratios in repeated measurement designs have exact F -distributions. *Journal of the American Statistical Association*, **65**, 1582–1589.
- HUYNH H., FELDT L.S. (1976) Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics*, **1**, 69–82.
- HUYNH H., MANDEVILLE G. (1979). Validity conditions in repeated measures designs. *Psychological Bulletin*, **86**, 964–973.
- JACCARD J., BECKER M.A., WOOD G. (1984). Pairwise multiple comparison procedures: A review. *Psychological Bulletin*, **94**, 589–596.
- JEYARATMAN S. (1982). A sufficient condition on the covariance matrix for F tests in linear models to be valid. *Biometrika*, **69**, 679–680.
- JOLLIFE L.T. (2003). *Principal component analysis*. New York: Springer-Verlag.
- KAISER L.D., BOWDEN D.C. (1983). Simultaneous confidence intervals for all linear contrasts of means with heterogeneous variances. *Communication in Statistical Theory and Method*, **12**, 73–88.
- KELLEY T.L. (1935). An unbiased correlation ratio measure. *Proceedings of the National Academy of Sciences USA*, **21**, 554–559.
- KEPPEL G. (1976). Words as random variables. *Journal of Verbal Learning and Verbal Behavior*, **15**, 263–265.
- KEPPEL G. (1982). *Design and analysis: a researcher's handbook*. Englewood Cliffs (NJ): Prentice-Hall.
- KEPPEL G., SAUFLEY W.H. (1980). *Introduction to design and analysis: a student's handbook*. San Francisco: Freeman.
- KEPPEL G., WICKENS T.D. (2004). *Design and analysis: a researcher's handbook*. Englewood Cliffs (NJ): Prentice Hall.
- KEREN G. (1982). *Statistical and methodological issues in psychology and social sciences research*. Hillsdale (NJ): Erlbaum.
- KEREN G., LEWIS C. (1969). Partial omega squared for anova design. *Educational and Psychological Measurement*, **39**, 119–117.
- KERLINGER F.N. (1973). *Foundations of behavioral research*. New York: Holt.
- KESELMAN H.J. (1982). Multiple comparisons for repeated measures means. *Multivariate Behavioral Research*, **17**, 87–92.

- KESELMAN H.J., ROGAN J.C., MENDOZA J.L., BREEN L.J. (1980). Testing the validity condition of repeated measures F tests. *Psychological Bulletin*, **87**, 479–481.
- KESELMAN H.J., ALGINA J., KOWALCHUK, R.K. (2001). The analysis of repeated measures designs: a review. *British Journal of Mathematical & Statistical Psychology*, **54**, 1–20.
- KEULS M. (1952). The use of studentized range in connection with an analysis of variance. *Euphysia*, **1**, 112–122.
- KIRK R.R. (1982). *Experimental design*. Belmont (CA): Brook-Cole.
- KIRK R.R. (1995). *Experimental design: procedures for the social sciences*. Belmont (CA): Brook-Cole.
- KRANTZ D.H., LUCE R.D., SUPPES P., TVERSKY A. (1971). *Foundations of measurement*. New York: Academic Press.
- LACHMAN R., LACHMAN J.L., BUTTERFIELD E.C. (1979). *Cognitive psychology and information processing: an introduction*. Hillsdale: Erlbaum.
- LANDAUER T.K., FREEDMAN J.L. (1968). Information retrieval from long term memory: category size and recognition time. *Journal of Verbal Learning and Verbal Behavior*, **7**, 291–295.
- LASAGNA L., VON FESLINGER J.M. (1954). The volunteer subject in research. *Science*, **120**, 359–461.
- LEACH C. (1979). *A nonparametric approach for the social sciences*. New York: Wiley.
- LECOUTRE B. (1984). *L'analyse bayésienne des comparaisons*. Lille: P.U.L.
- LEE W. (1966). Experimental design symbolization and model derivation. *Psychometrika*, **31**, 397–412.
- LEE W. (1975). *Experimental design and analysis*. San Francisco: Freeman.
- LEVENE H. (1950). Robust test for the equality of variances. In Olkin I. (Ed.): *Contributions to probability and statistics*. Palo Alto: Stanford University Press.
- LEVIN J.R., MARASCUILO L.A. (1972). Type IV errors. *Psychological Bulletin*, **78**, 368–374.
- LILLIEFORS H.W. (1967). On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, **64**, 387–389.
- LINDMAN H.R. (1974). *Analysis of variance in complex experimental designs*. San Francisco: Freeman.
- LINDSEY G., ARONSON E. (Eds) (1968). *The handbook of social psychology*. Reading: Addison-Wesley.
- LINQUIST E.F. (1953). *Design and analysis of experiments in psychology and education*. Boston: Houghton-Mifflin.
- LOEHLIN J.C. (1987). *Latent variable models*. Hillsdale (NJ): Erlbaum.
- LOEHLIN J.C. (2003). *Latent variable models* (4th edition). Hillsdale (NJ): Erlbaum.
- LOFTUS E.F., PALMER J.C. (1974). Reconstruction of automobile destruction: an example of the interaction between language and memory. *Journal of Experimental Psychology: Human Learning and Memory*, **4**, 19–31.
- LUNNEY G.H. (1970). Using analysis of variance with a dichotomous dependent variable: an empirical study. *Journal of Educational Measurement*, **7**, 263–269.

- MARASCUILLO L.A., LEVIN J.R. (1970). Appropriate post-hoc comparison for interaction and nested hypotheses in analysis of variance designs. *American Educational Research Journal*, 7, 397–421.
- MARTIN C.G., GAMES P.A. (1977). ANOVA tests for homogeneity of variance: nonnormality and unequal samples. *Journal of Educational Statistics*, 2, 187–206.
- MAXWELL S.E. (1980). Pairwise multiple comparisons in repeated measures designs. *Journal of Educational Statistics*, 5, 269–287.
- MAXWELL S.E., CAMP C.J., ARVEY R.D. (1981). Measure of strength of association: a comparative examination. *Journal of Applied Psychology*, 66, 525–534.
- MAXWELL S.E., DELANEY, H.D. (1980). *Designing experiments and analyzing data: a model comparison perspective* (second edition). Hillsdale: Erlbaum.
- MENARD W. (2001). *Applied logistic regression analysis*. Thousand Oaks (CA): Sage.
- MILLER R.G. (1966). *Simultaneous statistical inferences*. New York: McGraw-Hill.
- MILLER R.G. (1968). Jackknifing variances. *Annals of Mathematical Statistics*, 39, 567–582.
- MILLER R.G. (1981). *Simultaneous statistical inferences*. Berlin: Springer Verlag.
- MILLER G.A., NICELY P.E. (1965). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 27, 338–352.
- MILLMAN J., GLASS G.V. (1967). Rules of thumb for writing the ANOVA table. *Journal of Educational Measurement*, 4, 41–51.
- MOOD A., GRAYBILL F.A., BOSS D.C. (1974). *Introduction to the theory of statistics*. New York: McGraw-Hill.
- MOLIN P., ABDI H. (1998). New tables and numerical approximation for the Kolmogorov-Smirnov/Lilliefors/Van Soest test of normality. *Technical report, University of Bourgogne*. Available from www.utd.edu/~herve/MA/Lilliefors98.pdf
- MOSTELLER F., TUKEY J.W. (1977). *Data analysis and regression*. Reading: Addison-Wesley.
- MYERS J.L. (1979). *Fundamentals of experimental design*. Boston: Allyn & Bacon.
- McCALL R.B., APPELBAUM M.I. (1973). Bias in the analysis of repeated measures designs: some alternative approaches. *Child Development*, 44, 401–415.
- McCLOSKEY D.N., ZILIAK, S.T. (2008). *The cult of statistical significance: how the standard error costs us jobs, justice, and lives*. Minneapolis: University of Michigan Press.
- McCLOSKEY M. (1980). The stimulus familiarity problem in semantic memory research. *Journal of Verbal Learning and Verbal Behavior*, 19, 485–502.
- McGEOCH (1942). *The psychology of human learning*. New York: Longmans.
- McNEMAR Q. (1949). *Psychological statistics*. New York: Wiley.
- NEISSER U. (1982). *Memory observed*. San Francisco: Freeman.
- NEWMAN D. (1939). The distribution of the range in samples from a normal distribution, expressed in terms of an independent estimate of standard deviation. *Biometrika*, 31, 20–30.
- NUNALLY J.C. (1978) *Psychometric theory*. New York: McGraw-Hill.
- O'BRIEN P.C. (1983). The appropriateness of analysis of variance and multiple comparison procedures. *Biometrika*, 39, 787–788.

- O'BRIEN R.G. (1979). A general ANOVA method for robust tests of additive models for variance. *Journal of the American Statistical Association*, **74**, 877–880.
- O'BRIEN R.G. (1981). A simple test for variance effect in experimental designs. *Psychological Bulletin*, **89**, 570–574.
- O'GRADY R.E. (1982). Measure of explained variance: caution and limitations. *Psychological Bulletin*, **92**, 766–777.
- PAIVIO A. (1971). *Imagery and verbal processes*. New York: Holt.
- PAMPEL F.C. (2000). *Logistic regression: a primer*. Thousand Oaks (CA): Sage.
- PARSONS H.M. (1974). What happened in Hawthorne? *Science*, **183**, 922–932.
- PARSONS H.M. (1978). What caused the Hawthorne effect? A scientific detective story. *Administration and Society*, **10**, 259–283.
- PEARSON K. (1896). Mathematical contributions to the mathematical theory of evolution. III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London*, **187**, 253–318.
- PEDHAZUR E.J. (1982). *Multiple regression in behavioral research*. New York: Holt.
- PEDHAZUR E.J., PEDHAZUR-SCHMELKIN L. (1991). *Measurement, design, and analysis: an integrated approach*. Hillsdale (NJ): Erlbaum.
- PITMAN E.J.G. (1937). Significance tests which may be applied to samples from any population: (parts I and II). *Royal Statistical Society Supplement*, **3–4**, 119–130 & 225–232.
- PITMAN E.J.G. (1938). Significance tests which may be applied to samples from any population. Part III: The analysis of variance. *Biometrika*, **29**, 322–335.
- PLUTCHIK R. (1983). *Foundations of experimental research*. Cambridge: Harper & Row.
- RENY A. (1966). *Calcul des probabilités*. Paris: Dunod.
- ROBERT F.S. (1979). *Measurement theory with applications to decision making, utility and the social sciences*. Reading: Addison-Wesley.
- ROGAN J.C., KESELMAN H.J., MENDOZA J.L. (1979). Analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology*, **32**, 269–286.
- ROSENTHAL R. (1976). *Experimenter effects in behavioral science*. New York: Irvington.
- ROSENTHAL R. (1977). The pygmalion effect lives. In Schell R.E. (Ed.): *Readings in developmental psychology today*. New York: Random House.
- ROSENTHAL R. (1978). How often are our numbers wrong? *American Psychologist*, **33**, 1005–1008.
- ROSENTHAL R., GAITO J. (1963). The interpretation of levels of significance by psychological researchers. *Journal of Psychology*, **55**, 31–38.
- ROSENTHAL R., ROSNOW R.L. (1975). *The volunteer subject*. New York: Wiley.
- ROSENTHAL R., ROSNOW R.L. (1985). *Contrast analysis: focused comparisons in the analysis of variance*. Boston: Cambridge University Press.
- ROSENTHAL R., ROSNOW R.L. (2000). *Contrasts and effect sizes in behavioral research: a correlational approach*. Boston: Cambridge University Press.
- ROSNOW R.L., SULS J. (1970). Reactive effects of pretesting in attitude research. *Journal of Personality and Social Psychology*, **15**, 338–343.
- ROUANET H., LEPINE D. (1970). Comparison between treatments in a repeated-measures design: ANOVA and multivariate methods. *British Journal of Psychology*, **8**, 166–84.

- ROUANET H., LEPINE D. (1977). L'analyse des comparaisons pour le traitement des données expérimentales. *Informatique et Sciences Humaines*, 33–4, 10–125.
- ROUANET H., LEPINE D., HOLENDER D. (1978). Model acceptability and the use of Bayes–Fiducial methods for validating models. In Requin J. (Ed.): *Attention and Performance VII*. Hillsdale (NJ): Erlbaum.
- RUBIN D. (1985). Issues of regularity and control: confessions of a regularity freak. Paper presented at the *Third George A. Talland memorial conference on memory and aging*. New Seabury, Massachusetts.
- RUBIN G.S., BECKER C.A., FREEMAN R.H. (1979). Morphological structure and its effect on visual word recognition. *Journal of Verbal Learning and Verbal Behavior*, 18, 757–767.
- RYAN T.A. (1959). Multiple comparisons in psychological research. *Psychological Bulletin*, 56, 26–47.
- SANTA J.L., MILLER J.J., SHAW M.L. (1979). Using Quasi-*F* to prevent alpha inflation due to stimulus variation. *Psychological Bulletin*, 86, 37–46.
- SATTERTHWAITE F.E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2, 110–114.
- SCHEFFÉ H.A. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika*, 40, 87–104.
- SCHEFFÉ H. (1959). *The analysis of variance*. New York: Wiley.
- SCHULTZ E.F. (1955). Rules of thumb for determining expectations of mean squares. *Biometrics*, 11, 123–135.
- SEARLMAN A., HERRMAN (1994). *Memory from a broader perspective*. New York: McGraw-Hill.
- SEMON (1904). *Die mneme als erhaltendes Prinzip im Wechsel des organischen Geschehens*. Leipzig: Engelmann.
- SHEPARD R.N., METZLER J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701–3.
- SHOBEN E. (1982). Semantic and lexical decision. In Puff C.R. (Ed.) *Handbook of research methods in human memory and cognition*. New York: Academic Press.
- ŠIDÁK Z. (1967). Rectangular confidence region for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62, 626–633.
- SIEGEL S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.
- SLAMECKA N.J. (1960). Retroactive inhibition of connected discourse as a function of practice level. *Journal of Experimental Psychology*, 59, 104–108.
- SMART R.G. (1966). Subject selection bias in psychological research. *Canadian Psychology*, 7, 115–121.
- SMITH E.E., SHOBEN E.J., RIPS L.J. (1974). Structure and process in semantic memory: a featural model for semantic decisions. *Psychological Review*, 81, 214–241.
- SMITH P.L. (1982). Measures of variance accounted for: theory and practice. In Keren G. (Ed.): *Statistical and methodological issues in psychology and social sciences research*. Hillsdale (NJ): Erlbaum.
- SMITH S.M. (1979). Remembering in and out of context. *Journal of Experimental Psychology*, 6, 342–353.

- SNEDECOR G.W., COCHRAN W.C. (1971). *Méthodes statistiques*. Paris A.C.T.A.
- SOKAL R.R., ROHLF F.J. (1969). *Biometry*. San Francisco: Freeman.
- SOLSO R.L., MCCARTHY J.E. (1981). Prototype formation of faces: a case of pseudo-memory. *British Journal of Psychology*, 72, 499–503.
- STEIN B.S., BRANSFORD J.D. (1979). Quality of self-generated elaboration and subsequent retention. In J.D. Bransford (Ed.): *Human cognition*, p. 78 ff. Belmont (CA): Wadsworth.
- STERNBERG S. (1969). Memory scanning: mental processes revealed by reaction time experiments. *American Scientist*, 57, 421–457.
- STEVENS J.P. (1996). *Applied multivariate statistics for the social sciences*. New York: Psychology Press.
- STUDENT (1927). Errors of routine analysis. *Biometrika*, 19, 151–164.
- TABACHNICK B.G., FIDELL L.S. (2007). *Experimental designs using ANOVA*. Belmont (CA): Thomson.
- TAFT M., FOSTER K.I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14, 638–647.
- TAFT M., FOSTER K.I. (1976). Lexical storage and retrieval of polymorphic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior*, 15, 607–620.
- TAMHANE A.C. (1977) Multiple comparisons in model I one-way ANOVA with unequal variances. *Communications in Statistics*, A6, 15–32.
- TAMHANE A.C. (1979). A comparison of procedures for multiple comparisons of means with unequal variances. *Journal of the American Statistical Association*, 74, 471–480.
- TANG P.C. (1938). The power function of the analysis of variance test. *Statistical Research Memoirs*, 2, 126–149.
- TAYLOR J.A. (1953). A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology*, 48, 285–290.
- THOMPSON G.L. (1990). Efficiencies of interblock rank statistics for repeated measures designs. *Journal of the American Statistical Association*, 85, 519–528.
- THOMPSON L., AMMANN L. (1990) Efficiency of interblock rank statistics for repeated measures designs. *Journal of the American Statistical Association*, 85, 519–528.
- TONG Y.L. (1980). *Probability of inequalities in multivariate distributions*. New York: Academic Press.
- TREISMAN A. (1986). Features and objects in visual processing. *Scientific American*, 255, 114–125.
- TULVING E. (1983). *Elements of episodic memory*. Oxford: OUP
- TULVING E., PEARLSTONE Z. (1965). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 5, 381–391.
- TUKEY J.W. (1953). *The problem of multiple comparisons*. Princeton University (unpublished manuscript).
- UNDERWOOD B.J. (1975). Individual differences as a crucible in the construction. *American Psychologist*, 30, 128–34.
- UNDERWOOD B.J. (1983). *Studies in learning and memory*. New York: Praeger.
- UNDERWOOD B.J., SHAUGHNESSY J.J. (1983). *Experimentation in psychology*. Malabar: Krieger.

- URY H.K. (1976). A comparison of four procedures for multi comparisons among means (pairwise contrasts). *Technometrics*, **18**, 89–97.
- VAN SOEST J. (1967). Some experimental results concerning tests of normality. *Statistica Neerlandica*, **21**, 91–97.
- VOKEY J.R., ALLEN, S.W. (2005). *Thinking with data* (4th edition). Lethbridge (Alberta): Psyence Ink.
- WEISS J.M. (1972). Psychological factors in stress and disease. *Scientific American*, **226**, 104–113.
- WIKE E.L., CHURCH J.D. (1976). Comments on Clark's the language a fixed-effect fallacy. *Journal of Verbal Learning and Verbal Behavior*, 249–255.
- WILCOX R.R. (1987). New designs in analysis of variance. *Annual Review of Psychology*, **38**, 29–60.
- WILSON W. (1962). A note on the inconsistency inherent in the necessity to perform multiple comparisons. *Psychological Bulletin*, **59**, 296–303.
- WILSON W., MILLER H.A. (1964). A note on the inconclusiveness of accepting the null hypothesis. *Psychological Review*, **71**, 238–242.
- WILSON W., MILLER H.A., LOWER J.S. (1967). Much ado about the null hypothesis. *Psychological Review*, **67**, 188–196.
- WINER B.J. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.
- WINER B.J., BROWN D.R., MICHELS K.M. (1991). *Statistical principles in experimental design*. New York: McGraw-Hill.
- WISEMAN R. (2007). *Quirkology: how we discover the big truths in small things*. New York: Basic Books.
- YUILLE J.C., MARSCHARK M. (1983). Imagery effects on memory: theoretical interpretations. In A.A. Sheik (Ed.): *Imagery*. New York: Wiley.
- ZEICHMEISTER E.B., NYBERG S.E. (1982). *Human memory*. Monterey: Brooks-Cole.
- ZILIACK S.T., McCLOSKEY D.N. (2008). *The cult of statistical significance*. Michigan: University of Michigan Press.
- ZWICK R., MARASCUILLO L.A. (1984). Selection of pairwise multiple comparison procedures for parametric and nonparametric analysis of variance models. *Psychological Bulletin*, **95**, 148–155.

Index

A

α , 42, 43, 47, 149, 228,
474–476, 478
 α_a , 193
 $\alpha[PC]$, 230, 233, 240, 255,
274
 $\alpha[PF]$, 230, 233, 240, 254
a posteriori comparison,
225, 268
a priori, 165
a priori comparison, 225
abbreviated symbols, 168
Abdi, 287
active indices, 169
addition rule, 438, 439,
459
additivity
main effect, 296
sum of squares, 142
Adorno, 15, 345
Algina, 349
Allen, 47
alpha, *see* α
alpha level, 150, 476
alpha per comparison, *see*
 $\alpha[PC]$
alpha per family, *see* $\alpha[PF]$
alternative hypothesis, 40,
42–44, 148, 150, 472,
474, 475
contrast, 226
analysis of variance, 130, 262
 $S(A \times B)$, 290
 $S(A)$, 130, 147, 191
 $S(A) \times B$, 373
 $S \times A$, 335
 $S \times A \times B$, 358
 $S \times A(B)$ design, 387
goal, 130
intuitive approach, 130
regression approach,
172
repeated measures
design, 335
table, 144, 153
AND function, 430
ANOVA, *see* analysis of
variance
APA style, 154, 184, 243, 381,
393
Appelbaum, 349
Armor, 36
Aron, 10
Aronson, 345

assumption
circularity, 346, 348, 349
sphericity, 346, 348, 349
validity, 191, 199, 211, 371
astute coding, 130
Atkinson, 4

B

β , 43, 270, 478
Baddeley, 4, 11, 13, 112,
358, 370
balanced design, 134,
292, 298
bar graph, 411
Bartram, 4
barycenter, 20
basic ratios, 168
basic unit of measurement, 16
Bayes' theorem, 434
behavior, 470
Bennet, 404
Bernoulli, 456
best-fit line, 67
best-fit plane, 92
between-group
degrees of freedom,
141, 184
deviation, 136, 192, 298
distance, 136, 192
mean square, 140, 184
sum of squares, 138,
184, 195
variability, 132, 152
variation, 294
between-subjects
variability, 342
Bevan, 212
biased sample, 388
bimodal, 417
binary, 453, 471
binomial
coefficient, 456, 458,
460, 464
distribution, 443, 453, 455,
460, 462, 466, 480, 481
expected value, 455
variance, 456
test, 470, 480
Blake, 4
Bonferroni
correction, 253
inequality, 233, 254, 268
unequal allocation, 255
Boole inequality, 233, 254, 268

Bootstrap, 59
Box, 212, 349
Box's index of sphericity, 349
Brady, 7
Bransford, 13, 157, 158, 205
Breen, 349
Brown, 480, 482
Brunswick, 11, 207

C

Campbell, 9, 11, 345
canonical decomposition, 324
Captain's age, 231
carry-over effect, 345, 358
Cartesian product, 292
center of gravity, 20
centered graph, 20
central tendency, 415
mean, 416
measures of, 415
median, 416
mode, 416, 417
chance, 148
Chanquoy, 287
Chastaing, 207, 471, 481
Chateau, 37
Cherulnik, 9
chess, 4
Church, 207
circularity assumption, 346,
348, 349
Clark, 207
classificatory variable, 8,
287, 375
code, 176
coding
contrast, 188
dummy, 188
group, 188
coefficient of
amplification, 165
coefficient of correlation, *see*
correlation
coefficient of
determination, 74
coefficient of intensity,
384
coin, 427
Coleman, 36
Collins, 7
combinations, 457
comparison, 224
a posteriori, 225, 268
a posteriori factorial, 333

- comparison (*cont'd*)
 - a priori*, 225
 - independent, 225
 - multiple, 233
 - non-orthogonal, 225, 253
 - orthogonal, 225, 253
 - pairwise, 268, 273
 - planned, 224, 225, 245
 - planned non-orthogonal, 253
 - planned orthogonal, 224, 240
 - post hoc*, 224, 225, 268
 - post hoc* factorial, 333
- complementary event, 428
- complex factor, 399
- composed terms, 398
- comprehension formula, 167
 - development of, 402
 - from means and standard deviation, 205
- computational formula, 167, 352
 - development of, 402
- computational symbols, 168
- computer simulation, 212
- condensed notation, 134
- conditional probability
 - definition, 433
- confound, 5, 192, 201, 335, 345, 387
- Conover, 211
- Conrad, 11, 375
- conservative, 270
- contrast, 188, 224, 226, 324
 - alternative hypothesis, 226
 - classical approach, 253
 - coding, 188
 - coefficients, 228
 - definition of, 225
 - independence of, 236
 - interaction, 327
 - main effect, 326
 - mean square, 237
 - multiple orthogonal, 262
 - multiple regression
 - approach, 253, 258
 - non-orthogonal, 253
 - null hypothesis, 226
 - orthogonal, 253
 - maximum number of, 225
 - planned, 245
 - regression approach, 238
- control factor, 287, 288, 291, 310
- control group, 131, 157, 257
- Cook, 9, 345
- corner test, 22
- Cornfield, 404
- correction for continuity, 462, 464
- correlation, 162, 172, 179, 225
 - causality, 36
 - coefficient, 26, 70, 344
 - computation, 18
 - formula, 26
 - multiple coefficient, 96, 261
 - mutual, 36
 - negative linear relationship, 28
 - no linear relationship, 28
 - overview, 16
 - partial coefficient, 318, 384
 - Pearson's coefficient, 16, 39, 236
 - perfect, 28
 - positive linear relationship, 28
 - proportion of common variance, 37
 - rationale, 18
 - scatterplot, 18
 - semi-partial, 112, 121
 - semi-partial coefficient, 121, 125, 127, 254, 258, 262, 265
 - alternative formula, 123
 - similarity of distributions, 30
 - spurious, 36
 - square coefficient, 117
 - square multiple coefficient, 118
 - square semi-partial coefficient, 121
 - squared coefficient, 37, 74, 163, 179
 - unique contribution, 121
 - Z-scores, 32, 34
- covariance, 25, 26
 - definition of, 488
- Cowles, 15
- Cozby, 345
- criterion, 39
- critical region, 476
- critical value, 50, 52, 149, 476
- cross-product, 23, 24, 139, 236, 237, 489
- crossed, 374, 396, 397, 399
 - factor, 285, 292
 - relationship, 284
- Crowder, 349
- Crump, 398
- cup, 428
- de Moivre, 460
- decision making, 476
- decision rule, 151, 475
- degrees of freedom, 39, 75, 104, 137, 140, 141
 - $S \times A(B)$, 390
 - $S(A)$, 141
 - $S(A \times B)$, 300
 - $S \times A$, 340
 - $S(A) \times B$, 377
 - $S \times A \times B$, 359
 - between-group, 141, 184
 - developed, 402
 - formula, 401
 - interaction, 300
 - main effect, 300
 - regression, 184
 - residual, 184
 - total, 142
 - within-group, 141, 184, 300
- DeGroot, 4
- delta, 64
- dependent variable, 2, 12, 13, 479
 - good principles of, 13
- descriptive statistics, 162, 411
- design
 - $S(A \times B)$, 290
 - $S(A)$, 130, 147, 152, 172, 191
 - $S(A) \times B$, 373
 - $S \times A$, 335
 - $S \times A \times B$, 358
 - $S \times A(B)$, 387
 - balanced, 134, 292, 298
 - basic experimental, 286
 - factorial, 152, 185, 191, 291
 - contrast, 324
 - latin squares, 396
 - mixed, 373
 - multi-factorial, 282, 291
 - nested factorial, 387
 - notation, 283
 - one factor, 152
 - one-factor
 - between-subjects, 287
 - one-factor independent measures, 287
 - one-factor within-subjects, 287
 - randomized block, 310
 - repeated measures, 335, 373
 - split-plot, 287, 373
 - two-factor
 - notation, 293
 - two-factors
 - between-subjects, 287
 - two-factors independent measures, 287

D

- data snooping, 225
- Davenport, 371
- Davis, 15

- two-factors repeated measures, 287
- two-factors within-subjects, 287
- unbalanced, 185
- within-subjects, 335
- writing down, 285
- deviation
 - between-group, 136, 191, 192, 298
 - from the mean, 20
 - within-group, 136, 191, 192
- dichotomy, 453
- Dickinson, 371
- dissociation, 157
 - double, 13
- distance, 19
 - between-group, 136, 192
 - from the mean, 20
 - within-group, 136, 192
- dolls, 471
- dot notation, 134
- double notation, 180
- double-blind, 7, 232, 373
- double-centering, 350
- dummy coding, 188
- Duncan test, 268, 279
- Dunn inequality, 233, 254, 268
- Dunnett test, 254, 257
- Dutton, 10
- E**
 - $\hat{\varepsilon}$, 351
 - $\tilde{\varepsilon}$, 351
 - ε , 350, 351
 - η^2 , 162
 - Ebbinghaus, 3
 - ecological validity, 11, 282, 288
 - Edwards, 196
 - effect
 - carry-over, 358
 - estimation of intensity, 162
 - intensity of, 161, 162
 - number of subjects, 166
 - perverse, 348
 - Efron, 59
 - Eisenhart, 197
 - elementary event, 427
 - empirical approach, 45
 - empty set, 429
 - equiprobability, 431
 - error
 - experimental, 43
 - measurement, 43, 192
 - random, 346
 - error term, 154, 324, 330
 - Estes, 226
 - estimation, 227
 - estimator
 - unbiased, 195
 - Euler, 461
 - event, 427
 - complement, 431
 - complementary, 229, 428
 - compound, 427, 431
 - disjoint, 429
 - elementary, 427, 431, 443
 - exclusive, 429, 431, 439
 - exhaustive, 429, 431
 - impossible, 431
 - inclusive, 439
 - independent, 229, 430, 436
 - joint, 430
 - mutually exclusive, 429
 - mutually independent, 453
 - non-exclusive, 440
 - probability (of an), 431
 - exact statistical hypothesis, 41
 - examples
 - $S \times A$ design, 340
 - $S \times A$ design
 - computational formulas, 352
 - age, speech rate and memory span
 - non-orthogonal multiple regression, 112
 - bat and hat, 375
 - comprehension formulas, 378
 - computational formula, 380
 - context and memory
 - contrasts, 234
 - contrast
 - rank order, 228
 - cute cued recall, 292, 303
 - computation, 305
 - computational formulas, 320
 - non-orthogonal factorial
 - contrasts, 332
 - orthogonal factorial
 - contrasts, 330
 - faces in space, 388
 - computation, 392
 - Kiwi and Koowoo
 - binomial test, 471
 - mental imagery, 130, 173
 - computation, 142, 173
 - score model, 194
 - mental imagery
 - regression approach, 176
 - plungin', 358
 - comprehension formulas, 363
 - computation, 363
 - computational formulas, 365
 - proactive interference, 353
 - proactive interference
 - computation, 354
 - reaction time and memory
 - Set, 65
 - computation, 68
 - retroactive interference
 - computation, 93, 105
 - orthogonal multiple
 - regression, 90
 - Romeo and Juliet, 157
 - Bonferonni test, 256
 - computation, 158
 - computational formulas, 168
 - Dunnett test, 258
 - Newman-Keuls test, 279
 - non-orthogonal
 - contrasts, 263
 - orthogonal contrasts, 258
 - regression approach, 182
 - Scheffé's test, 272
 - Tukey test, 275
 - taking off ...
 - Newman-Keuls test, 277
 - video game performance
 - contrast, 247
 - word length and number of lines
 - computation, 25
 - correlation, 16
 - expected value, 192, 195, 302, 330, 443, 445, 447, 485
 - definition of, 447, 486
 - derivation of, 485
 - mean square, 196, 302, 494
 - rules for, 404
 - sum of squares, 196, 494
 - experiment, 427
 - experimental
 - effect, 191, 343
 - error, 43, 132
 - factor, 2
 - group, 131, 157
 - treatment, 1
 - validity, 345
 - experimental design, 1
 - basic, 286
 - notation, 283
 - writing down, 285
 - external validity, 11, 282
 - F**
 - F_{range} , 273, 274, 276
 - F ratio, 39, 47, 71, 75, 96, 104, 133, 142, 182, 316, 337, 340, 344, 360, 396, 485

- F ratio**
 $S \times A(B)$, 391
 $S(A \times B)$, 300
 $S \times A$, 340
 $S(A) \times B$, 377
 $S \times A \times B$, 360
 $S(A)$, 142
 interaction, 300
 main effect, 300
 omnibus, 244–246, 270
 smaller than 1, 199
- F test**, 39, 47
 correlation, 100
 omnibus, 261
 semi-partial correlation, 124
- F''** , 370
- F'** , 369, 370
- factor**
 control, 287, 288, 291
 crossed, 285, 292, 374, 397
 elementary, 398
 experimental, 2
 fixed, 192, 194, 197, 207, 301, 400
 nested, 284, 373
 of interest, 287, 288
 random, 192, 194, 197, 207, 400
- factor analysis**, 36
- factorial design**, 185, 191
 contrast, 324
 partially repeated measures, 373
 repeated measures, 358
- factorial number**, 456, 457
- false alarm**, 268, 477
- family of comparisons**, 228
- family of contrasts**, 228
- Feldt**, 349, 351
- Fermat**, 426
- Fidell**, 396
- Fisher**, 47, 53, 149
- Fisher distribution**, 196
- Fisher's F** , 133, 142
- fixed factor**, 192, 194, 196, 197, 207, 301, 330, 387, 400
- Foster**, 283, 371
- Franklin**, 404
- Freedman**, 7
- frequency polygon**, 411
- frequency table**, 411
- Friedman and Kruskal–Wallis tests**, 211
- frivolous**, 446
- frontier value**, 52
- functional relationship**, 66
- Games**, 212, 254, 273
- games**, 426
- Gauss**, 456, 460
- Gaussian approximation**, 482
- Gaussian distribution**, 460
- Gaylor**, 371
- Geisser**, 349, 351
- generalization**, 288, 292
- Glass**, 398
- global effect**, 161, 224
- global indices**, 311
- Godden**, 358, 370
- good old rule**, 22
- Gosset**, 273
- grand mean**, 131, 134, 343
- Greenhouse**, 349, 351
- Greenhouse–Geisser extreme correction**, 352
- group coding**, 188
- group mean**, 131, 134
- H**
 H_0 , 472
 H_1 , 472
 Hand, 349
 Hays, 196
 Herrman, 13
 Hertzog, 349
 hierarchical analysis, 325
 hierarchical partition, 328
 histogram, 411, 455
 proportion, 412
 history, 345
 Hochberg, 268
 Hoel, 196
 homogeneity, 200, 212
 homogeneity of covariance, 346, 384
 homogeneity of variance, 346, 489
 homoscedasticity, 148, 212, 274, 319, 404
 Honeck, 398
 Hopper, 371
 horizontal axis, 18, 66
 how to get rid of the book, 396
 Howell, 273
 Hudson, 371
 Hulme, 112
 Huynh, 349, 351
 hypothesis
 alternative, 41, 42, 472
 null, 41, 42, 472
 rival, 157
 Hyunh–Feldt sphericity correction, 351
- I**
 imprecise statistical hypothesis, 41, 472
- independence**, 233, 436
 definition of, 488
 independence of error, 489
 independent comparison, 225
 independent event, 229, 436
 independent observations, 141
- independent variable**
 confounded, 5
 manipulated, 4
- indices**
 global, 311
 partial, 311
- individual differences**, 43
- inexact statistical hypothesis**, 41
- inference**, 41, 192
- inferential statistics**, 41, 162
- intensity of the relationship**, 43
- interaction**, 192, 282, 283, 288, 294, 295, 297, 299, 361
 F ratio, 300
 contrast, 327
 degrees of freedom, 300
 first-order, 399
 mean square, 300
 plot, 299
 second-order, 361, 389, 399
 terms, 398, 399
- intercept**, 63, 67, 68, 77, 87, 107, 115, 178, 185, 187
- internal validity**, 11
- intersection**, 429
- interval scale**, 211
- item factor**, 387, 388
- J**
 Jaccard, 273
 Johnson, 157
- K**
 Keppel, 168, 396
 Kerrich, 465
 Keselman, 349
 Kirk, 168, 273, 349, 396
 Kiwi, 471
 Koowoo, 471, 479
 Krantz, 211
 Krutchkoff, 371
- L**
 λ_a , 226
 Lépine, 349
 Lachman, 11
 Landauer, 7
 large N , 480
 Lasagna, 15
 latent variables, 36
 latin squares design, 396
 Lawrence, 112

- Leach, 211
- least squares criterion, 92
- least squares estimation, 177
- Lee, 168, 398, 404
- likelihood, 471
- Lilliefors, 212
- Lindsey, 345
- linear function, 63, 88
- linear relationship
 - negative, 28
 - positive, 28
- Linguist, 212
- Loehlin, 36
- Loftus, 206, 207
- long run, 447, 465
- Lucas, 212
- M**
- main effect, 283, 294, 295, 298, 299, 360, 398
 - F* ratio, 300
 - additivity, 296
 - contrast, 326
 - degrees of freedom, 300
 - mean square, 300
- Mandeville, 349
- Marascuilo, 273
- marginal mean, 295, 298
- Marschark, 7
- Martin, 212
- matched sample, 285
- mathematical expectation, 302, 347, 361, 445, 447
- maturation, 345
- Maul, 11
- McCall, 349
- McCloskey, 2, 7, 47
- McNemar, 14
- mean, 80, 416, 445, 447, 481
 - as predictor, 177
 - distance from, 135
 - grand, 131, 134
 - group, 131, 134
 - marginal, 295, 298
 - model, 185, 188
 - population, 130, 401, 416
 - sample, 416
- mean square, 75, 104, 140, 338, 340, 488
 - between-group, 140, 184
 - contrast, 237
 - expected value, 494
 - interaction, 300, 360
 - main effect, 300
 - mathematical expectation, 347
 - regression, 75, 184
 - residual, 75, 184
 - test, 330, 360
 - within-group, 140, 184
- measure of dispersion, 417
 - range, 417
 - standard deviation, 421
 - sum of squares, 417
 - variance, 419
- measurement, 470
- measurement error, 43, 67, 92
- median, 416
- memory set, 65
- Mendoza, 349
- Miller, 268, 291
- Millman, 398
- miss, 477
- mixed design, 373
- mode, 416, 417
- Monte-Carlo, 212, 230, 465, 466, 479, 480
- Mood, 195, 196
- $MS_{\text{test},A}$, 383
- $MS'_{\text{test},A}$, 369
- Muir, 112
- multi-factorial design, 282, 283
- multicollinearity, 126, 259
- multimodal, 417
- multiple comparisons, 233
- multiple regression, 86, 262
 - 2D representation, 92
 - 3D projection, 89
 - 3D representation, 91
 - approach, 188
 - best-fit plane, 92
 - classical approach
 - non-orthogonal
 - contrasts, 262
 - contrasts, 253
 - equation of a plane, 87
 - hierarchical approach, 118
 - importance of each
 - independent variable, 98, 99
 - intercept, 87, 93
 - least squares criterion, 92
 - multiple variables, 125
 - orthogonal, 86, 112
 - orthogonality, 95
 - part correlation, 121
 - plane, 86, 91
 - quality of prediction, 96, 98, 116, 117
 - quality of prediction of *T*, 98
 - quality of prediction of *X*, 98
 - residual, 92
 - semi-partial coefficient, 125, 127
 - semi-partial contribution, 118
 - semi-partial correlation, 120–122
 - alternative formula, 123
 - slope, 87, 88, 93
 - specific contribution, 119
 - square semi-partial
 - correlation, 121
- multiplication rule, 22
- Myers, 168, 349, 371
- N**
- v_1 , 48, 52, 97, 149
- v_2 , 48, 52, 97, 149, 370
- negative linear
 - relationship, 28
- Neisser, 14
- nested, 152, 284, 373, 387, 396, 397
 - factor, 152, 284, 286, 291, 373
 - relationship, 284
- nested factorial design, 387
- nested variables, 185
- Newman–Keuls test, 268, 273, 276
- Nicely, 291
- noisy, 470
- nominal variable, 130, 172
- non-experimental research, 16
- non-orthogonal, 117, 126
- non-orthogonal comparison, 225, 253
- non-orthogonal
 - contrasts, 258
- non-parametric, 211
- non-significant
 - ns*, 153
- normal approximation, 460, 482
- normal distribution, 460, 462, 480, 481
 - standardized, 460, 461
- normality assumption, 148, 236, 274, 320
- notation
 - score, 414
- notational equivalences, 180
- notational shortcut, 402
- ns* (non-significant), 153
- null hypothesis, 40, 42, 43, 148, 149, 302, 472, 474, 475
 - contrast, 226, 235
- number in the square, 167, 402
 - construction of, 168
- number of combinations, 458
- number of observations, 24, 43
- number of subjects, 165
- Nunally, 126
- O**
- ω , 162
- O'Brien, 212

- observations
 - number of, 43
- observed value, 63
- omnibus test, 161, 224, 271
- one-tailed test, 200, 474, 482
- operationalize, 157, 388
- or function, 430
- ordinal scale, 211
- orthogonal comparison, 225, 253
- orthogonal multiple
 - regression, 250
- orthogonal polynomial, 248
- orthogonal variables, 112
- orthogonality, 86, 233
 - consequences, 104
 - contrast and correlation, 236
- P**
- p*-value, 473
- pairwise comparison, 268, 273
- Paivio, 7
- Palmer, 206
- parameters, 40, 193, 481
- parasitic variable, 310
- Parsons, 7
- partial indices, 311, 317
- partition, 135
- Pascal, 426, 456
- Pearlstone, 292, 325, 330
- Pearson's coefficient of
 - correlation, *see* correlation
- Pedhazur, 36, 126
- Pedhazur-Schmelkin, 36
- permutation, 456
- permutation test, 53
 - approximation, 55
 - exact, 55
- perverse effects, 348
- phonetic symbolism, 481
- Pitman, 53
- planned comparison, 224, 225, 245
 - .01, 476
 - .05, 476
- polygon, 455
- population, 41, 43, 148, 193, 196
- positive linear relationship, 28
- post hoc* comparison, 224, 225, 268
- post-test, 345, 346
- potatoids, 428
- power, 44, 478
- power analysis, 165, 166
- pre-test, 345, 346
- precise statistical hypothesis, 41, 472
- predicted value, 63, 178
- prediction, 178
- principal component
 - analysis, 36
- probability, 148, 426, 430
 - associated, 184, 452, 460, 473
 - axioms, 430
 - conditional, 432, 433, 436
 - distribution, 443–445, 467
 - plotting, 445
 - event, 430
 - in the family, 229
 - level, 161
 - posterior distribution, 433
 - prior distribution, 433
- probability distribution, 443–445
 - definition of, 486
- product rule, 438
- proportion histogram, 412
- ψ , 226, 227
- $\hat{\psi}$, 227, 326
- Pygmalion effect, 7
- Pythagorean theorem, 447
- Q**
- quadrants, 20
- quadratic function, 79
- quantitative variable, 63
- quasi-*F*, 368, 369, 383, 391, 396
 - ν_2 for, 370
- quasi-*F'*, 391
- quasi-mean square, 391
- Quillian, 7
- R**
- random
 - assignment, 5, 11, 131
 - error, 67, 346
 - fluctuation, 132, 472
 - sample, 14, 194, 336, 388
 - variable, 443
- random factor, 185, 192, 194, 197, 207, 212, 330, 387, 400
- random variable, 445
 - definition of, 485
- randomization, 5, 282
 - Joy of, 200
- randomized block design, 310
- range, 417
- rank order, 211, 227
- raw data, 411
- re-test, 346
- rectangle, 23
- Refresher: correlation, 16
- region
 - critical, 476
 - non-rejection, 476
- rejection, 50–52, 149, 151, 476
 - suspension of judgment, 50–52, 149, 476
- regression, 130, 172
 - best-fit line, 67
 - best-fit plane, 92
 - degrees of freedom, 184
 - equation of a line, 63
 - equation of a plane, 87
 - generalities, 63
 - intercept, 63, 67, 68, 87, 115
 - least squares criterion, 92
 - line, 66, 86
 - linear function, 63
 - mean square, 184
 - measurement error, 67
 - model, 102
 - perfect line, 64
 - plane, 86, 91, 113
 - predicted value, 63
 - quality of the prediction, 70
 - rate of change, 63
 - residual, 67, 92
 - residual of prediction, 67
 - scatterplot, 66
 - simple, 63
 - slope, 63, 67, 68, 88, 113
 - straight line, 63, 67
 - sum of squares, 74, 184
 - toward the mean, 346
 - two factors, 312
 - variance, 75
- related sample, 285
- Reny, 196
- repeated measures design, 335
- repetition, 147
- replication, 147, 194, 231, 387
- residual, 67, 92, 119, 339
 - degrees of freedom, 184
 - mean square, 184
 - of prediction, 92, 119
 - sum of squares, 74, 184
 - variance, 73, 75
- retroactive interference, 90
- rival
 - hypotheses, 157, 241
 - theories, 157
- Roberts, 211
- robust, 211
- Rogan, 349
- Rosenthal, 7, 15, 161, 226, 255
- Rosnow, 15, 226, 255
- Rouanet, 349
- Rovine, 349
- Rubin, 283
- rules
 - random and fixed factors, 194

S

 Σ , 414 $S \times A$ design, 335 $S(A \times B)$ design, 290 $S(A)$ design, 152 $S(A) \times B$ design, 373 $S \times A \times B$ design, 358 $S \times A(B)$ design, 387

sample, 41, 148

biased, 388

matched, 285

point, 427

related, 285

representative, 207

space, 427, 430

yoked, 285

sampling

replacement, 59

sampling distribution, 44, 148,

150, 155, 191, 196, 230,

319, 475–477, 480

 Fisher's F , 44, 148

Santa, 371

Satterthwaite, 370, 371

Saugnessy, 12

scalar-product, 236

scatterplot, 18, 22, 30, 66

Scheffé, 212, 273

Scheffé's test, 268, 270, 272

Schultz, 404

Schwartz inequality, 27

score model, 102, 191, 193,

212, 314, 404, 489

 $S(A)$, 191 $S \times A$

fixed factors, 346

random factors, 347

 $S \times A \times B$

fixed factors, 361

 $S(A) \times B$

fixed factors, 381

mixed, 383

random factors, 382

experimental effect, 191

fixed, 489

Greek letters, 400

mixed factors, 310, 368

Model I, 197, 301, 346, 381

 $S \times A$, 343 $S \times A \times B$, 361

Model II, 197, 198, 307, 347,

367, 382

 $S \times A$, 343 $S \times A \times B$, 367

Model III, 310, 368, 383

 $S \times A \times B$, 368

random factor, 306, 307

refresher, 489

repeated measures design,
 343

Roman letters, 400

sum of squares, 195

two-factor design, 301

writing of, 400

score notation, 414

SCP, 24

Searleman, 13

Seattle, 479

semi-partial coefficients of

correlation, 254

semi-partial contribution, 118

semi-partial correlation

alternative formula, 123

sensitization, 345

set theory, 428–431

sex (again?), 285

sex ...

Sorry not in this book, 5

Shaughnessy, 6

Shoben, 6, 283

short term memory, 65

shower, 479

Šidák

correction, 234, 253, 255,

265

critical values, 234

equation, 233, 254

inequality, 254, 268

Siegel, 211

 σ_a^2 , 197

signal to noise ratio, 165

significance level, 42, 43, 228

simple effect, 295

simulation, 465

single-blind, 7

Slamecka, 90, 96

slope, 63, 67, 68, 77, 87, 88,

107, 113, 178, 185, 187

Smart, 14

Smith, 7, 234

Snedecor, 149

sources of variation, 398

sphericity, 349

index, 349

sphericity assumption, 346,

348, 349

split-plot design, 287, 373

squared quantities, 168

standard decomposition, 324

standard deviation, 26, 421,

447, 448, 481

population, 421

sample, 421

standardized normal

distribution, 460, 461

standardized score, 422

standardized variable, 449

Stanley, 9, 11

statistical

decision, 44

index, 150

model, 475

statistical hypothesis

alternative, 472

exact, 41, 148

imprecise, 41, 472

inexact, 41, 148

null, 472

precise, 41, 472

rival, 225

statistical test, 39, 471, 472

binomial, 480

formal steps, 475

omnibus, 161

one-tailed, 474, 482

procedure, 53, 309, 341

refresher, 147

two-tailed, 475, 482

statistics, 40

descriptive, 1, 162, 411

inferential, 1, 41, 162

Stein, 205

Sternberg, 65

strawberry basket, 199, 200

Student, 53, 149

Student's t , 154, 273

studentized range, 273

subdesign, 242, 246, 295, 324,

327

subject factor, 152, 285

subject variability, 338

sum of cross-product, 24

sum of squares, 417

 $S \times A(B)$, 389 $S(A)$, 137 $S(A \times B)$, 298, 302 $S \times A$, 339 $S(A) \times B$, 377 $S \times A \times B$, 359

additivity, 139, 142, 244

between, 184

between-group, 138, 195,

298

between-subjects, 339

contrast, 237, 326

elements, 168

expected value, 494

general formulas, 401

partition, 297

partition of, 297

regression, 74, 102, 103, 184

residual, 74, 102, 184, 339

score model, 195

total, 72, 73, 137, 337

partition of, 138

within, 184

within-group, 137, 195

partition, 340

within-subjects, 339

sum of squares, 137

summation sign, 414
properties of, 424

T

t^2 , 155
2D space, 87
3D plot, 115
3D projection, 89
3D space, 87
Tabachnick, 396
table of critical values, 52
Taft, 283
tag variable, 8, 112, 287, 288, 375, 385, 387
Tamhane, 268, 273
test mean square, 330
theoretical approach, 47
theory, 470
 ϑ_a^2 , 197
Thomson, 112
Tibshirani, 59
total sum of squares, 72–74
partition, 102, 138
transitivity, 398
tree diagram, 439, 454, 481
Treisman, 290
trend analysis, 248
trick, 172, 176
tricky wording, 479
Tukey, 404
Tukey test, 268, 273, 275
Tulving, 283, 292, 325, 330
Tulving and Pearlstone, 303
two-factors repeated
measures design, 358
two-tailed test, 475, 482
Type I error, 42, 43, 193, 224, 229, 230, 255, 270, 476, 478
per contrast, 240
per family, 240
Type II error, 42, 43, 270, 476, 478
Type III error, 478

Type IV error, 478

U

unbalanced design, 185
unbiased estimator, 195
uncorrelated, 28
Underwood, 2, 3, 6, 9, 12, 37
union, 428
unit of measurement, 16, 24
universal set, 428
Universe of discourse, 428
Ury, 273

V

validity, 12
assumption, 191, 199, 211, 371
ecological, 11, 282, 288, 291
experimental, 345
external, 11, 282
internal, 11
variability, 470
between-group, 152
between-subjects, 342
subject, 338
within-group, 152, 338
within-subjects, 342
variable
random, 443, 445
standardized, 449
variance, 72, 121, 136, 196, 419, 445, 447, 448
definition of, 448, 486
estimated, 196
experimental source, 324
non-experimental source, 324
overestimating, 118
population, 419
proportion of common
variance, 37
refresher, 136
regression, 75
residual, 75

sample, 419
unexplained, 122
variation term, 398
vector, 236
Venn, 428
Venn diagram, 122, 428
vertical axis, 18, 66
Vokey, 47
Von Felsinger, 15

W

Webster, 371
weighted average, 447
weighted sum, 226
Weiss, 7
Wickens, 396
Wike, 207
Wine tasting, 250
Winer, 168, 273, 274, 349, 396
Wiseman, 10
within-group
degrees of freedom, 141, 184
deviation, 136, 192
distance, 136, 192
mean square, 140, 184
sum of squares, 137, 184, 195
variability, 132, 152, 338
variation, 294
within-subjects
design, 335
variability, 342

Y

yoked sample, 285
Yuille, 7

Z

Z score, 33, 82, 422, 423, 449, 482
formula, 422
Ziliak, 2, 47
Zwick, 273

