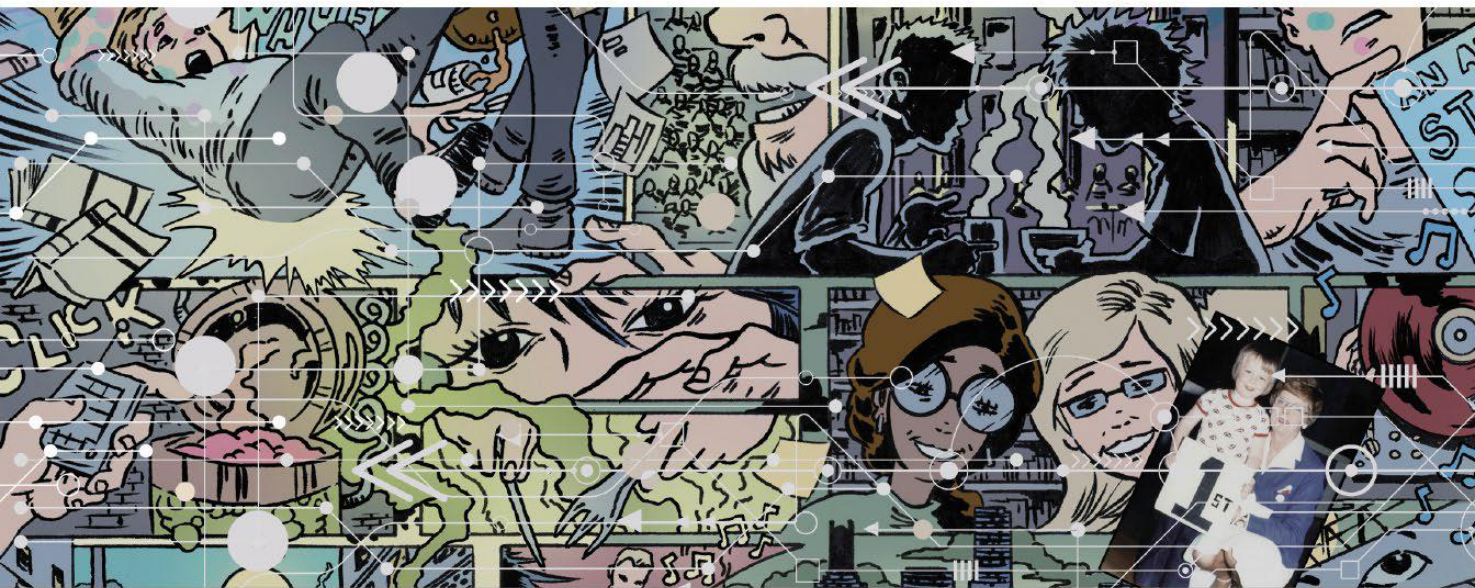


**5<sup>TH</sup>**  
**EDITION**

# **DISCOVERING STATISTICS USING IBM SPSS STATISTICS**

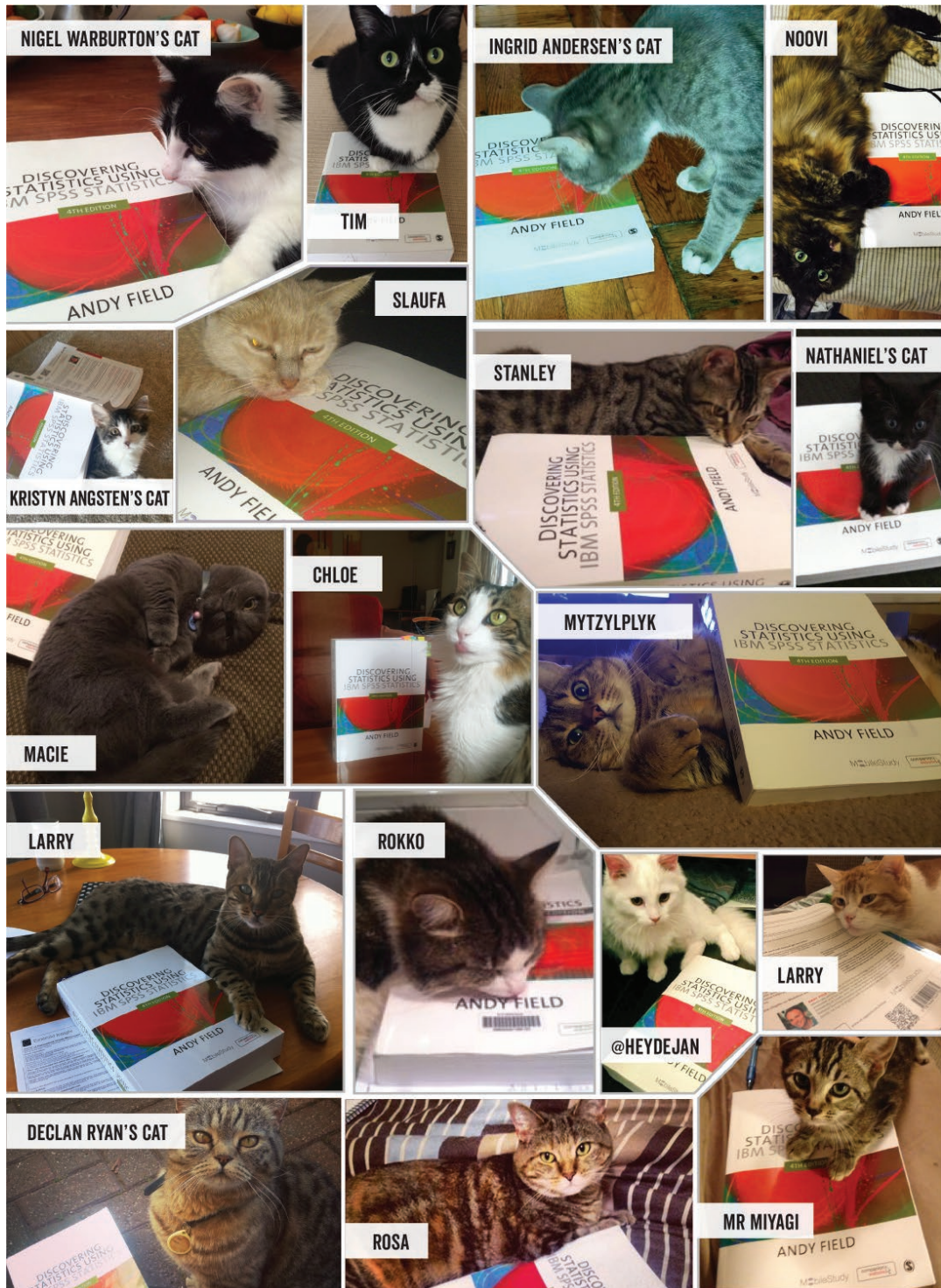


# **ANDY FIELD**

# DISCOVERING STATISTICS USING IBM SPSS STATISTICS

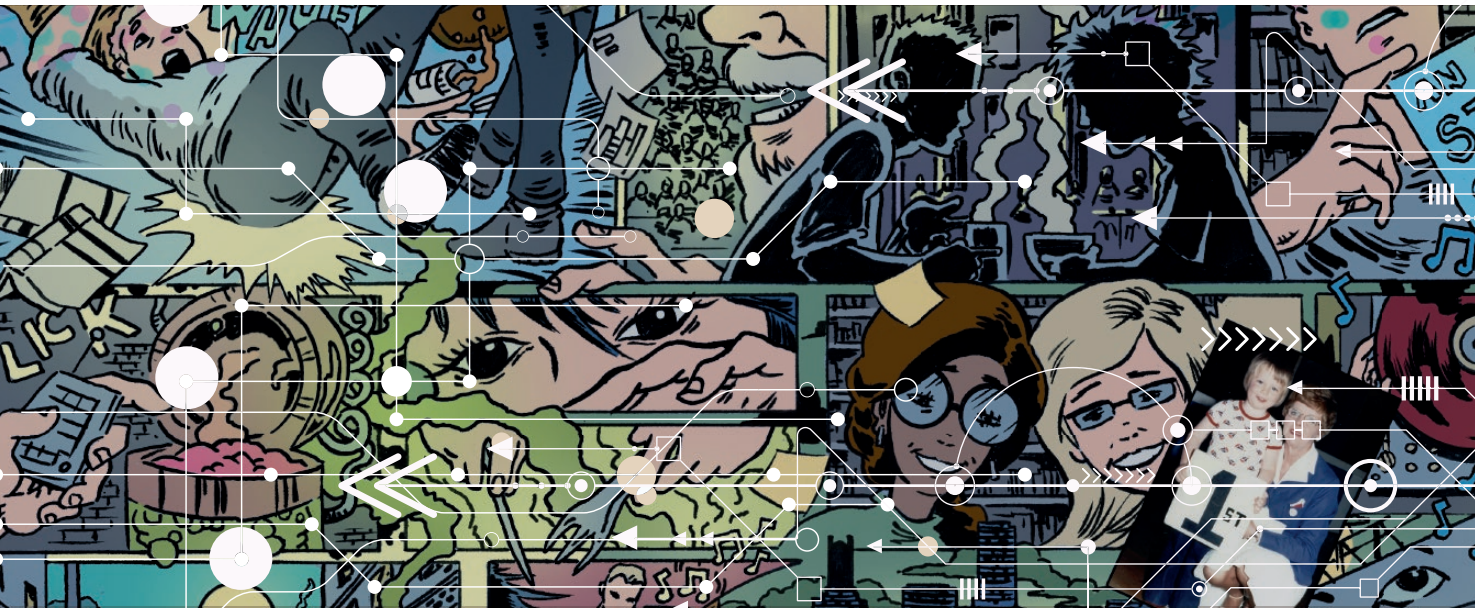


# CATISFIED CUSTOMERS



5<sup>TH</sup>  
EDITION

# DISCOVERING STATISTICS USING IBM SPSS STATISTICS



# ANDY FIELD

 SAGE

Los Angeles | London | New Delhi  
Singapore | Washington DC | Melbourne





Los Angeles | London | New Delhi  
Singapore | Washington DC | Melbourne

SAGE Publications Ltd

1 Oliver's Yard  
55 City Road  
London EC1Y 1SP

SAGE Publications Inc.

2455 Teller Road  
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd

B 1/I 1 Mohan Cooperative Industrial Area  
Mathura Road  
New Delhi 110 044

SAGE Publications Asia-Pacific Pte Ltd

3 Church Street  
#10-04 Samsung Hub  
Singapore 049483

---

Editor: Jai Seaman

Development editors: Sarah Turpie & Nina Smith

Assistant editor, digital: Chloe Statham

Production editor: Ian Antcliff

Copyeditor: Richard Leigh

Indexer: David Rudeforth

Marketing manager: Ben Griffin-Sherwood

Cover design: Wendy Scott

Typeset by: C&M Digital (P) Ltd, Chennai, India

Printed in Germany by: Mohn Media

Mohndruck GmbH

Illustration: James Iles

© Andy Field 2018

First edition published 2000

Second edition published 2005

Third edition published 2009. Reprinted 2009, 2010, 2011  
(twice), 2012

Fourth edition published 2013. Reprinted 2014 (twice), 2015,  
2016, 2017

Throughout the book, screenshots and images from IBM®  
SPSS® Statistics software ('SPSS') are reprinted courtesy of  
International Business Machines Corporation, © International  
Business Machines Corporation. SPSS Inc. was acquired by  
IBM in October 2009.

Apart from any fair dealing for the purposes of research or  
private study, or criticism or review, as permitted under the  
Copyright, Designs and Patents Act, 1988, this publication  
may be reproduced, stored or transmitted in any form, or by  
any means, only with the prior permission in writing of the  
publishers, or in the case of reprographic reproduction, in  
accordance with the terms of licences issued by the Copyright  
Licensing Agency. Enquiries concerning reproduction outside  
those terms should be sent to the publishers.

**Library of Congress Control Number: 2017954636**

**British Library Cataloguing in Publication data**

A catalogue record for this book is available from  
the British Library

ISBN 978-1-5264-1951-4

ISBN 978-1-5264-1952-1 (pbk)

At SAGE we take sustainability seriously. We print most of our products in the UK. These are produced using FSC papers and boards. We undertake an annual audit on materials used to ensure that we monitor our sustainability in what we are doing. When we print overseas, we ensure that sustainable papers are used, as measured by the PREPS grading system.

# CONTENTS

PREFACE	xv
HOW TO USE THIS BOOK	xxi
THANK YOU	xxv
SYMBOLS USED IN THIS BOOK	xxviii
SOME MATHS REVISION	xxx
1 WHY IS MY EVIL LECTURER FORCING ME TO LEARN STATISTICS?	1
2 THE SPINE OF STATISTICS	47
3 THE PHOENIX OF STATISTICS	95
4 THE IBM SPSS STATISTICS ENVIRONMENT	135
5 EXPLORING DATA WITH GRAPHS	177
6 THE BEAST OF BIAS	225
7 NON-PARAMETRIC MODELS	281
8 CORRELATION	333
9 THE LINEAR MODEL (REGRESSION)	369
10 COMPARING TWO MEANS	437
11 MODERATION, MEDIATION AND MULTICATEGORY PREDICTORS	481
12 GLM 1: COMPARING SEVERAL INDEPENDENT MEANS	519
13 GLM 2: COMPARING MEANS ADJUSTED FOR OTHER PREDICTORS (ANALYSIS OF COVARIANCE)	573
14 GLM 3: FACTORIAL DESIGNS	607
15 GLM 4: REPEATED-MEASURES DESIGNS	649
16 GLM 5: MIXED DESIGNS	703
17 MULTIVARIATE ANALYSIS OF VARIANCE (MANOVA)	735
18 EXPLORATORY FACTOR ANALYSIS	777
19 CATEGORICAL OUTCOMES: CHI-SQUARE AND LOGLINEAR ANALYSIS	835
20 CATEGORICAL OUTCOMES: LOGISTIC REGRESSION	877
21 MULTILEVEL LINEAR MODELS	935
EPILOGUE	991
APPENDIX	995
GLOSSARY	1007
REFERENCES	1043
INDEX	1057



# EXTENDED CONTENTS

<b>PREFACE</b>	<b>xv</b>
<b>HOW TO USE THIS BOOK</b>	<b>xxi</b>
<b>THANK YOU</b>	<b>xxv</b>
<b>SYMBOLS USED IN THIS BOOK</b>	<b>xxix</b>
<b>SOME MATHS REVISION</b>	<b>xxx</b>
<b>1 WHY IS MY EVIL LECTURER FORCING ME TO LEARN STATISTICS?</b>	<b>1</b>
1.1 What will this chapter tell me?	2
1.2 What the hell am I doing here? I don't belong here	3
1.3 The research process	3
1.4 Initial observation: finding something that needs explaining	4
1.5 Generating and testing theories and hypotheses	5
1.6 Collecting data: measurement	9
1.7 Collecting data: research design	16
1.8 Analysing data	22
1.9 Reporting data	40
1.10 Brian's attempt to woo Jane	44
1.11 What next?	44
1.12 Key terms that I've discovered	44
Smart Alex's tasks	45
<b>2 THE SPINE OF STATISTICS</b>	<b>47</b>
2.1 What will this chapter tell me?	48
2.2 What is the SPINE of statistics?	49
2.3 Statistical models	49
2.4 Populations and samples	53
2.5 P is for parameters	54
2.6 E is for estimating parameters	60
2.7 S is for standard error	61
2.8 I is for (confidence) interval	64
2.9 N is for null hypothesis significance testing	72
2.10 Reporting significance tests	90
2.11 Brian's attempt to woo Jane	92
2.12 What next?	92
2.13 Key terms that I've discovered	92
Smart Alex's tasks	93



<b>3</b>	<b>THE PHOENIX OF STATISTICS</b>	<b>95</b>
3.1	What will this chapter tell me?	96
3.2	Problems with NHST	97
3.3	NHST as part of wider problems with science	104
3.4	A phoenix from the EMBERS	111
3.5	Sense, and how to use it	111
3.6	Pre-registering research and open science	112
3.7	Effect sizes	113
3.8	Bayesian approaches	122
3.9	Reporting effect sizes and Bayes factors	131
3.10	Brian's attempt to woo Jane	132
3.11	What next?	133
3.12	Key terms that I've discovered	133
	Smart Alex's tasks	134
<b>4</b>	<b>THE IBM SPSS STATISTICS ENVIRONMENT</b>	<b>135</b>
4.1	What will this chapter tell me?	136
4.2	Versions of IBM SPSS Statistics	137
4.3	Windows, Mac OS, and Linux	137
4.4	Getting started	138
4.5	The data editor	139
4.6	Entering data into IBM SPSS Statistics	144
4.7	Importing data	156
4.8	The SPSS viewer	157
4.9	Exporting SPSS output	162
4.10	The syntax editor	162
4.11	Saving files	164
4.12	Opening files	165
4.13	Extending IBM SPSS Statistics	166
4.14	Brian's attempt to woo Jane	171
4.15	What next?	172
4.16	Key terms that I've discovered	172
	Smart Alex's tasks	173
<b>5</b>	<b>EXPLORING DATA WITH GRAPHS</b>	<b>177</b>
5.1	What will this chapter tell me?	178
5.2	The art of presenting data	179
5.3	The SPSS Chart Builder	181
5.4	Histograms	184
5.5	Boxplots (box-whisker diagrams)	192
5.6	Graphing means: bar charts and error bars	194
5.7	Line charts	207
5.8	Graphing relationships: the scatterplot	208
5.9	Editing graphs	218
5.10	Brian's attempt to woo Jane	221
5.11	What next?	222

5.12	Key terms that I've discovered	222
	Smart Alex's tasks	223
<b>6</b>	<b>THE BEAST OF BIAS</b>	<b>225</b>
6.1	What will this chapter tell me?	226
6.2	What is bias?	227
6.3	Outliers	227
6.4	Overview of assumptions	229
6.5	Additivity and linearity	230
6.6	Normally distributed something or other	230
6.7	Homoscedasticity/homogeneity of variance	237
6.8	Independence	239
6.9	Spotting outliers	240
6.10	Spotting normality	243
6.11	Spotting linearity and heteroscedasticity/heterogeneity of variance	257
6.12	Reducing bias	262
6.13	Brian's attempt to woo Jane	276
6.14	What next?	277
6.15	Key terms that I've discovered	278
	Smart Alex's tasks	278
<b>7</b>	<b>NON-PARAMETRIC MODELS</b>	<b>281</b>
7.1	What will this chapter tell me?	282
7.2	When to use non-parametric tests	283
7.3	General procedure of non-parametric tests using SPSS Statistics	284
7.4	Comparing two independent conditions: the Wilcoxon rank-sum test and Mann–Whitney test	285
7.5	Comparing two related conditions: the Wilcoxon signed-rank test	297
7.6	Differences between several independent groups: the Kruskal–Wallis test	306
7.7	Differences between several related groups: Friedman's ANOVA	321
7.8	Brian's attempt to woo Jane	329
7.9	What next?	331
7.10	Key terms that I've discovered	331
	Smart Alex's tasks	331
<b>8</b>	<b>CORRELATION</b>	<b>333</b>
8.1	What will this chapter tell me?	334
8.2	Modelling relationships	335
8.3	Data entry for correlation analysis	342
8.4	Bivariate correlation	344
8.5	Partial and semi-partial correlation	355
8.6	Comparing correlations	362
8.7	Calculating the effect size	363
8.8	How to report correlation coefficients	364
8.9	Brian's attempt to woo Jane	366
8.10	What next?	367

8.11	Key terms that I've discovered	367
	Smart Alex's tasks	367
<b>9</b>	<b>THE LINEAR MODEL (REGRESSION)</b>	<b>369</b>
9.1	What will this chapter tell me?	370
9.2	An introduction to the linear model (regression)	371
9.3	Bias in linear models?	381
9.4	Generalizing the model	385
9.5	Sample size and the linear model	389
9.6	Fitting linear models: the general procedure	391
9.7	Using SPSS Statistics to fit a linear model with one predictor	391
9.8	Interpreting a linear model with one predictor	394
9.9	The linear model with two or more predictors (multiple regression)	397
9.10	Using SPSS Statistics to fit a linear model with several predictors	402
9.11	Interpreting a linear model with several predictors	408
9.12	Robust regression	425
9.13	Bayesian regression	429
9.14	Reporting linear models	431
9.15	Brian's attempt to woo Jane	434
9.16	What next?	434
9.17	Key terms that I've discovered	435
	Smart Alex's tasks	435
<b>10</b>	<b>COMPARING TWO MEANS</b>	<b>437</b>
10.1	What will this chapter tell me?	438
10.2	Looking at differences	439
10.3	A mischievous example	439
10.4	Categorical predictors in the linear model	443
10.5	The <i>t</i> -test	445
10.6	Assumptions of the <i>t</i> -test	453
10.7	Comparing two means: general procedure	453
10.8	Comparing two independent means using SPSS Statistics	453
10.9	Comparing two related means using SPSS Statistics	463
10.10	Reporting comparisons between two means	476
10.11	Between groups or repeated measures?	477
10.12	Brian's attempt to woo Jane	477
10.13	What next?	479
10.14	Key terms that I've discovered	479
	Smart Alex's tasks	479
<b>11</b>	<b>MODERATION, MEDIATION AND MULTICATEGORY PREDICTORS</b>	<b>481</b>
11.1	What will this chapter tell me?	482
11.2	The <i>PROCESS</i> tool	483
11.3	Moderation: interactions in the linear model	483
11.4	Mediation	497

## EXTENDED CONTENTS

11.5	Categorical predictors in regression	508
11.6	Brian's attempt to woo Jane	516
11.7	What next?	516
11.8	Key terms that I've discovered	517
	Smart Alex's tasks	518
<b>12</b>	<b>GLM 1: COMPARING SEVERAL INDEPENDENT MEANS</b>	<b>519</b>
12.1	What will this chapter tell me?	520
12.2	Using a linear model to compare several means	521
12.3	Assumptions when comparing means	534
12.4	Planned contrasts (contrast coding)	537
12.5	<i>Post hoc</i> procedures	549
12.6	Comparing several means using SPSS Statistics	551
12.7	Output from one-way independent ANOVA	556
12.8	Robust comparisons of several means	564
12.9	Bayesian comparison of several means	566
12.10	Calculating the effect size	567
12.11	Reporting results from one-way independent ANOVA	568
12.12	Brian's attempt to woo Jane	568
12.13	What next?	570
12.14	Key terms that I've discovered	570
	Smart Alex's tasks	570
<b>13</b>	<b>GLM 2: COMPARING MEANS ADJUSTED FOR OTHER PREDICTORS (ANALYSIS OF COVARIANCE)</b>	<b>573</b>
13.1	What will this chapter tell me?	574
13.2	What is ANCOVA?	575
13.3	ANCOVA and the general linear model	576
13.4	Assumptions and issues in ANCOVA	580
13.5	Conducting ANCOVA using SPSS Statistics	584
13.6	Interpreting ANCOVA	591
13.7	Testing the assumption of homogeneity of regression slopes	598
13.8	Robust ANCOVA	600
13.9	Bayesian analysis with covariates	601
13.10	Calculating the effect size	602
13.11	Reporting results	603
13.12	Brian's attempt to woo Jane	603
13.13	What next?	604
13.14	Key terms that I've discovered	605
	Smart Alex's tasks	605
<b>14</b>	<b>GLM 3: FACTORIAL DESIGNS</b>	<b>607</b>
14.1	What will this chapter tell me?	608
14.2	Factorial designs	609
14.3	Independent factorial designs and the linear model	609
14.4	Model assumptions in factorial designs	620
14.5	Factorial designs using SPSS Statistics	620



14.6	Output from factorial designs	627
14.7	Interpreting interaction graphs	635
14.8	Robust models of factorial designs	638
14.9	Bayesian models of factorial designs	641
14.10	Calculating effect sizes	643
14.11	Reporting the results of factorial designs	645
14.12	Brian's attempt to woo Jane	646
14.13	What next?	647
14.14	Key terms that I've discovered	647
	Smart Alex's tasks	647
<b>15</b>	<b>GLM 4: REPEATED-MEASURES DESIGNS</b>	<b>649</b>
15.1	What will this chapter tell me?	650
15.2	Introduction to repeated-measures designs	651
15.3	A grubby example	652
15.4	Repeated-measures and the linear model	652
15.5	The ANOVA approach to repeated-measures designs	654
15.6	The $F$ -statistic for repeated-measures designs	658
15.7	Assumptions in repeated-measures designs	663
15.8	One-way repeated-measures designs using SPSS	664
15.9	Output for one-way repeated-measures designs	668
15.10	Robust tests of one-way repeated-measures designs	676
15.11	Effect sizes for one-way repeated-measures designs	678
15.12	Reporting one-way repeated-measures designs	679
15.13	A boozy example: a factorial repeated-measures design	680
15.14	Factorial repeated-measures designs using SPSS Statistics	681
15.15	Interpreting factorial repeated-measures designs	687
15.16	Effect sizes for factorial repeated-measures designs	698
15.17	Reporting the results from factorial repeated-measures designs	698
15.18	Brian's attempt to woo Jane	700
15.19	What next?	701
15.20	Key terms that I've discovered	701
	Smart Alex's tasks	701
<b>16</b>	<b>GLM 5: MIXED DESIGNS</b>	<b>703</b>
16.1	What will this chapter tell me?	704
16.2	Mixed designs	705
16.3	Assumptions in mixed designs	705
16.4	A speed-dating example	706
16.5	Mixed designs using SPSS Statistics	708
16.6	Output for mixed factorial designs	713
16.7	Calculating effect sizes	727
16.8	Reporting the results of mixed designs	729
16.9	Brian's attempt to woo Jane	732
16.10	What next?	733
16.11	Key terms that I've discovered	733
	Smart Alex's tasks	733

<b>17</b>	<b>MULTIVARIATE ANALYSIS OF VARIANCE (MANOVA)</b>	<b>735</b>
17.1	What will this chapter tell me?	736
17.2	Introducing MANOVA	737
17.3	Introducing matrices	739
17.4	The theory behind MANOVA	741
17.5	Practical issues when conducting MANOVA	753
17.6	MANOVA using SPSS Statistics	755
17.7	Interpreting MANOVA	757
17.8	Reporting results from MANOVA	762
17.9	Following up MANOVA with discriminant analysis	765
17.10	Interpreting discriminant analysis	767
17.11	Reporting results from discriminant analysis	771
17.12	The final interpretation	771
17.13	Brian's attempt to woo Jane	772
17.14	What next?	773
17.15	Key terms that I've discovered	775
	Smart Alex's tasks	775
<b>18</b>	<b>EXPLORATORY FACTOR ANALYSIS</b>	<b>777</b>
18.1	What will this chapter tell me?	778
18.2	When to use factor analysis	779
18.3	Factors and components	780
18.4	Discovering factors	787
18.5	An anxious example	795
18.6	Factor analysis using SPSS Statistics	800
18.7	Interpreting factor analysis	806
18.8	How to report factor analysis	819
18.9	Reliability analysis	821
18.10	Reliability analysis using SPSS Statistics	824
18.11	Interpreting reliability analysis	826
18.12	How to report reliability analysis	830
18.13	Brian's attempt to woo Jane	831
18.14	What next?	831
18.15	Key terms that I've discovered	831
	Smart Alex's tasks	832
<b>19</b>	<b>CATEGORICAL OUTCOMES: CHI-SQUARE AND LOGLINEAR ANALYSIS</b>	<b>835</b>
19.1	What will this chapter tell me?	836
19.2	Analysing categorical data	837
19.3	Associations between two categorical variables	837
19.4	Associations between several categorical variables: loglinear analysis	846
19.5	Assumptions when analysing categorical data	849
19.6	General procedure for analysing categorical outcomes	850
19.7	Doing chi-square using SPSS Statistics	850
19.8	Interpreting the chi-square test	854
19.9	Loglinear analysis using SPSS Statistics	864

19.10	Interpreting loglinear analysis	866
19.11	Reporting the results of loglinear analysis	872
19.12	Brian's attempt to woo Jane	872
19.13	What next?	874
19.14	Key terms that I've discovered	875
	Smart Alex's tasks	875
<b>20</b>	<b>CATEGORICAL OUTCOMES: LOGISTIC REGRESSION</b>	<b>877</b>
20.1	What will this chapter tell me?	878
20.2	What is logistic regression?	879
20.3	Theory of logistic regression	879
20.4	Sources of bias and common problems	886
20.5	Binary logistic regression	891
20.6	Interpreting logistic regression	900
20.7	Reporting logistic regression	911
20.8	Testing assumptions: another example	911
20.9	Predicting several categories: multinomial logistic regression	916
20.10	Reporting multinomial logistic regression	930
20.11	Brian's attempt to woo Jane	930
20.12	What next?	932
20.13	Key terms that I've discovered	932
	Smart Alex's tasks	933
<b>21</b>	<b>MULTILEVEL LINEAR MODELS</b>	<b>935</b>
21.1	What will this chapter tell me?	936
21.2	Hierarchical data	937
21.3	Theory of multilevel linear models	941
21.4	The multilevel model	944
21.5	Some practical issues	948
21.6	Multilevel modelling using SPSS Statistics	951
21.7	Growth models	972
21.8	How to report a multilevel model	987
21.9	A message from the octopus of inescapable despair	988
21.10	Brian's attempt to woo Jane	988
21.11	What next?	988
21.12	Key terms that I've discovered	990
	Smart Alex's tasks	990
	<b>EPILOGUE</b>	<b>991</b>
	<b>APPENDIX</b>	<b>995</b>
	<b>GLOSSARY</b>	<b>1007</b>
	<b>REFERENCES</b>	<b>1043</b>
	<b>INDEX</b>	<b>1057</b>

# PREFACE

*Karma Police, arrest this man, he talks in maths, he buzzes like a fridge, he's like a detuned radio*  
Radiohead, 'Karma Police', *OK Computer* (1997)

## Introduction

Many behavioural and social science students (and researchers for that matter) despise statistics. Most of us have a non-mathematical background, which makes understanding complex statistical equations very difficult. Nevertheless, the evil goat-warriors of Satan force our non-mathematical brains to apply themselves to what is the very complex task of becoming a statistics expert. The end result, as you might expect, can be quite messy. The one weapon that we have is the computer, which allows us to neatly circumvent the considerable disability of not understanding mathematics. Computer programs such as IBM SPSS Statistics, SAS, R, JASP and the like provide an opportunity to teach statistics at a conceptual level without getting too bogged down in equations. The computer to a goat-warrior of Satan is like catnip to a cat: it makes them rub their heads along the ground and purr and dribble ceaselessly. The only downside of the computer is that it makes it really easy to make a complete idiot of yourself if you don't understand what you're doing. Using a computer without any statistical knowledge at all can be a dangerous thing. Hence this book.

My first aim is to strike a balance between theory and practice: I want to use the computer as a tool for teaching statistical concepts in the hope that you will gain a better understanding of both theory and practice. If you want theory and you like equations then there are certainly more technical books. However, if you want a stats book that also discusses digital rectal stimulation, then you have just spent your money wisely.

Too many books create the impression that there is a 'right' and 'wrong' way to do statistics. Data analysis is more subjective than is often made out. Therefore, although I make recommendations, within the limits imposed by the senseless destruction of rainforests, I hope to give you enough background in theory to enable you to make your own decisions about how best to conduct your analysis.

A second (ridiculously ambitious) aim is to make this the only statistics book that you'll ever need to buy (sort of). It's a book that I hope will become your friend from your first year at university right through to your professorship. The start of the book is aimed at first-year undergraduates (Chapters 1–10), and then we move onto second-year undergraduate-level material (Chapters 6, 9 and 11–16) before a dramatic climax that should keep postgraduates tickled (Chapters 17–21). There should be something for everyone in each chapter, and to help you gauge the difficulty of material, I flag the level of each section within each chapter (more on that later).

My final and most important aim is to make the learning process fun. I have a sticky history with maths. This extract is from my school report at the age of 11:

MATHEMATICS ADDL. MATHS.	43	59	27	D	C	His work shows lack of discipline in thought and presentation. I hope it will mature next year.	...
CHEMISTRY							



The '27=' in the report is to say that I came equal 27th with another student out of a class of 29. That's pretty much bottom of the class. The 43 is my exam mark as a percentage. Oh dear. Four years later (at 15), this was my school report:

NAME Andrew Field FORM 4D SUBJECT Mathematics

Andrew's progress in Mathematics has been remarkable. From being a weaker candidate who lacked confidence he has developed into a budding Mathematician. He should achieve a good grade.

EXAM	
ATTAINMENT	
EFFORT	

Date 27/6/88 B.A. Oreste Subject Teacher

The catalyst of this remarkable change was a good teacher: my brother, Paul. I owe my life as an academic to Paul's ability to teach me stuff in an engaging way – something my maths teachers failed to do. Paul's a great teacher because he cares about bringing out the best in people, and he was able to make things interesting and relevant to me. Everyone should have a brother Paul to teach them stuff when they're throwing their maths book at their bedroom wall, and I will attempt to be yours.

I strongly believe that people appreciate the human touch, and so I inject a lot of my own personality and sense of humour (or lack of) into *Discovering Statistics Using ...* books. Many of the examples in this book, although inspired by some of the craziness that you find in the real world, are designed to reflect topics that play on the minds of the average student (i.e., sex, drugs, rock and roll, celebrity, people doing crazy stuff). There are also some examples that are there simply because they made me laugh. So, the examples are light-hearted (some have said 'smutty', but I prefer 'light-hearted') and by the end, for better or worse, I think you will have some idea of what goes on in my head on a daily basis. I apologize to those who think it's crass, hate it, or think that I'm undermining the seriousness of science, but, come on, what's not funny about a man putting an eel up his anus?

I never believe that I meet my aims, but previous editions have certainly been popular. I enjoy the rare luxury of having complete strangers emailing me to tell me how wonderful I am. (Admittedly, there are also emails accusing me of all sorts of unpleasant things, but I've usually got over them after a couple of months.) With every new edition, I fear that the changes I make will ruin my previous hard work. Let's see what you're going to get and what's different this time around.

## What do you get for your money?

This book takes you on a journey (I try my best to make it a pleasant one) not just of statistics but also of the weird and wonderful contents of the world and my brain. It's full of daft examples, bad jokes, and smut. Aside from the smut, I have been forced, reluctantly, to include some academic content. It contains everything I know about statistics (actually, more than I know ...). It also has these features:

- **Everything you'll ever need to know:** I want this book to be good value for money, so it guides you from complete ignorance (Chapter 1 tells you the basics of doing research) to being an expert in multilevel linear modelling (Chapter 21). Of course, no book can contain everything, but I think this one has a fair crack. It's pretty good for developing your biceps also.
- **Stupid faces:** You'll notice that the book is riddled with 'characters', some of them my own. You can find out more about the pedagogic function of these 'characters' in the next section.
- **Data sets:** There are about 132 data files associated with this book on the companion website. Not unusual in itself for a statistics book, but my data sets contain more sperm (not literally) than other books. I'll let you judge for yourself whether this is a good thing.
- **My life story:** Each chapter is book-ended by a chronological story from my life. Does this help you to learn about statistics? Probably not, but it might provide light relief between chapters.
- **SPSS tips:** SPSS does confusing things sometimes. In each chapter, there are boxes containing tips, hints and pitfalls related to SPSS.
- **Self-test questions:** Given how much students hate tests, I thought that the best way to commit commercial suicide was to liberally scatter tests throughout each chapter. These range from simple questions to test what you have just learned to going back to a technique that you read about several chapters before and applying it in a new context. All of these questions have answers so that you can check on your progress.
- **Online resources:** The website contains an insane amount of additional material, which no one reads, but it is described in the section about the online resources so that you know what you're ignoring.
- **Digital stimulation:** No, not the aforementioned type of digital stimulation, but brain stimulation. Many of the features on the website will be accessible from tablets and smartphones, so that when you're bored in the cinema you can read about the fascinating world of heteroscedasticity instead.
- **Reporting your analysis:** Every chapter has a guide to writing up your analysis. How one writes up an analysis varies a bit from one discipline to another, but my guides should get you heading in the right direction.
- **Glossary:** Writing the glossary was so horribly painful that it made me stick a vacuum cleaner into my ear to suck out my own brain. You can find my brain in the bottom of the vacuum cleaner in my house.
- **Real-world data:** Students like to have 'real data' to play with. The trouble is that real research can be quite boring. I trawled the world for examples of research on really fascinating topics (in my opinion). I then stalked the authors of the research until they gave me their data. Every chapter has a real research example.

## What do you get that you didn't get last time?

I suppose if you have spent your hard-earned money on the previous edition, it's reasonable that you want a good reason to spend more of your hard-earned money on this edition. In some respects, it's hard to quantify all of the changes in a list: I'm a better writer than I was five years ago, so there is a lot of me rewriting things because I think I can do it better than before. I spent 6 months solidly on the updates, so, suffice it to say that a lot has changed; but anything you might have liked about the previous edition probably hasn't changed:

- **IBM SPSS compliance:** This edition was written using version 25 of IBM SPSS Statistics. IBM releases new editions of SPSS Statistics more often than I bring out new editions of this book, so,

depending on when you buy the book, it may not reflect the latest version. This shouldn't worry you because the procedures covered in this book are unlikely to be affected (see Section 4.12).

- **New! Chapter:** In the past four years the open science movement has gained a lot of momentum. Chapter 3 is new and discusses issues relevant to this movement such as *p*-hacking, HARKing, researcher degrees of freedom, and pre-registration of research. It also has an introduction to Bayesian statistics.
- **New! Bayes:** Statistical times are a-changing, and it's more common than it was four years ago to encounter Bayesian methods in social science research. IBM SPSS Statistics doesn't really do Bayesian estimation, but you can implement Bayes factors. Several chapters now include sections that show how to obtain and interpret Bayes factors. Chapter 3 also explains what a Bayes factor is.
- **New! Robust methods:** Statistical times are a-changing ... oh, hang on, I just said that. Although IBM SPSS Statistics does bootstrap (if you have the premium version), there are a bunch of statistics based on trimmed data that are available in R. I have included several sections on robust tests and syntax to do them (using R).
- **New! Pointless fiction:** Having got quite into writing a statistics textbook in the form of a fictional narrative (*An Adventure in Statistics*) I staved off boredom by fleshing out Brian and Jane's story (which goes with the diagrammatic summaries at the end of each chapter). Of course, it is utterly pointless, but maybe someone will enjoy the break from the stats.
- **New! Misconceptions:** Since the last edition my cat of 20 years died, so I needed to give him a more spiritual role. He has become the Correcting Cat, and he needed a foil, so I created the Misconception Mutt, who has a lot of common misconceptions about statistics. So, the mutt (based on my cocker spaniel) gets stuff wrong and the cat appears from the spiritual ether to correct him. All of which is an overly elaborate way to point out some common misconceptions.
- **New-ish! The linear model theme:** In the past couple of editions of this book I've been keen to scaffold the content on the linear model to focus on the commonalities between models traditionally labelled as regression, ANOVA, ANCOVA, *t*-tests, etc. I've always been mindful of trying not to alienate teachers who are used to the historical labels, but I have again cranked up a level the general linear model theme.
- **New-ish! Characters:** I loved working with James Iles on *An Adventure in Statistics* so much that I worked with him to create new versions of the characters in the book (and other design features like their boxes). They look awesome.

Every chapter had a thorough edit/rewrite, I've redone all of the figures, and obviously updated the SPSS Statistics screenshots and output. Here is a chapter-by-chapter rundown of the more substantial changes:

- **Chapter 1 (Doing research):** I changed the way I discuss hypotheses. I changed my suicide example to be about memes.
- **Chapter 2 (Statistical theory):** I restructured this chapter around the acronym of SPINE,<sup>1</sup> so you'll notice that subheadings have changed and so on. The content is all there, just rewritten and reorganized into a better narrative. I've expanded my description of null hypothesis significance testing (NHST).

---

<sup>1</sup> Thanks to a colleague, Jennifer Mankin, for distracting me from the acronym that more immediately sprang to my childish mind.

- **Chapter 3 (Current thinking in statistics):** This chapter is completely new. It co-opts some of the critique of NHST that used to be in Chapter 2 but moves this into a discussion of open science, *p*-hacking, *HARK*ing, researcher degrees of freedom, pre-registration, and ultimately Bayesian statistics (primarily Bayes factors).
- **Chapter 4 (IBM SPSS Statistics):** Obviously reflects changes to SPSS Statistics since the previous edition. There's a new section on 'extending' SPSS Statistics that covers installing the PROCESS tool, the *Essentials for R* plugin and *WRS2* package (for robust tests).
- **Chapter 5 (Graphs):** No substantial changes, I just tweaked a few examples.
- **Chapter 6 (Assumptions):** The content is more or less as it was. I have a much stronger steer away from tests of normality and homogeneity (although I still cover them) because I now offer some robust alternatives to common tests.
- **Chapter 7 (Nonparametric models):** No substantial changes to content.
- **Chapter 8 (Correlation):** I completely rewrote the section on partial correlations.
- **Chapter 9 (The linear model):** I restructured this chapter a bit and wrote new sections on robust regression and Bayesian regression.
- **Chapter 10 (*t*-tests):** I did an overhaul of the theory section to tie it in more with the linear model theme. I wrote new sections on robust and Bayesian tests of two means.
- **Chapter 11 (Mediation and moderation):** No substantial changes to content.
- **Chapters 12–13 (GLM 1–2):** I changed the main example to be about puppy therapy. I thought that the Viagra example was a bit dated, and I needed an excuse to get some photos of my spaniel into the book. This was the perfect solution. I wrote new sections on robust and Bayesian (Chapter 12 only) variants of these models.
- **Chapter 14 (GLM 3):** I tweaked the example – it's still about the beer-goggles effect, but I linked it to some real research so that the findings now reflect some actual science that's been done. I added sections on robust and Bayesian variants of models for factorial designs.
- **Chapters 15–16 (GLM 4–5):** I added some theory to Chapter 14 to link it more closely to the linear model (and to the content of Chapter 21). I give a clearer steer now to ignoring Mauchly's test and routinely applying a correction to *F* (although, if you happen to like Mauchly's test, I doubt that the change is dramatic enough to upset you). I added sections on robust variants of models for repeated-measures designs. I added some stuff on pivoting trays in tables. I tweaked the example in Chapter 16 a bit so that it doesn't compare males and females but instead links to some real research on dating strategies.
- For **Chapters 17 (MANOVA); 18 (Factor analysis); 19 (Categorical data); 20 (Logistic regression); and 21 (Multilevel models)** there are no major changes, except to improve the structure in Chapter 19.



## Goodbye

The first edition of this book was the result of two years (give or take a few weeks to write up my PhD) of trying to write a statistics book that I would enjoy reading. With each new edition I try not just to make superficial changes but also to rewrite and improve everything (one of the problems with getting older is that you look back at your past work and think you can do things better). This fifth edition is the culmination of about seven years of full-time work (on top of my actual job). This book has consumed the last 20 years or so of my life, and each time I get a nice email from someone who found it useful, I am reminded that it is the most useful thing I'll ever do with my academic life. It began and continues to be a labour of love. It still isn't perfect, and I still love to have feedback (good or bad) from the people who matter most: you.

Andy



[www.facebook.com/profandyfield](http://www.facebook.com/profandyfield)



[@ProfAndyField](https://twitter.com/ProfAndyField)



[www.youtube.com/user/ProfAndyField](http://www.youtube.com/user/ProfAndyField)



[www.discoveringstatistics.com/category/blog/](http://www.discoveringstatistics.com/category/blog/)



# HOW TO USE THIS BOOK

When the publishers asked me to write a section on ‘How to use this book’ it was tempting to write ‘Buy a large bottle of Olay anti-wrinkle cream (which you’ll need to fend off the effects of ageing while you read), find a comfy chair, sit down, fold back the front cover, begin reading and stop when you reach the back cover.’ However, I think they wanted something more useful. 😊

## What background knowledge do I need?

In essence, I assume that you know nothing about statistics, but that you have a very basic grasp of computers (I won’t be telling you how to switch them on, for example) and maths (although I have included a quick revision of some very basic concepts).

## Do the chapters get more difficult as I go through the book?

Yes, more or less: Chapters 1–10 are first-year degree level, Chapters 9–16 move into second-year degree level, and Chapters 17–21 discuss more technical topics. However, my aim is to tell a statistical story rather than worry about what level a topic is at. Many books teach different tests in isolation and never really give you a grasp of the similarities between them; this, I think, creates an unnecessary mystery. Most of the tests in this book are the same thing expressed in slightly different ways. I want the book to tell this story, and I see it as consisting of seven parts:

- Part 1 (Doing research and introducing linear models): Chapters 1–4.
- Part 2: (Exploring data): Chapters 5–7.
- Part 3: (Linear models with continuous predictors): Chapters 8–9.
- Part 4: (Linear models with continuous or categorical predictors): Chapters 10–16.
- Part 5: (Linear models with multiple outcomes): Chapter 17–18.
- Part 6 (Linear models with categorical outcomes): Chapters 19–20.
- Part 7 (Linear models with hierarchical data structures): Chapter 21.

This structure might help you to see the method in my madness. If not, to help you on your journey, I’ve coded each section with an icon. These icons are designed to give you an idea of the difficulty of the section. It doesn’t mean that you can skip the sections, but it will let you know whether a section is at about your level, or whether it’s going to push you. It’s based on a wonderful categorization system using the letter ‘T’:



**Introductory**, which I hope means that everyone will understand these sections. These are for people just starting their undergraduate courses.

- |||| **Intermediate.** Anyone with a bit of background in statistics should be able to get to grips with these sections. They are aimed at people who are perhaps in the second year of their degree, but they can still be quite challenging in places.
- |||| **In at the deep end.** These topics are difficult. I'd expect final-year undergraduates and recent postgraduate students to be able to tackle these sections.
- |||| **Incinerate your brain.** These are difficult topics. I would expect these sections to be challenging for undergraduates, but postgraduates with a reasonable background in research methods shouldn't find them too much of a problem.

## Why do I keep seeing silly faces everywhere?



**Brian Haemorrhage:** Brian is a really nice guy, and he has a massive crush on Jane Superbrain. He's seen her around the university campus carrying her jars of brains (see below). Whenever he sees her, he gets a knot in his stomach and he imagines slipping a ring onto her finger on a beach in Hawaii, as their friends and family watch through gooey eyes. Jane never even notices him; this makes him very sad. His friends have told him that the only way she'll marry him is if he becomes a statistics genius (and changes his surname). Therefore, he's on a mission to learn statistics. It's his last hope of impressing Jane, settling down and living happily ever after. At the moment he knows nothing, but he's about to go on a journey that will take him from statistically challenged to a genius, in 1104 pages. Along his journey he pops up and asks questions, and at the end of each chapter he flaunts his newly found knowledge to Jane in the hope that she'll go on a date with him.



**Confusius:** The great philosopher Confucius had a lesser-known brother called Confusius. Jealous of his brother's great wisdom and modesty, Confusius vowed to bring confusion to the world. To this end, he built the Confusion machine. He puts statistical terms into it, and out of it come different names for the same concept. When you see Confusius he will be alerting you to statistical terms that mean the same thing.



**Correcting Cat:** This cat lives in the ether and appears to taunt the Misconception Mutt by correcting his misconceptions. He also appears when I want to make a bad cat-related pun. He exists in loving memory of my own ginger cat who, after 20 years as the star of this book, sadly passed away, which he promised me he'd never do. You can't trust a cat.



**Cramming Sam:** Samantha thinks that statistics is a boring waste of time. She just wants to pass her exam and forget that she ever had to know anything about normal distributions. She appears and gives you a summary of the key points that you need to know. If, like Samantha, you're cramming for an exam, she will tell you the essential information to save you having to trawl through hundreds of pages of my drivel.



**Jane Superbrain:** Jane is the cleverest person in the universe. She has acquired a vast statistical knowledge, but no one knows how. She is an enigma, an outcast, and a mystery. Brian has a massive crush on her. Jane appears to tell you advanced things that are a bit tangential to the main text. Can Brian win his way into her heart? You'll have to read and find out.



**Labcoat Leni:** Leni is a budding young scientist and he's fascinated by real research. He says, 'Andy, I like an example about using an eel as a cure for constipation as much as the next guy, but all of your data are made up. We need some real examples, buddy!' Leni walked the globe, a lone data warrior in a thankless quest for real data. When you see Leni you know that you will get some real data, from a real research study, to analyse.



**Misconception Mutt:** Since the last edition, I acquired a spaniel called Ramsey. I needed some way to get him into the book, so here he is as the Misconception Mutt. He follows his owner to statistics lectures and finds himself learning about stats. Sometimes, he gets things wrong, though, and, when he does, something very strange happens. A ginger cat materializes out of nowhere and corrects him.



**Oditi's Lantern:** Oditi believes that the secret to life is hidden in numbers and that only by large-scale analysis of those numbers shall the secrets be found. He didn't have time to enter, analyse, and interpret all of the data in the world, so he established the cult of undiscovered numerical truths. Working on the principle that if you gave a million monkeys typewriters, one of them would recreate Shakespeare, members of the cult sit at their computers crunching numbers in the hope that one of them will unearth the hidden meaning of life. To help his cult Oditi has set up a visual vortex called 'Oditi's Lantern'. When Oditi appears, it is to implore you to stare into the lantern, which basically means that there is a video tutorial to guide you.



**Oliver Twisted:** With apologies to Charles Dickens, Oliver, like the more famous fictional London urchin, asks, 'Please, Sir, can I have some more?' Unlike Master Twist, though, Master Twisted always wants more statistics information. Who wouldn't? Let us not be the ones to disappoint a young, dirty, slightly smelly boy who dines on gruel. When Oliver appears, he's telling you that there is additional information on the companion website. (It took a long time to write, so, someone, please, actually read it.)



**Satan's Personal Statistics Slave:** Satan is a busy boy – he has all of the lost souls to torture in hell, then there are the fires to keep fuelled, not to mention organizing enough carnage on the planet's surface to keep Norwegian black metal bands inspired. Like many of us, this leaves little time for him to analyse data, and this makes him very sad. So, he has his own personal slave, who, also like some of us, spends all day dressed in a gimp mask and tight leather pants in front of IBM SPSS Statistics, analysing Satan's data. Consequently, he knows a thing or two about SPSS, and when Satan's busy spanking a goat, he pops up in a box with SPSS tips.



**Smart Alex:** Alex was aptly named because she's, like, super smart. She likes teaching people, and her hobby is posing people questions so that she can explain the answers to them. Alex appears at the end of each chapter to pose you some questions. Her answers are on the companion website.

## What online resources do you get with the book?

I've put a cornucopia of additional funk on that worldwide interweb thing. To enter my world of delights, go to <https://edge.sagepub.com/field5e>. The website contains resources for students and lecturers alike, with additional content from some of the characters from the book:

- **Testbank:** There is a (hopefully) comprehensive testbank of multiple-choice and numeracy-based/algorithmic questions for your instructors to use. It comes as a file that you can upload into your institution's online teaching system. Furthermore, there are additional testbanks of multiple-choice questions for your instructors.
- **Data files:** You need data files to work through the examples in the book, and they are on the website. We did this to force you to go there and, once you're there, SAGE will flash up subliminal messages that make you buy more of their books.
- **Resources for other subject areas:** I am a psychologist and, although I tend to base my examples around the weird and wonderful, I do have a nasty habit of resorting to psychology when I don't have any better ideas. My publishers have recruited some non-psychologists to provide data files and an instructor's testbank of multiple-choice questions for those studying or teaching in business and management, education, sport sciences and health sciences. You have no idea how happy I am that I didn't have to write those.
- **YouTube:** Whenever you see Oditi in the book it means that there is a screencast to accompany the chapter. These are hosted on my YouTube channel ([www.youtube.com/user/ProfAndyField](http://www.youtube.com/user/ProfAndyField)), which I have amusingly called  $\mu$ -Tube (see what I did there?). 
- **Self-assessment multiple-choice questions:** Organized by difficulty, or what you need to practice, these allow you to test whether wasting your life reading this book has paid off so that you can annoy your friends by walking with an air of confidence into the examination. If you fail said exam, please don't sue me.
- **Flashcard glossary:** As if a printed glossary wasn't enough, my publishers insisted that you'd like an electronic one too. Have fun here flipping through terms and definitions covered in the textbook; it's better than actually learning something.
- **Oliver Twisted's pot of gruel:** Oliver Twisted will draw your attention to the 300 pages or so of more information that we have put online so that (1) the planet suffers a little less, and (2) you won't die when the book falls from your bookshelf onto your head.
- **Labcoat Leni solutions:** There are full answers to the Labcoat Leni tasks.
- **Smart Alex answers:** Each chapter ends with a set of tasks for you to test your newly acquired expertise. The chapters are also littered with self-test questions. The companion website contains detailed answers. Will I ever stop writing?
- **PowerPoint slides:** I can't come and teach you all in person (although you can watch my lectures on YouTube). Instead, I rely on a crack team of highly skilled and super-intelligent pan-dimensional beings called 'lecturers'. I have personally grown each and every one of them in a greenhouse in my garden. To assist in their mission to spread the joy of statistics I have provided them with PowerPoint slides for each chapter. If you see something weird on their slides that upsets you, then remember that's probably my fault.
- **Links:** There are the obligatory links to other useful sites.
- **SAGE:** My publishers are giving you a tonne of free material from their books, journals and digital products. If you want it.
- **SAGE Research Methods** is a digital platform full of  **research methods** research methods stuff. Some of it, including videos and a  **datasets** test yourself maths diagnostic tool, is available for free on the companion website.
- **Cyberworms of knowledge:** I have used nanotechnology to create cyberworms that crawl down your broadband, wifi or 4G, pop out of a port on your computer, tablet, iPad or phone, and fly through space into your brain. They rearrange your neurons so that you understand statistics. You don't believe me? You'll never know for sure unless you visit the online resources ...

Happy reading, and don't get distracted by social media.



# THANK YOU

**Colleagues:** This book (in the SPSS, SAS, and R version) wouldn't have happened if not for Dan Wright's unwarranted faith in a then postgraduate to write the first SPSS edition. Numerous other people have contributed to previous editions of this book. I don't have room to list them all, but particular thanks to Dan (again), David Hitchin, Laura Murray, Gareth Williams, Lynne Slocombe, Kate Lester, Maria de Ridder, Thom Baguley, Michael Spezio, and my wife Zoë who have given me invaluable feedback during the life of this book. Special thanks to Jeremy Miles. Part of his 'help' involves ranting on at me about things I've written being, and I quote, 'bollocks'. Nevertheless, working on the SAS and R versions of this book with him has influenced me enormously. He's also been a very nice person to know over the past few years (apart from when he's ranting on at me about ... ). For this edition, J. W. Jacobs, Ann-Will Kruijt, Johannes Petzold, and E.-J. Wagenmakers provided particularly useful feedback.

Thanks to the following for allowing me to use their raw data – it's an honour for me to include their fascinating research in my book: Rebecca Ang, Philippe Bernard, Hakan Çetinkaya, Tomas Chamorro-Premuzic, Graham Davey, Mike Domjan, Gordon Gallup, Nicolas Guéguen, Sarah Johns, Eric Lacourse, Nate Lambert, Sarah Marzillier, Karlijn Massar, Geoffrey Miller, Peter Muris, Laura Nichols, Nick Perham, Achim Schützwohl, Mirjam Tuk, and Lara Zibarras.

I appreciate everyone who has taken time to write nice reviews of this book on the various Amazon (and other) websites around the world; the success of this book has been in no small part due to these people being so positive and constructive in their feedback. Thanks also to everyone who participates so enthusiastically in my Facebook and Twitter pages: I always hit motivational dark times when I'm writing, but feeling the positive vibes from readers always gets me back on track (especially the photos of cats, dogs, parrots, and lizards with my books 😊). I continue to be amazed and bowled over by the nice things that people say about the book (and disproportionately upset by the less positive things).

Not all contributions are as tangible as those above. Very early in my career, Graham Hole made me realize that teaching research methods didn't have to be dull. My approach to teaching has been to steal his good ideas, and he has had the good grace not to ask for them back! He is a rarity in being brilliant, funny, and nice.

**Software:** This book wouldn't exist without the generous support of International Business Machines Corporation (IBM), who allow me to beta test IBM® SPSS® Statistics software ('SPSS'), kept me up to date with the software while I wrote this update, and kindly granted permission for me to include screenshots and images from SPSS. I wrote this edition on MacOS but used Windows for the screenshots. Mac and Mac OS are trademarks of Apple Inc., registered in the United States and other countries; Windows is a registered trademark of Microsoft Corporation in the United States and other countries. I don't get any incentives for saying this (perhaps I should, hint, hint ...) but the following software packages are invaluable to me when writing: TechSmith's ([www.techsmith.com](http://www.techsmith.com)) Camtasia (which I use to produce videos) and Snagit (which I use for screenshots) for Mac; the Omnigroup's ([www.omnigroup.com](http://www.omnigroup.com)) OmniGraffle, which I use to create most of the diagrams and flowcharts (it is awesome); and R (in particular, Hadley Wickham's *ggplot2* package) and R Studio, which I use for data visualizations.



**Publishers:** My publishers, SAGE, are rare in being a large, successful company that manages to maintain a family feel. For this edition I was particularly grateful for them trusting me enough to leave me alone to get on with things because my deadline was insane. Now that I have emerged from my attic, I'm fairly sure that I'm going to be grateful to Jai Seaman and Sarah Turpie for what they have been doing and will do to support the book. A long-overdue thank you to Richard Leigh, who has copyedited my books over many years and never gets thanked because his job begins after I've written the acknowledgements! My long-suffering production editor, Ian Antcliff, deserves special mention not only for the fantastic job he does but also for being the embodiment of calm when the pressure is on. I'm also grateful to Karen and Ziyad who don't work directly on my books but are such an important part of my fantastic relationship with SAGE.

James Iles redesigned the characters in this book and produced the artwork for the pedagogic boxes. I worked with James on another book where there was a lot more artwork (*An Adventure in Statistics*) and it was an incredible experience. I'm delighted that that experience didn't put him off working with me again. It's an honour to have his artwork in another of my books.

**Music:** I always write while listening to music. For this edition, I predominantly enjoyed (my neighbours less so): AC/DC, A Forest of Stars, Alice Cooper, Alter of Plagues, Anathema, Animals as Leaders, Anthrax, Billy Cobham, Blackfield, Deafheaven, Deathspell Omega, Deep Purple, Enslaved, Faith No More, Genesis (Peter Gabriel era), Ghost, Ghost Bath, Glenn Hughes, Gojira, Gorguts, Iced Earth, Ihsahn, The Infernal Sea, Iron Maiden, Judas Priest, Katatonia, Kiss, Marillion, Meshuggah, Metallica, MGLA, Motörhead, Primal Rock Rebellion, Opeth, Oranssi Pazuzu, Rebirth of Nefast, Royal Thunder, Satyricon, Skuggsjä, Status Quo (R.I.P. Rick 😞), Steven Wilson, Thin Lizzy, Wolves in the Throne Room.

**Friends and family:** All this book-writing nonsense requires many lonely hours of typing. Without some wonderful friends to drag me out of my dimly lit room from time to time I'd be even more of a gibbering cabbage than I already am. Across many editions, my eternal gratitude goes to Graham Davey, Ben Dyson, Kate Lester, Mark Franklin, and their lovely families for reminding me that there is more to life than work. I throw a robust set of horns to my brothers of metal, Rob Mephram, Nick Paddy, and Ben Anderson, for letting me deafen them with my drumming. Thanks to my parents and Paul and Julie for being my family. Special cute thanks to my niece and nephew, Oscar and Melody: I hope to teach you many things that will annoy your parents.

For someone who spends his life writing, I'm constantly surprised at how incapable I am of finding words to express how wonderful my wife Zoë is. She has a never-ending supply of patience, love, support, and optimism (even when her husband is a grumpy, sleep-deprived, withered, self-doubting husk). I never forget, not even for a nanosecond, how lucky I am. Finally, since the last edition, I made a trivial contribution to creating two humans: Zach and Arlo. I thank them for the realization of how utterly pointless work is and for the permanent feeling that my heart has expanded to bursting point from trying to contain my love for them.

Like the previous editions, this book is dedicated to my brother Paul and my cat Fuzzy (now in the spirit cat world), because one of them was an intellectual inspiration and the other woke me up in the morning by sitting on me and purring in my face until I gave him cat food: mornings were considerably more pleasant when my brother got over his love of cat food for breakfast. 😊



# SYMBOLS USED IN THIS BOOK

## Mathematical operators

$\Sigma$	This symbol (called sigma) means ‘add everything up’. So, if you see something like $\sum x_i$ it means ‘add up all of the scores you’ve collected’.
$\Pi$	This symbol means ‘multiply everything’. So, if you see something like $\prod x_i$ it means ‘multiply all of the scores you’ve collected’.
$\sqrt{x}$	This means ‘take the square root of $x$ ’.

## Greek symbols

$\alpha$	Alpha, the probability of making a Type I error
$\beta$	Beta, the probability of making a Type II error
$\beta_i$	Beta, the standardized regression coefficient
$\varepsilon$	Epsilon, usually stands for ‘error’, but is also used to denote sphericity
$\eta^2$	Eta squared, an effect size measure
$\mu$	Mu, the mean of a population of scores
$\rho$	Rho, the correlation in the population; also used to denote Spearman’s correlation coefficient
$\sigma$	Sigma, the standard deviation in a population of data
$\sigma^2$	Sigma squared, the variance in a population of data
$\sigma_{\bar{x}}$	Another variation on sigma, which represents the standard error of the mean
$\tau$	Kendall’s tau (non-parametric correlation coefficient)
$\phi$	Phi, a measure of association between two categorical variables, but also used to denote the dispersion parameter in logistic regression
$\chi^2$	Chi-square, a test statistic that quantifies the association between two categorical variables
$\chi_F^2$	Another use of the letter chi, but this time as the test statistic in Friedman’s ANOVA, a non-parametric test of differences between related means
$\omega^2$	Omega squared (an effect size measure). This symbol also means ‘expel the contents of your intestine immediately into your trousers’; you will understand why in due course

## Latin symbols

$b_i$	The regression coefficient (unstandardized); I tend to use it for any coefficient in a linear model
$df$	Degrees of freedom
$e_i$	The error associated with the $i$ th person
$F$	$F$ -statistic
$H$	Kruskal–Wallis test statistic
$k$	The number of levels of a variable (i.e., the number of treatment conditions), or the number of predictors in a regression model
$\ln$	Natural logarithm
$MS$	The mean squared error (mean square): the average variability in the data
$N, n, n_i$	The sample size. $N$ usually denotes the total sample size, whereas $n$ usually denotes the size of a particular group
$P$	Probability (the probability value, $p$ -value or significance of a test are usually denoted by $p$ )
$r$	Pearson's correlation coefficient
$r_s$	Spearman's rank correlation coefficient
$r_b, r_{pb}$	Biserial correlation coefficient and point-biserial correlation coefficient, respectively
$R$	The multiple correlation coefficient
$R^2$	The coefficient of determination (i.e., the proportion of data explained by the model)
$s$	The standard deviation of a sample of data
$s^2$	The variance of a sample of data
$SS$	The sum of squares, or sum of squared errors to give it its full title
$SS_A$	The sum of squares for variable $A$
$SS_M$	The model sum of squares (i.e., the variability explained by the model fitted to the data)
$SS_R$	The residual sum of squares (i.e., the variability that the model can't explain – the error in the model)
$SS_T$	The total sum of squares (i.e., the total variability within the data)
$t$	Testicle statistic for a $t$ -test. Yes, I did that deliberately to check whether you're paying attention
$T$	Test statistic for Wilcoxon's matched-pairs signed-rank test
$U$	Test statistic for the Mann–Whitney test
$W_s$	Test statistic for Rilcoxon's wank-sum test. See what I did there? It doesn't matter because no one reads this page
$\bar{X}$	The mean of a sample of scores
$z$	A data point expressed in standard deviation units

# SOME MATHS REVISION

There are good websites that can help you if any of the maths in this book confuses you. The pages at [study-maths.co.uk](http://study-maths.co.uk), [www.gcflernfree.org/math](http://www.gcflernfree.org/math), and [www.mathsisfun.com](http://www.mathsisfun.com) look useful, but there are many others, so use a search engine to find something that suits you. Some resources are also available on the book's website so you can try there if you run out of inspiration. I will quickly remind you of three important things:

**Two negatives make a positive:** Although in life two wrongs don't make a right, in mathematics they do. When we multiply a negative number by another negative number, the result is a positive number. For example,  $-2 \times -4 = 8$ .

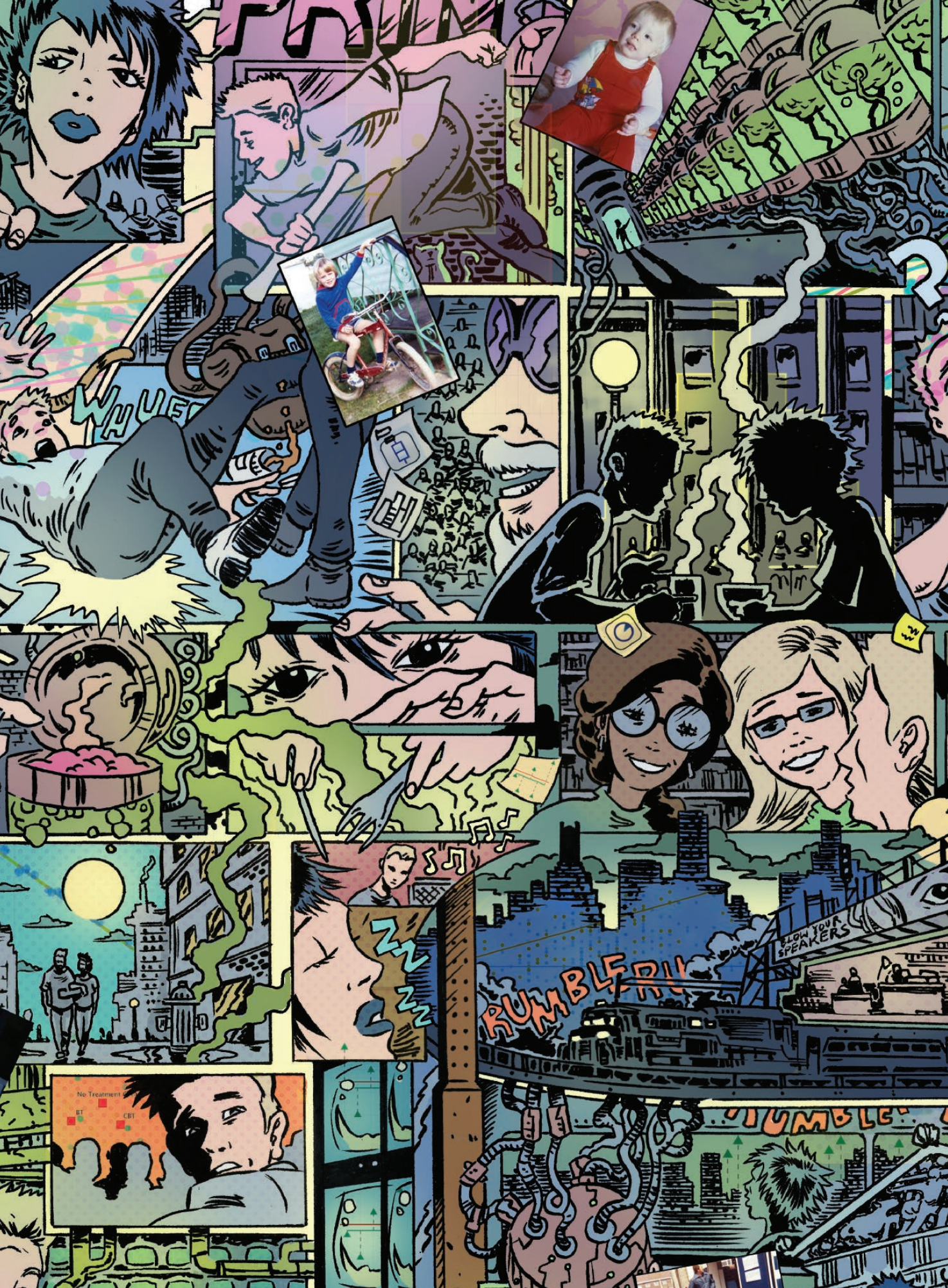
**A negative number multiplied by a positive one make a negative number:** If you multiply a positive number by a negative number then the result is another negative number. For example,  $2 \times -4 = -8$ , or  $-2 \times 6 = -12$ .

**BODMAS and PEMDAS:** These two acronyms are different ways of remembering the order in which mathematical operations are performed. BODMAS stands for Brackets, Order, Division, Multiplication, Addition, and Subtraction; whereas PEMDAS stems from Parentheses, Exponents, Multiplication, Division, Addition, and Subtraction. Having two widely used acronyms is confusing (especially because multiplication and division are the opposite way around), but they do mean the same thing:

- Brackets/Parentheses: When solving any expression or equation you deal with anything in brackets/parentheses first.
- Order/Exponents: Having dealt with anything in brackets, you next deal with any order terms/exponents. These refer to power terms such as squares. Four squared, or  $4^2$ , used to be called four raised to the order of 2, hence the word 'order' in BODMAS. These days, the term 'exponents' is more common (so by all means use BEDMAS as your acronym if you find that easier).
- Division and Multiplication: The next things to evaluate are any division or multiplication terms. The order in which you handle them is from the left to the right of the expression/equation. That's why BODMAS and PEMDAS can list them the opposite way around, because they are considered at the same time (so, BOMDAS or PEDMAS would work as acronyms, too).
- Addition and Subtraction: Finally, deal with any addition or subtraction. Again, go from left to right, doing any addition or subtraction in the order that you meet the terms. (So, BODMSA would work as an acronym too, but it's hard to say.)

Let's look at an example of BODMAS/PEMDAS in action: what would be the result of  $1 + 3 \times 5^2$ ? The answer is 76 (not 100 as some of you might have thought). There are no brackets, so the first thing is to deal with the order/exponent:  $5^2$  is 25, so the equation becomes  $1 + 3 \times 25$ . Moving from left to right, there is no division, so we do the multiplication:  $3 \times 25$ , which gives us 75. Again, going from left to right, we look for addition and subtraction terms – there are no subtractions, so the first thing we come across is the addition:  $1 + 75$ , which gives us 76 and the expression is solved. If I'd written the expression as  $(1 + 3) \times 5^2$ , then the answer would have been 100 because we deal with the brackets first:  $(1 + 3) = 4$ , so the expression becomes  $4 \times 5^2$ . We then deal with the order/exponent ( $5^2$  is 25), which results in  $4 \times 25 = 100$ .









# WHY IS MY EVIL LECTURER FORCING ME TO LEARN STATISTICS?

- 1.1 What will this chapter tell me? 2
- 1.2 What the hell am I doing here? I don't belong here 3
- 1.3 The research process 3
- 1.4 Initial observation: finding something that needs explaining 4
- 1.5 Generating and testing theories and hypotheses 5
- 1.6 Collecting data: measurement 9
- 1.7 Collecting data: research design 16
- 1.8 Analysing data 22
- 1.9 Reporting data 40
- 1.10 Brian's attempt to woo Jane 44
- 1.11 What next? 44
- 1.12 Key terms that I've discovered 44
  - Smart Alex's tasks 45

## 1.1 What will this chapter tell me?

I was born on 21 June 1973. Like most people, I don't remember anything about the first few years of life and, like most children, I went through a phase of driving my dad mad by asking 'Why?' every five seconds. With every question, the word 'dad' got longer and whinier: 'Dad, why is the sky blue?', 'Daaad, why don't worms have legs?', 'Daaaaaaaad, where do babies come from?' Eventually, my dad could take no more and whacked me around the face with a golf club.<sup>1</sup>

My torrent of questions reflected the natural curiosity that children have: we all begin our voyage through life as inquisitive little scientists. At the age of 3, I was at my friend Obe's party (just before he left England to return to Nigeria, much to my distress). It was a hot day, and there was an electric fan blowing cold air around the room. My 'curious little scientist' brain was working through what seemed like a particularly pressing question: 'What happens when you stick your finger in a fan?' The answer, as it turned out, was that it hurts – a lot.<sup>2</sup> At the age of 3, we intuitively know that to answer questions you need to collect data, even if it causes us pain.

My curiosity to explain the world never went away, which is why I'm a scientist. The fact that you're reading this book means that the inquisitive 3-year-old in you is alive and well and wants to answer new and exciting questions, too. To answer these questions you need 'science' and science has a **pilot fish** called 'statistics' that hides under its belly eating ectoparasites. That's why your evil lecturer is forcing you to learn statistics. Statistics is a bit like sticking your finger into a revolving fan blade: sometimes it's very painful, but it does give you answers to interesting questions. I'm going to try to convince you in this chapter that statistics are an important part of doing research. We will overview the whole research process, from why we conduct research in the first place, through how theories are generated, to why we need data to test these theories. If that doesn't convince you to read on then maybe the fact that we discover whether Coca-Cola kills sperm will. Or perhaps not.



**Figure 1.1** When I grow up, please don't let me be a statistics lecturer

- 1 He was practising in the garden when I unexpectedly wandered behind him at the exact moment he took a back swing. It's rare that a parent enjoys the sound of their child crying, but, on this day, it filled my dad with joy because my wailing was tangible evidence that he hadn't killed me, which he thought he might have done. Had he hit me with the club end rather than the shaft he probably would have. Fortunately (for me, but not for you), I survived, although some might argue that this incident explains the way my brain functions.
- 2 In the 1970s, fans didn't have helpful protective cages around them to prevent idiotic 3-year-olds sticking their fingers into the blades.



## 1.2 What the hell am I doing here? I don't belong here ■■■■

You're probably wondering why you have bought this book. Maybe you liked the pictures, maybe you fancied doing some weight training (it *is* heavy), or perhaps you needed to reach something in a high place (it *is* thick). The chances are, though, that given the choice of spending your hard-earned cash on a statistics book or something more entertaining (a nice novel, a trip to the cinema, etc.), you'd choose the latter. So, why have you bought the book (or downloaded an illegal PDF of it from someone who has way too much time on their hands if they're scanning 900 pages for fun)? It's likely that you obtained it because you're doing a course on statistics, or you're doing some research, and you need to know how to analyse data. It's possible that you didn't realize when you started your course or research that you'd have to know about statistics but now find yourself inexplicably wading, neck high, through the Victorian sewer that is data analysis. The reason why you're in the mess that you find yourself in is that you have a curious mind. You might have asked yourself questions like why people behave the way they do (psychology) or why behaviours differ across cultures (anthropology), how businesses maximize their profit (business), how the dinosaurs died (palaeontology), whether eating tomatoes protects you against cancer (medicine, biology), whether it is possible to build a quantum computer (physics, chemistry), whether the planet is hotter than it used to be and in what regions (geography, environmental studies). Whatever it is you're studying or researching, the reason why you're studying it is probably that you're interested in answering questions. Scientists are curious people, and you probably are too. However, it might not have occurred to you that to answer interesting questions, you need data and explanations for those data.

The answer to 'What the hell are you doing here?' is simple: to answer interesting questions you need data. One of the reasons why your evil statistics lecturer is forcing you to learn about numbers is that they are a form of data and are vital to the research process. Of course, there are forms of data other than numbers that can be used to test and generate theories. When numbers are involved, the research involves **quantitative methods**, but you can also generate and test theories by analysing language (such as conversations, magazine articles and media broadcasts). This involves **qualitative methods** and it is a topic for another book not written by me. People can get quite passionate about which of these methods is *best*, which is a bit silly because they are complementary, not competing, approaches and there are much more important issues in the world to get upset about. Having said that, all qualitative research is rubbish.<sup>3</sup>

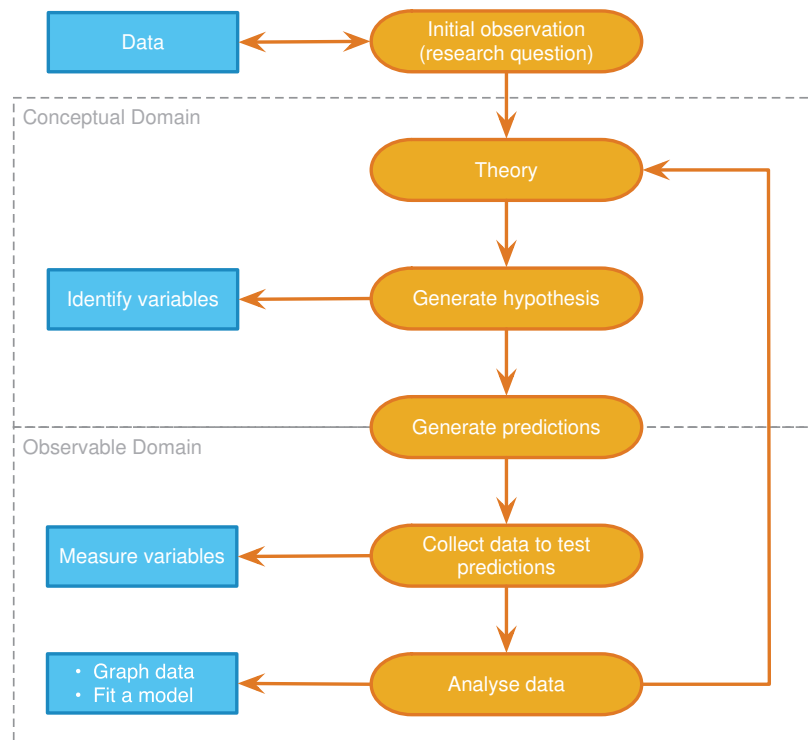
## 1.3 The research process ■■■■

How do you go about answering an interesting question? The research process is broadly summarized in Figure 1.2. You begin with an observation that you want to understand, and this observation could be anecdotal (you've noticed that your cat watches birds when they're on TV but not when jellyfish are on)<sup>4</sup> or could be based on some data (you've got several cat owners to keep diaries of their cat's TV habits and noticed that lots of them watch birds). From your initial observation you consult relevant theories and generate explanations (hypotheses) for those observations, from which you can make predictions. To



<sup>3</sup> This is a joke. Like many of my jokes, there are people who won't find it remotely funny. Passions run high between qualitative and quantitative researchers, so its inclusion will likely result in me being hunted down, locked in a room and forced to do discourse analysis by a horde of rabid qualitative researchers.

<sup>4</sup> In his younger days my cat actually did climb up and stare at the TV when birds were being shown.



**Figure 1.2** The research process

test your predictions you need data. First you collect some relevant data (and to do that you need to identify things that can be measured) and then you analyse those data. The analysis of the data may support your hypothesis, or generate a new one, which, in turn, might lead you to revise the theory. As such, the processes of data collection and analysis and generating theories are intrinsically linked: theories lead to data collection/analysis and data collection/analysis informs theories. This chapter explains this research process in more detail.

## 1.4 Initial observation: finding something that needs explaining ■■■■

The first step in Figure 1.2 was to come up with a question that needs an answer. I spend rather more time than I should watching reality TV. Over many years, I used to swear that I wouldn't get hooked on reality TV, and yet year upon year I would find myself glued to the TV screen waiting for the next contestant's meltdown (I am a psychologist, so really this is just research). I used to wonder why there is so much arguing in these shows, and why so many contestants have really unpleasant personalities (my money is on narcissistic personality disorder).<sup>5</sup> A lot of scientific endeavour starts this way: not by watching reality TV, but by observing something in the world and wondering why it happens.

Having made a casual observation about the world (reality TV contestants on the whole have extreme personalities and argue a lot), I need to collect some data to see whether this observation is true (and not a biased observation). To do this, I need to define one or more **variables** to measure that quantify the thing

<sup>5</sup> This disorder is characterized by (among other things) a grandiose sense of self-importance, arrogance, lack of empathy for others, envy of others and belief that others envy them, excessive fantasies of brilliance or beauty, the need for excessive admiration, and exploitation of others.

I'm trying to measure. There's one variable in this example: the personality of the contestant. I could measure this variable by giving them one of the many well-established questionnaires that measure personality characteristics. Let's say that I did this and I found that 75% of contestants did have narcissistic personality disorder. These data support my observation: a lot of reality TV contestants have extreme personalities.

## 1.5 Generating and testing theories and hypotheses

The next logical thing to do is to explain these data (Figure 1.2). The first step is to look for relevant theories. A **theory** is an explanation or set of principles that is well substantiated by repeated testing and explains a broad phenomenon. We might begin by looking at theories of narcissistic personality disorder, of which there are currently very few. One theory of personality disorders in general links them to early attachment (put simplistically, the bond formed between a child and their main caregiver). Broadly speaking, a child can form a secure (a good thing) or an insecure (not so good) attachment to their caregiver, and the theory goes that insecure attachment explains later personality disorders (Levy, Johnson, Clouthier, Scala, & Temes, 2015). This is a theory because it is a set of principles (early problems in forming interpersonal bonds) that explains a general broad phenomenon (disorders characterized by dysfunctional interpersonal relations). There is also a critical mass of evidence to support the idea. Theory also tells us that those with narcissistic personality disorder tend to engage in conflict with others despite craving their attention, which perhaps explains their difficulty in forming close bonds.

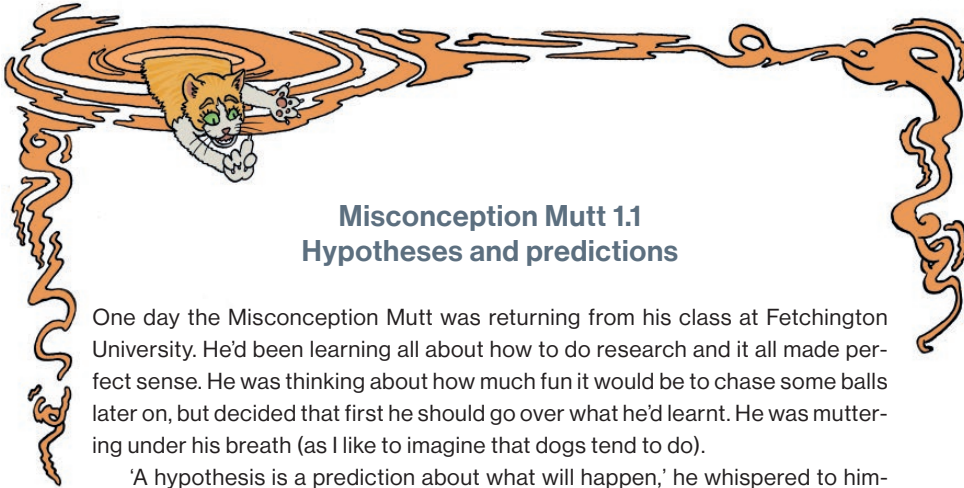
Given this theory, we might generate a **hypothesis** about our earlier observation (see Jane Superbrain Box 1.1). A hypothesis is a proposed explanation for a fairly narrow phenomenon or set of observations. It is not a guess, but an informed, theory-driven attempt to explain what has been observed. Both theories and hypotheses seek to explain the world, but a theory explains a wide set of phenomena with a small set of well-established principles, whereas a hypothesis typically seeks to explain a narrower phenomenon and is, as yet, untested. Both theories and hypotheses exist in the conceptual domain, and you cannot observe them directly.

To continue the example, having studied the attachment theory of personality disorders, we might decide that this theory implies that people with personality disorders seek out the attention that a TV appearance provides because they lack close interpersonal relationships. From this we can generate a hypothesis: people with narcissistic personality disorder use reality TV to satisfy their craving for attention from others. This is a conceptual statement that explains our original observation (that rates of narcissistic personality disorder are high on reality TV shows).

To test this hypothesis, we need to move from the conceptual domain into the observable domain. That is, we need to operationalize our hypothesis in a way that enables us to collect and analyse data that have a bearing on the hypothesis (Figure 1.2). We do this using predictions. Predictions emerge from a hypothesis (Misconception Mutt 1.1), and transform it from something unobservable into something that is. If our hypothesis is that people with narcissistic personality disorder use reality TV to satisfy their craving for attention from others, then a prediction we could make based on this hypothesis is that people with narcissistic personality disorder are more likely to audition for reality TV than those without. In making this prediction we can move from the conceptual domain into the observable domain, where we can collect evidence.

In this example, our prediction is that people with narcissistic personality disorder are more likely to audition for reality TV than those without. We can measure this prediction by getting a team of clinical psychologists to interview each person at a reality TV audition and diagnose them as having narcissistic personality disorder or not. The population rates of narcissistic personality disorder are





### Misconception Mutt 1.1

#### Hypotheses and predictions

One day the Misconception Mutt was returning from his class at Fetchington University. He'd been learning all about how to do research and it all made perfect sense. He was thinking about how much fun it would be to chase some balls later on, but decided that first he should go over what he'd learnt. He was muttering under his breath (as I like to imagine that dogs tend to do).

'A hypothesis is a prediction about what will happen,' he whispered to himself in his deep, wheezy, jowly dog voice. Before he could finish, the ground before him became viscous, as though the earth had transformed into liquid. A slightly irritated-looking ginger cat rose slowly from the puddle.

'Don't even think about chasing me,' he said in his whiny cat voice.

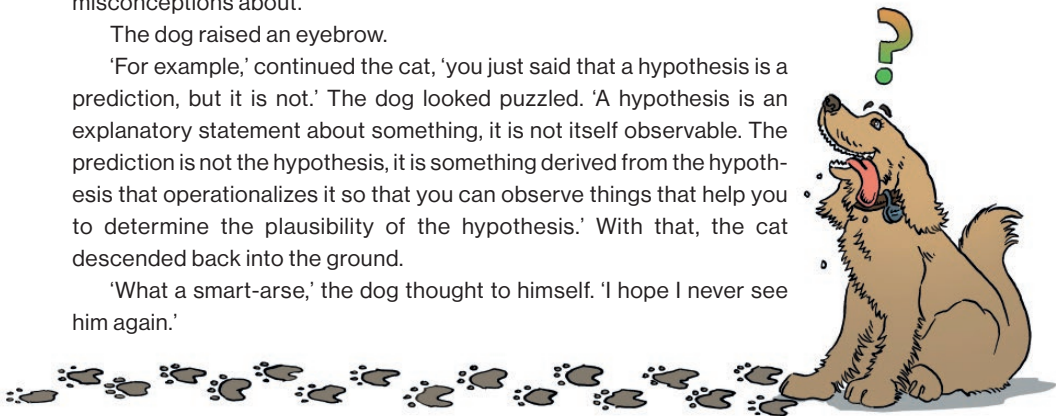
The mutt twitched as he inhibited the urge to chase the cat. 'Who are you?' he asked.

'I am the Correcting Cat,' said the cat wearily. 'I travel the ether trying to correct people's statistical misconceptions. It's very hard work, there are a lot of misconceptions about.'

The dog raised an eyebrow.

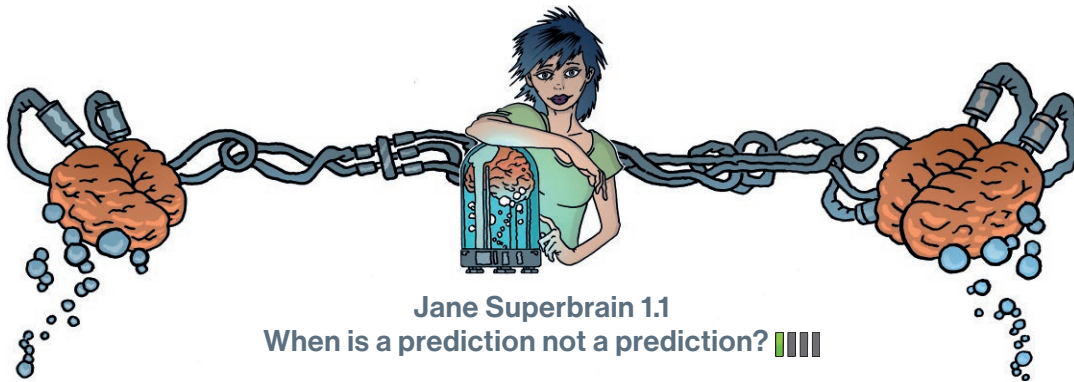
'For example,' continued the cat, 'you just said that a hypothesis is a prediction, but it is not.' The dog looked puzzled. 'A hypothesis is an explanatory statement about something, it is not itself observable. The prediction is not the hypothesis, it is something derived from the hypothesis that operationalizes it so that you can observe things that help you to determine the plausibility of the hypothesis.' With that, the cat descended back into the ground.

'What a smart-arse,' the dog thought to himself. 'I hope I never see him again.'

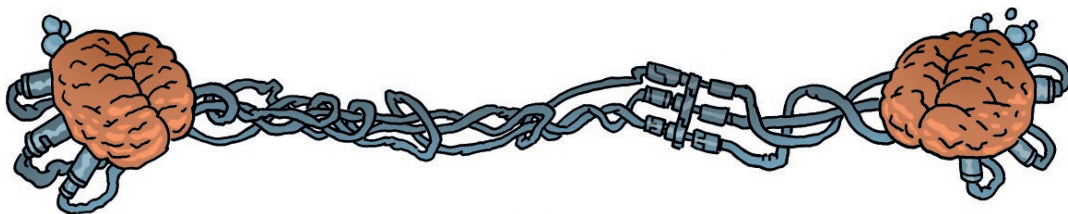


about 1%, so we'd be able to see whether the ratio of narcissistic personality disorder to not is higher at the audition than in the general population. If it is higher then our prediction is correct: a disproportionate number of people with narcissistic personality disorder turned up at the audition. Our prediction, in turn, tells us something about the hypothesis from which it derived.

This is tricky stuff, so let's look at another example. Imagine that, based on a different theory, we generated a different hypothesis. I mentioned earlier that people with narcissistic personality disorder tend to engage in conflict, so a different hypothesis is that producers of reality TV shows select people who have narcissistic personality disorder to be contestants because they believe that conflict makes good TV. As before, to test this hypothesis we need to bring it into the observable domain by generating a prediction from it. The prediction would be that (assuming no bias in the number of people with narcissistic personality disorder applying for the show) a disproportionate number of people with narcissistic personality disorder will be selected by producers to go on the show.



A good theory should allow us to make statements about the state of the world. Statements about the world are good things: they allow us to make sense of our world, and to make decisions that affect our future. One current example is global warming. Being able to make a definitive statement that global warming is happening, and that it is caused by certain practices in society, allows us to change these practices and, hopefully, avert catastrophe. However, not all statements can be tested using science. Scientific statements are ones that can be verified with reference to empirical evidence, whereas non-scientific statements are ones that cannot be empirically tested. So, statements such as 'The Led Zeppelin reunion concert in London in 2007 was the best gig ever',<sup>6</sup> 'Lindt chocolate is the best food' and 'This is the worst statistics book in the world' are all non-scientific; they cannot be proved or disproved. Scientific statements can be confirmed or disconfirmed empirically. 'Watching *Curb Your Enthusiasm* makes you happy,' 'Having sex increases levels of the neurotransmitter dopamine' and 'Velociraptors ate meat' are all things that can be tested empirically (provided you can quantify and measure the variables concerned). Non-scientific statements can sometimes be altered to become scientific statements, so 'The Beatles were the most influential band ever' is non-scientific (because it is probably impossible to quantify 'influence' in any meaningful way) but by changing the statement to 'The Beatles were the best-selling band ever,' it becomes testable (we can collect data about worldwide album sales and establish whether the Beatles have, in fact, sold more records than any other music artist). Karl Popper, the famous philosopher of science, believed that non-scientific statements were nonsense and had no place in science. Good theories and hypotheses should, therefore, produce predictions that are scientific statements.



Imagine we collected the data in Table 1.1, which shows how many people auditioning to be on a reality TV show had narcissistic personality disorder or not. In total, 7662 people turned up for the audition. Our first prediction (derived from our first hypothesis) was that the percentage of people with narcissistic personality disorder will be higher at the audition than the general level in the population. We can see in the table that of the 7662 people at the audition, 854 were diagnosed with the disorder; this is about 11% ( $854/7662 \times 100$ ), which is much higher than the 1% we'd expect in the general

**Table 1.1** The number of people at the TV audition split by whether they had narcissistic personality disorder and whether they were selected as contestants by the producers

	No Disorder	Disorder	Total
Selected	3	9	12
Rejected	6805	845	7650
Total	6808	854	7662

<sup>6</sup> It was pretty awesome actually.

population. Therefore, prediction 1 is correct, which in turn supports hypothesis 1. The second prediction was that the producers of reality TV have a bias towards choosing people with narcissistic personality disorder. If we look at the 12 contestants that they selected, 9 of them had the disorder (a massive 75%). If the producers did not have a bias we would have expected only 11% of the contestants to have the disorder (the same rate as was found when we considered everyone who turned up for the audition). The data are in line with prediction 2 which supports our second hypothesis. Therefore, my initial observation that contestants have personality disorders was verified by data, and then using theory I generated specific hypotheses that were operationalized by generating predictions that could be tested using data. Data are *very* important.

I would now be smugly sitting in my office with a contented grin on my face because my hypotheses were well supported by the data. Perhaps I would quit while I was ahead and retire. It's more likely, though, that having solved one great mystery, my excited mind would turn to another. I would lock myself in a room to watch more reality TV. I might wonder at why contestants with narcissistic personality disorder, despite their obvious character flaws, enter a situation that will put them under intense public scrutiny.<sup>7</sup> Days later, the door would open, and a stale odour would waft out like steam rising from the New York subway. Through this green cloud, my bearded face would emerge, my eyes squinting at the shards of light that cut into my pupils. Stumbling forwards, I would open my mouth to lay waste to my scientific rivals with my latest profound hypothesis: 'Contestants with narcissistic personality disorder believe that they will win'. I would croak before collapsing on the floor. The prediction from this hypothesis is that if I ask the contestants if they think that they will win, the people with a personality disorder will say 'yes'.

Let's imagine I tested my hypothesis by measuring contestants' expectations of success in the show, by asking them, 'Do you think you will win?' Let's say that 7 of 9 contestants with narcissistic personality disorder said that they thought that they would win, which confirms my hypothesis. At this point I might start to try to bring my hypotheses together into a theory of reality TV contestants that revolves around the idea that people with narcissistic personalities are drawn towards this kind of show because it fulfils their need for approval and they have unrealistic expectations about their likely success because they don't realize how unpleasant their personalities are to other people. In parallel, producers tend to select contestants with narcissistic tendencies because they tend to generate interpersonal conflict.

One part of my theory is untested, which is the bit about contestants with narcissistic personalities not realizing how others perceive their personality. I could operationalize this hypothesis through a prediction that if I ask these contestants whether their personalities were different from those of other people they would say 'no'. As before, I would collect more data and ask the contestants with narcissistic personality disorder whether they believed that their personalities were different from the norm. Imagine that all 9 of them said that they thought their personalities *were* different from the norm. These data contradict my hypothesis. This is known as **falsification**, which is the act of disproving a hypothesis or theory.

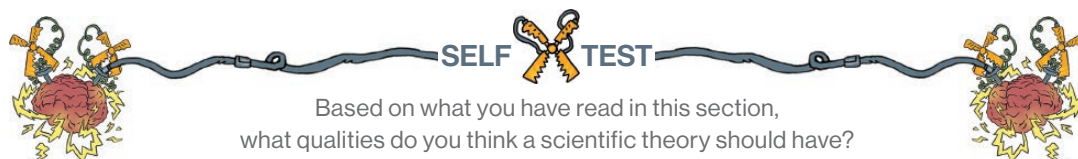
It's unlikely that we would be the only people interested in why individuals who go on reality TV have extreme personalities. Imagine that these other researchers discovered that: (1) people with narcissistic personality disorder think that they are more interesting than others; (2) they also think that they deserve success more than others; and (3) they also think that others like them because they have 'special' personalities.

This additional research is even worse news for my theory: if contestants didn't realize that they had a personality different from the norm, then you wouldn't expect them to think that they were more interesting than others, and you certainly wouldn't expect them to think that others will *like* their unusual personalities. In general, this means that this part of my theory sucks: it cannot explain all of the data,

<sup>7</sup> One of the things I like about many reality TV shows in the UK is that the winners are very often nice people, and the odious people tend to get voted out quickly, which gives me faith that humanity favours the nice.

predictions from the theory are not supported by subsequent data, and it cannot explain other research findings. At this point I would start to feel intellectually inadequate and people would find me curled up on my desk in floods of tears, wailing and moaning about my failing career (no change there then).

At this point, a rival scientist, Fester Ingplant-Stain, appears on the scene adapting my theory to suggest that the problem is not that personality-disordered contestants don't realize that they have a personality disorder (or at least a personality that is unusual), but that they falsely believe that this special personality is perceived positively by other people. One prediction from this model is that if personality-disordered contestants are asked to evaluate what other people think of them, then they will overestimate other people's positive perceptions. You guessed it, Fester Ingplant-Stain collected yet more data. He asked each contestant to fill out a questionnaire evaluating all of the other contestants' personalities, and also to complete the questionnaire about themselves but answering from the perspective of each of their housemates. (So, for every contestant there is a measure of what they thought of every other contestant, and also a measure of what they believed every other contestant thought of them.) He found out that the contestants with personality disorders did overestimate their housemates' opinions of them; conversely, the contestants without personality disorders had relatively accurate impressions of what others thought of them. These data, irritating as it would be for me, support Fester Ingplant-Stain's theory more than mine: contestants with personality disorders do realize that they have unusual personalities but believe that these characteristics are ones that others would feel positive about. Fester Ingplant-Stain's theory is quite good: it explains the initial observations and brings together a range of research findings. The end result of this whole process (and my career) is that we should be able to make a general statement about the state of the world. In this case we could state 'Reality TV contestants who have personality disorders overestimate how much other people like their personality characteristics'.



## 1.6 Collecting data: measurement ■■■

In looking at the process of generating theories and hypotheses, we have seen the importance of data in testing those hypotheses or deciding between competing theories. This section looks at data collection in more detail. First we'll look at measurement.

### 1.6.1 Independent and dependent variables ■■■

To test hypotheses we need to measure variables. Variables are things that can change (or vary); they might vary between people (e.g., IQ, behaviour) or locations (e.g., unemployment) or even time (e.g., mood, profit, number of cancerous cells). Most hypotheses can be expressed in terms of two variables: a proposed cause and a proposed outcome. For example, if we take the scientific statement, 'Coca-Cola is an effective spermicide'<sup>8</sup> then the proposed cause is 'Coca-Cola' and the proposed effect is dead

<sup>8</sup> Actually, there is a long-standing urban myth that a post-coital douche with the contents of a bottle of Coke is an effective contraceptive. Unbelievably, this hypothesis has been tested and Coke does affect sperm motility (movement), and some types of Coke are more effective than others – Diet Coke is best, apparently (Umpierre, Hill & Anderson, 1985). In case you decide to try this method out, I feel it worth mentioning that despite the effects on sperm motility a Coke douche is ineffective at preventing pregnancy.





## Cramming Sam's Tips Variables

When doing and reading research you're likely to encounter these terms:

- *Independent variable*: A variable thought to be the cause of some effect. This term is usually used in experimental research to describe a variable that the experimenter has manipulated.
- *Dependent variable*: A variable thought to be affected by changes in an independent variable. You can think of this variable as an outcome.
- *Predictor variable*: A variable thought to predict an outcome variable. This term is basically another way of saying 'independent variable'. (Although some people won't like me saying that; I think life would be easier if we talked only about predictors and outcomes.)
- *Outcome variable*: A variable thought to change as a function of changes in a predictor variable. For the sake of an easy life this term could be synonymous with 'dependent variable'.



sperm. Both the cause and the outcome are variables: for the cause we could vary the type of drink, and for the outcome, these drinks will kill different amounts of sperm. The key to testing scientific statements is to measure these two variables.

A variable that we think is a cause is known as an **independent variable** (because its value does not depend on any other variables). A variable that we think is an effect is called a **dependent variable** because the value of this variable depends on the cause (independent variable). These terms are very closely tied to experimental methods in which the cause is manipulated by the experimenter (as we will see in Section 1.7.2). However, researchers can't always manipulate variables (for example, if you wanted see whether smoking causes lung cancer you wouldn't lock a bunch of people in a room for 30 years and force them to smoke). Instead, they sometimes use correlational methods (Section 1.7), for which it doesn't make sense to talk of dependent and independent variables because all variables are essentially dependent variables. I prefer to use the terms **predictor variable** and **outcome variable** in place of dependent and independent variable. This is not a personal whimsy: in experimental work the cause (independent variable) is a predictor, and the effect (dependent variable) is an outcome, and in correlational work we can talk of one or more (predictor) variables predicting (statistically at least) one or more outcome variables.

## 1.6.2 Levels of measurement

Variables can take on many different forms and levels of sophistication. The relationship between what is being measured and the numbers that represent what is being measured is known as the **level of measurement**. Broadly speaking, variables can be categorical or continuous, and can have different levels of measurement.

A **categorical variable** is made up of categories. A categorical variable that you should be familiar with already is your species (e.g., human, domestic cat, fruit bat, etc.). You are a human or a cat or a fruit bat: you cannot be a bit of a cat and a bit of a bat, and neither a batman nor (despite many fantasies to the contrary) a catwoman exist (not even one in a PVC suit). A categorical variable is one that names distinct entities. In its simplest form it names just two distinct types of things, for example male or female. This is known as a **binary variable**. Other examples of binary variables are being alive or dead, pregnant or not, and responding 'yes' or 'no' to a question. In all cases there are just two categories and an entity can be placed into only one of the two categories. When two things that are equivalent in some sense are given the same name (or number), but there are more than two possibilities, the variable is said to be a **nominal variable**.

It should be obvious that if the variable is made up of names it is pointless to do arithmetic on them (if you multiply a human by a cat, you do not get a hat). However, sometimes numbers are used to denote categories. For example, the numbers worn by players in a sports team. In rugby, the numbers on shirts denote specific field positions, so the number 10 is always worn by the fly-half<sup>9</sup> and the number 2 is always the hooker (the ugly-looking player at the front of the scrum). These numbers do not tell us anything other than what position the player plays. We could equally have shirts with FH and H instead of 10 and 2. A number 10 player is not necessarily better than a number 2 (most managers would not want their fly-half stuck in the front of the scrum!). It is equally daft to try to do arithmetic with nominal scales where the categories are denoted by numbers: the number 10 takes penalty kicks, and if the coach found that his number 10 was injured, he would not get his number 4 to give number 6



A lot of self-report data are ordinal. Imagine two judges on *The X Factor* were asked to rate Billie's singing on a 10-point scale. We might be confident that a judge who gives a rating of 10 found Billie more talented than one who gave a rating of 2, but can we be certain that the first judge found her five times more talented than the second? What if both judges gave a rating of 8; could we be sure that they found her equally talented? Probably not: their ratings will depend on their subjective feelings about what constitutes talent (the quality of singing? showmanship? dancing?). For these reasons, in any situation in which we ask people to rate something subjective (e.g., their preference for a product, their confidence about an answer, how much they have understood some medical instructions) we should probably regard these data as ordinal, although many scientists do not.



9 Unlike, for example, NFL football where a quarterback could wear any number from 1 to 19.



a piggy-back and then take the kick. The only way that nominal data can be used is to consider frequencies. For example, we could look at how frequently number 10s score compared to number 4s.

So far, the categorical variables we have considered have been unordered (e.g., different brands of Coke with which you're trying to kill sperm), but they can be ordered too (e.g., increasing concentrations of Coke with which you're trying to skill sperm). When categories are ordered, the variable is known as an **ordinal variable**. Ordinal data tell us not only that things have occurred, but also the order in which they occurred. However, these data tell us nothing about the differences between values. In TV shows like *The X Factor*, *American Idol*, and *The Voice*, hopeful singers compete to win a recording contract. They are hugely popular shows, which could (if you take a depressing view) reflect the fact that Western society values 'luck' more than hard work.<sup>10</sup> Imagine that the three winners of a particular *X Factor* series were Billie, Freema and Elizabeth. The names of the winners don't provide any information about where they came in the contest; however, labelling them according to their performance does – first, second and third. These categories are ordered. In using ordered categories we now know that the woman who won was better than the women who came second and third. We still know nothing about the differences between categories, though. We don't, for example, know how much better the winner was than the runners-up: Billie might have been an easy victor, getting many more votes than Freema and Elizabeth, or it might have been a very close contest that she won by only a single vote. Ordinal data, therefore, tell us more than nominal data (they tell us the order in which things happened) but they still do not tell us about the differences between points on a scale.

The next level of measurement moves us away from categorical variables and into continuous variables. A **continuous variable** is one that gives us a score for each person and can take on any value on the measurement scale that we are using. The first type of continuous variable that you might encounter is an **interval variable**. Interval data are considerably more useful than ordinal data, and most of the statistical tests in this book rely on having data measured at this level at least. To say that data are interval, we must be certain that equal intervals on the scale represent equal differences in the property being measured. For example, on [www.ratemyprofessors.com](http://www.ratemyprofessors.com), students are encouraged to rate their lecturers on several dimensions (some of the lecturers' rebuttals of their negative evaluations are worth a look). Each dimension (helpfulness, clarity, etc.) is evaluated using a 5-point scale. For this scale to be interval it must be the case that the difference between helpfulness ratings of 1 and 2 is the same as the difference between (say) 3 and 4, or 4 and 5. Similarly, the difference in helpfulness between ratings of 1 and 3 should be identical to the difference between ratings of 3 and 5. Variables like this that look interval (and are treated as interval) are often ordinal – see Jane Superbrain Box 1.2.

**Ratio variables** go a step further than interval data by requiring that in addition to the measurement scale meeting the requirements of an interval variable, the ratios of values along the scale should be meaningful. For this to be true, the scale must have a true and meaningful zero point. In our lecturer ratings this would mean that a lecturer rated as 4 would be twice as helpful as a lecturer rated with a 2 (who would, in turn, be twice as helpful as a lecturer rated as 1). The time to respond to something is a good example of a ratio variable. When we measure a reaction time, not only is it true that, say, the difference between 300 and 350 ms (a difference of 50 ms) is the same as the difference between 210 and 260 ms or between 422 and 472 ms, but it is also true that distances along the scale are divisible: a reaction time of 200 ms is twice as long as a reaction time of 100 ms and half as long as a reaction time of 400 ms. Time also has a meaningful zero point: 0 ms does mean a complete absence of time.

*Continuous variables* can be, well, continuous (obviously) but also discrete. This is quite a tricky distinction (Jane Superbrain Box 1.3). A truly continuous variable can be measured to any level of

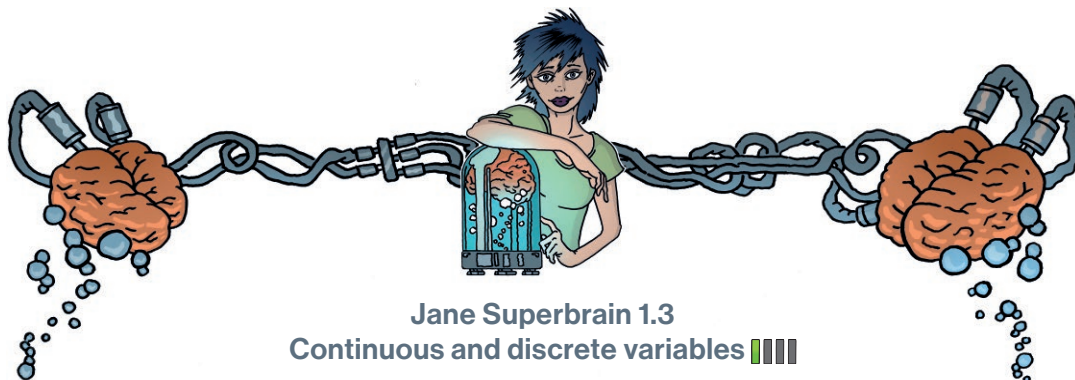
<sup>10</sup> I am in no way bitter about spending years learning musical instruments and trying to create original music, only to be beaten to musical fame and fortune by 15-year-olds who can sing, sort of.

precision, whereas a **discrete variable** can take on only certain values (usually whole numbers) on the scale. What does this actually mean? Well, our example of rating lecturers on a 5-point scale is an example of a discrete variable. The range of the scale is 1–5, but you can enter only values of 1, 2, 3, 4 or 5; you cannot enter a value of 4.32 or 2.18. Although a continuum exists underneath the scale (i.e., a rating of 3.24 makes sense), the actual values that the variable takes on are limited. A continuous variable would be something like age, which can be measured at an infinite level of precision (you could be 34 years, 7 months, 21 days, 10 hours, 55 minutes, 10 seconds, 100 milliseconds, 63 microseconds, 1 nanosecond old).

### 1.6.3 Measurement error

It's one thing to measure variables, but it's another thing to measure them accurately. Ideally we want our measure to be calibrated such that values have the same meaning over time and across situations. Weight is one example: we would expect to weigh the same amount regardless of who weighs us, or where we take the measurement (assuming it's on Earth and not in an anti-gravity chamber). Sometimes, variables can be measured directly (profit, weight, height) but in other cases we are forced to use indirect measures such as self-report, questionnaires, and computerized tasks (to name a few).

It's been a while since I mentioned sperm, so let's go back to our Coke as a spermicide example. Imagine we took some Coke and some water and added them to two test tubes of sperm. After several minutes, we measured the motility (movement) of the sperm in the two samples and discovered no difference. A few years passed, as you might expect given that Coke and sperm rarely top scientists' research lists, before another scientist, Dr Jack Q. Late, replicated the study. Dr Late found that sperm motility



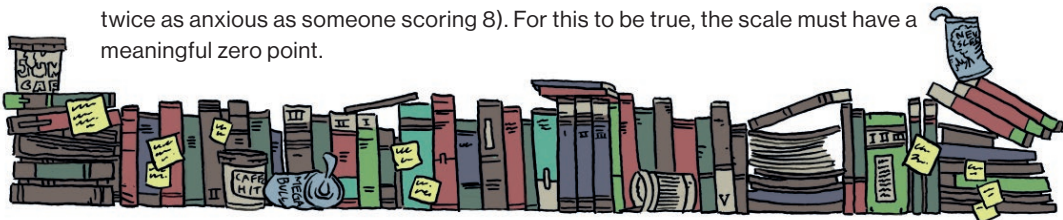
The distinction between continuous and discrete variables can be blurred. For one thing, continuous variables can be measured in discrete terms; for example, when we measure age we rarely use nanoseconds but use years (or possibly years and months). In doing so we turn a continuous variable into a discrete one (the only acceptable values are years). Also, we often treat discrete variables as if they were continuous. For example, the number of boyfriends/girlfriends that you have had is a discrete variable (it will be, in all but the very weirdest cases, a whole number). However, you might read a magazine that says 'The average number of boyfriends that women in their 20s have has increased from 4.6 to 8.9'. This assumes that the variable is continuous, and of course these averages are meaningless: no one in their sample actually had 8.9 boyfriends.





## Cramming Sam's Tips Levels of measurement

- Variables can be split into categorical and continuous, and within these types there are different levels of measurement:
- Categorical (entities are divided into distinct categories):
  - Binary variable: There are only two categories (e.g., dead or alive).
  - Nominal variable: There are more than two categories (e.g., whether someone is an omnivore, vegetarian, vegan, or fruitarian).
  - Ordinal variable: The same as a nominal variable but the categories have a logical order (e.g., whether people got a fail, a pass, a merit or a distinction in their exam).
- Continuous (entities get a distinct score):
  - Interval variable: Equal intervals on the variable represent equal differences in the property being measured (e.g., the difference between 6 and 8 is equivalent to the difference between 13 and 15).
  - Ratio variable: The same as an interval variable, but the ratios of scores on the scale must also make sense (e.g., a score of 16 on an anxiety scale means that the person is, in reality, twice as anxious as someone scoring 8). For this to be true, the scale must have a meaningful zero point.



was worse in the Coke sample. There are two measurement-related issues that could explain his success and our failure: (1) Dr Late might have used more Coke in the test tubes (sperm might need a critical mass of Coke before they are affected); (2) Dr Late measured the outcome (motility) differently than us.

The former point explains why chemists and physicists have devoted many hours to developing standard units of measurement. If you had reported that you'd used 100ml of Coke and 5ml of sperm, then Dr Late could have ensured that he had used the same amount – because millilitres are a standard unit of measurement – we would know that Dr Late used exactly the same amount of Coke that we used. Direct measurements such as the millilitre provide an objective standard: 100ml of a liquid is known to be twice as much as only 50ml.

The second reason for the difference in results between the studies could have been to do with how sperm motility was measured. Perhaps in our original study we measured motility using absorption spectrophotometry, whereas Dr Late used laser light-scattering techniques.<sup>11</sup> Perhaps his measure is more sensitive than ours.

<sup>11</sup> In the course of writing this chapter I have discovered more than I think is healthy about the measurement of sperm motility.

There will often be a discrepancy between the numbers we use to represent the thing we're measuring and the actual value of the thing we're measuring (i.e., the value we would get if we could measure it directly). This discrepancy is known as **measurement error**. For example, imagine that you know as an absolute truth that you weigh 83kg. One day you step on the bathroom scales and they read 80kg. There is a difference of 3kg between your actual weight and the weight given by your measurement tool (the scales): this is a measurement error of 3kg. Although properly calibrated bathroom scales should produce only very small measurement errors (despite what we might want to believe when it says we have gained 3kg), self-report measures will produce larger measurement error because factors other than the one you're trying to measure will influence how people respond to our measures. For example, if you were completing a questionnaire that asked you whether you had stolen from a shop, would you admit it, or might you be tempted to conceal this fact?

### 1.6.4 Validity and reliability

One way to try to ensure that measurement error is kept to a minimum is to determine properties of the measure that give us confidence that it is doing its job properly. The first property is **validity**, which is whether an instrument measures what it sets out to measure. The second is **reliability**, which is whether an instrument can be interpreted consistently across different situations.

Validity refers to whether an instrument measures what it was designed to measure (e.g., does your lecturer helpfulness rating scale actually measure lecturers' helpfulness?); a device for measuring sperm *motility* that actually measures sperm *count* is not valid. Things like reaction times and physiological measures are valid in the sense that a reaction time does, in fact, measure the time taken to react and skin conductance does measure the conductivity of your skin. However, if we're using these things to infer other things (e.g., using skin conductance to measure anxiety), then they will be valid only if there are no other factors other than the one we're interested in that can influence them.

**Criterion validity** is whether you can establish that an instrument measures what it claims to measure through comparison to objective criteria. In an ideal world, you assess this by relating scores on your measure to real-world observations. For example, we could take an objective measure of how helpful lecturers were and compare these observations to students' ratings of helpfulness on [ratemyprofessor.com](http://ratemyprofessor.com). When data are recorded simultaneously using the new instrument and existing criteria, then this is said to assess **concurrent validity**; when data from the new instrument are used to predict observations at a later point in time, this is said to assess **predictive validity**.

Assessing criterion validity (whether concurrently or predictively) is often impractical because objective criteria that can be measured easily may not exist. Also, with measuring attitudes, you might be interested in the person's perception of reality and not reality itself (you might not care whether a person *is* a psychopath but whether they *think* they are a psychopath). With self-report measures/questionnaires we can also assess the degree to which individual items represent the construct being measured, and cover the full range of the construct (**content validity**).

Validity is a necessary but not sufficient condition of a measure. A second consideration is reliability, which is the ability of the measure to produce the same results under the same conditions. To be valid the instrument must first be reliable. The easiest way to assess reliability is to test the same group of people twice: a reliable instrument will produce similar scores at both points in time (**test-retest reliability**). Sometimes, however, you will want to measure something that does vary over time (e.g., moods, blood-sugar levels, productivity). Statistical methods can also be used to determine reliability (we will discover these in Chapter 18).

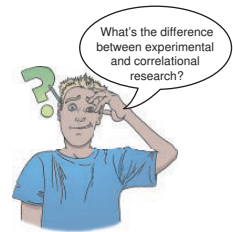


## 1.7 Collecting data: research design ■■■

We've looked at the question of *what* to measure and discovered that to answer scientific questions we measure variables (which can be collections of numbers or words). We also saw that to get accurate answers we need accurate measures. We move on now to look at research design: *how* data are collected. If we simplify things quite a lot then there are two ways to test a hypothesis: either by observing what naturally happens, or by manipulating some aspect of the environment and observing the effect it has on the variable that interests us. In **correlational** or **cross-sectional research** we observe what naturally goes on in the world without directly interfering with it, whereas in **experimental research** we manipulate one variable to see its effect on another.

### 1.7.1 Correlational research methods ■■■

In correlational research we observe natural events; we can do this by either taking a snapshot of many variables at a single point in time, or by measuring variables repeatedly at different time points (known as **longitudinal research**). For example, we might measure pollution levels in a stream and the numbers of certain types of fish living there; lifestyle variables (smoking, exercise, food intake) and disease (cancer, diabetes); workers' job satisfaction under different managers; or children's school performance across regions with different demographics. Correlational research provides a very natural view of the question we're researching because we're not influencing what happens and the measures of the variables should not be biased by the researcher being there (this is an important aspect of **ecological validity**).



At the risk of sounding like I'm absolutely obsessed with using Coke as a contraceptive (I'm not, but my discovery that people in the 1950s and 1960s actually tried this has, I admit, intrigued me), let's return to that example. If we wanted to answer the question, 'Is Coke an effective contraceptive?' we could administer questionnaires about sexual practices (quantity of sexual activity, use of contraceptives, use of fizzy drinks as contraceptives, pregnancy, etc.). By looking at these variables, we could see which variables correlate with pregnancy and, in particular, whether those reliant on Coca-Cola as a form of contraceptive were more likely to end up pregnant than those using other contraceptives, and less likely than those using no contraceptives at all. This is the only way to answer a question like this because we cannot manipulate any of these variables particularly easily. Even if we could, it would be totally unethical to insist on some people using Coke as a contraceptive (or indeed to do anything that would make a person likely to produce a child that they didn't intend to produce). However, there is a price to pay, which relates to causality: correlational research tells us nothing about the causal influence of variables.

### 1.7.2 Experimental research methods ■■■

Most scientific questions imply a causal link between variables; we have seen already that dependent and independent variables are named such that a causal connection is implied (the dependent variable



*depends* on the independent variable). Sometimes the causal link is very obvious in the research question, ‘Does low self-esteem cause dating anxiety?’ Sometimes the implication might be subtler; for example, in ‘Is dating anxiety all in the mind?’ the implication is that a person’s mental outlook causes them to be anxious when dating. Even when the cause–effect relationship is not explicitly stated, most research questions can be broken down into a proposed cause (in this case, mental outlook) and a proposed outcome (dating anxiety). Both the cause and the outcome are variables: for the cause, some people will perceive themselves in a negative way (so it is something that varies); and, for the outcome, some people will get more anxious on dates than others (again, this is something that varies). The key to answering the research question is to uncover how the proposed cause and the proposed outcome relate to each other; are the people who have a low opinion of themselves the same people who are more anxious on dates?

David Hume, an influential philosopher, defined a cause as ‘An object precedent and contiguous to another, and where all the objects resembling the former are placed in like relations of precedency and contiguity to those objects that resemble the latter’ (1739–40/1965).<sup>12</sup> This definition implies that (1) the cause needs to precede the effect, and (2) causality is equated to high degrees of correlation between contiguous events. In our dating example, to infer that low self-esteem caused dating anxiety, it would be sufficient to find that low self-esteem and feeling anxious when on a date co-occur, and that the low self-esteem emerged before the dating anxiety did.

In correlational research variables are often measured simultaneously. The first problem with doing this is that it provides no information about the contiguity between different variables: we might find from a questionnaire study that people with low self-esteem also have dating anxiety but we wouldn’t know whether it was the low self-esteem or the dating anxiety that came first. Longitudinal research addresses this issue to some extent, but there is still a problem with Hume’s idea that causality can be inferred from corroborating evidence, which is that it doesn’t distinguish between what you might call an ‘accidental’ conjunction and a causal one. For example, it could be that both low self-esteem and dating anxiety are caused by a third variable (e.g., poor social skills which might make you feel generally worthless but also puts pressure on you in dating situations). Therefore, low self-esteem and dating anxiety do always co-occur (meeting Hume’s definition of cause) but only because poor social skills causes them both.

This example illustrates an important limitation of correlational research: the **tertium quid** (‘A third person or thing of indeterminate character’). For example, a correlation has been found between having breast implants and suicide (Koot, Peeters, Granath, Grobbee, & Nyren, 2003). However, it is unlikely that having breast implants causes you to commit suicide – presumably, there is an external factor (or factors) that causes both; for example, low self-esteem might lead you to have breast implants and also attempt suicide. These extraneous factors are sometimes called **confounding variables**, or confounds for short.

The shortcomings of Hume’s definition led John Stuart Mill (1865) to suggest that, in addition to a correlation between events, all other explanations of the cause–effect relationship must be ruled out. To rule out confounding variables, Mill proposed that an effect should be present when the cause is present and that when the cause is absent, the effect should be absent also. In other words, the only way to infer causality is through comparing two controlled situations: one in which the cause is present and one in which the cause is absent. This is what *experimental methods* strive to do: to provide a comparison of situations (usually called *treatments* or *conditions*) in which the proposed cause is present or absent.

---

<sup>12</sup> As you might imagine, his view was a lot more complicated than this definition alone, but let’s not get sucked down that particular wormhole.

As a simple case, we might want to look at the effect of feedback style on learning about statistics. I might, therefore, randomly split<sup>13</sup> some students into three different groups, in which I change my style of feedback in the seminars on my course:

- **Group 1 (supportive feedback):** During seminars I congratulate all students in this group on their hard work and success. Even when they get things wrong, I am supportive and say things like ‘that was very nearly the right answer, you’re coming along really well’ and then give them a nice piece of chocolate.
- **Group 2 (harsh feedback):** This group receives seminars in which I give relentless verbal abuse to all of the students even when they give the correct answer. I demean their contributions and am patronizing and dismissive of everything they say. I tell students that they are stupid, worthless, and shouldn’t be doing the course at all. In other words, this group receives normal university-style seminars.☹
- **Group 3 (no feedback):** Students are not praised or punished, instead I give them no feedback at all.

The thing that I have manipulated is the feedback style (supportive, harsh or none). As we have seen, this variable is known as the independent variable and, in this situation, it is said to have three levels, because it has been manipulated in three ways (i.e., the feedback style has been split into three types: supportive, harsh and none). The outcome in which I am interested is statistical ability, and I could measure this variable using a statistics exam after the last seminar. As we have seen, this outcome variable is the dependent variable because we assume that these scores will depend upon the type of teaching method used (the independent variable). The critical thing here is the inclusion of the ‘no feedback’ group because this is a group in which our proposed cause (feedback) is absent, and we can compare the outcome in this group against the two situations in which the proposed cause is present. If the statistics scores are different in each of the feedback groups (cause is present) compared to the group for which no feedback was given (cause is absent), then this difference can be attributed to the type of feedback used. In other words, the style of feedback used caused a difference in statistics scores (Jane Superbrain Box 1.4).

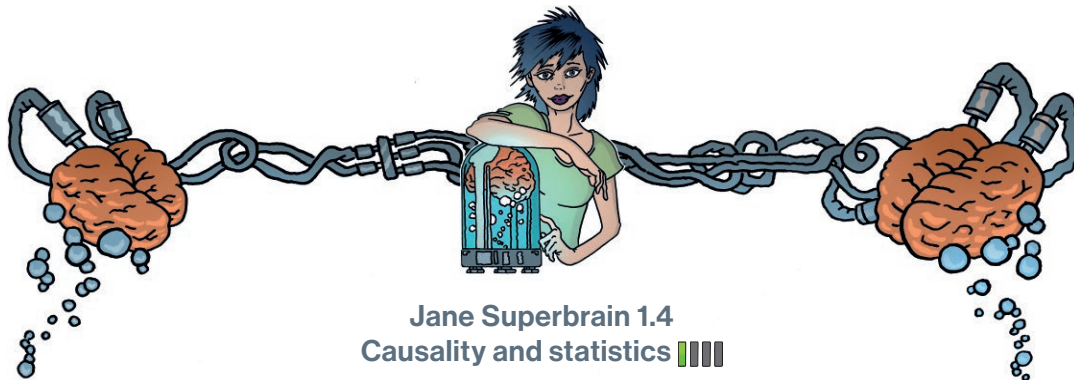
### 1.7.3 Two methods of data collection

When we use an experiment to collect data, there are two ways to manipulate the independent variable. The first is to test different entities. This method is the one described above, in which different groups of entities take part in each experimental condition (a **between-groups**, **between-subjects**, or **independent design**). An alternative is to manipulate the independent variable using the same entities. In our motivation example, this means that we give a group of students supportive feedback for a few weeks and test their statistical abilities and then give this same group harsh feedback for a few weeks before testing them again and, then, finally, give them no feedback and test them for a third time (a **within-subject** or **repeated-measures design**). As you will discover, the way in which the data are collected determines the type of test that is used to analyse the data.

### 1.7.4 Two types of variation

Imagine we were trying to see whether you could train chimpanzees to run the economy. In one training phase they are sat in front of a chimp-friendly computer and press buttons that change various parameters of the economy; once these parameters have been changed a figure appears on the screen

<sup>13</sup> This random assignment of students is important, but we’ll get to that later.



## Jane Superbrain 1.4

### Causality and statistics

People sometimes get confused and think that certain statistical procedures allow causal inferences and others don't. This isn't true, it's the fact that in experiments we manipulate the causal variable systematically to see its effect on an outcome (the effect). In correlational research we observe the co-occurrence of variables; we do not manipulate the causal variable first and then measure the effect, therefore we cannot compare the effect when the causal variable is present against when it is absent. In short, we cannot say which variable causes a change in the other; we can merely say that the variables co-occur in a certain way. The reason why some people think that certain statistical tests allow causal inferences is that, historically, certain tests (e.g., ANOVA, *t*-tests, etc.) have been used to analyse experimental research, whereas others (e.g., regression, correlation) have been used to analyse correlational research (Cronbach, 1957). As you'll discover, these statistical procedures are, in fact, mathematically identical.



indicating the economic growth resulting from those parameters. Now, chimps can't read (I don't think) so this feedback is meaningless. A second training phase is the same, except that if the economic growth is good, they get a banana (if growth is bad they do not) – this feedback is valuable to the average chimp. This is a repeated-measures design with two conditions: the same chimps participate in condition 1 *and* in condition 2.

Let's take a step back and think what would happen if we did *not* introduce an experimental manipulation (i.e., there were no bananas in the second training phase, so condition 1 and condition 2 were identical). If there is no experimental manipulation then we expect a chimp's behaviour to be similar in both conditions. We expect this because external factors such as age, sex, IQ, motivation and arousal will be the same for both conditions (a chimp's biological sex, etc. will not change from when they are tested in condition 1 to when they are tested in condition 2). If the performance measure (i.e., our test of how well they run the economy) is reliable, and the variable or characteristic that we are measuring (in this case ability to run an economy) remains stable over time, then a participant's performance in condition 1 should be very highly related to their performance in condition 2. So, chimps who score highly in condition 1 will also score highly in condition 2, and those who have low scores for condition 1 will have low scores in condition 2. However, performance won't be *identical*, there will be small differences in performance created by unknown factors. This variation in performance is known as **unsystematic variation**.

If we introduce an experimental manipulation (i.e., provide bananas as feedback in one of the training sessions), then we do something different to participants in condition 1 than in condition 2. So, the

*only* difference between conditions 1 and 2 is the manipulation that the experimenter has made (in this case that the chimps get bananas as a positive reward in one condition but not in the other).<sup>14</sup> Therefore, any differences between the means of the two conditions are probably due to the experimental manipulation. So, if the chimps perform better in one training phase than in the other, this *has* to be due to the fact that bananas were used to provide feedback in one training phase but not in the other. Differences in performance created by a specific experimental manipulation are known as **systematic variation**.

Now let's think about what happens when we use different participants – an independent design. In this design we still have two conditions, but this time different participants participate in each condition. Going back to our example, one group of chimps receives training without feedback, whereas a second group of different chimps does receive feedback on their performance via bananas.<sup>15</sup> Imagine again that we didn't have an experimental manipulation. If we did nothing to the groups, then we would still find some variation in behaviour between the groups because they contain different chimps who will vary in their ability, motivation, propensity to get distracted from running the economy by throwing their own faeces, and other factors. In short, the factors that were held constant in the repeated-measures design are free to vary in the independent design. So, the unsystematic variation will be bigger than for a repeated-measures design. As before, if we introduce a manipulation (i.e., bananas), then we will see additional variation created by this manipulation. As such, in both the repeated-measures design and the independent design there are always two sources of variation:

- **Systematic variation:** This variation is due to the experimenter doing something in one condition but not in the other condition.
- **Unsystematic variation:** This variation results from random factors that exist between the experimental conditions (such as natural differences in ability, the time of day, etc.).

Statistical tests are often based on the idea of estimating how much variation there is in performance, and comparing how much of this is systematic to how much is unsystematic.

In a repeated-measures design, differences between two conditions can be caused by only two things: (1) the manipulation that was carried out on the participants, or (2) any other factor that might affect the way in which an entity performs from one time to the next. The latter factor is likely to be fairly minor compared to the influence of the experimental manipulation. In an independent design, differences between the two conditions can also be caused by one of two things: (1) the manipulation that was carried out on the participants, or (2) differences between the characteristics of the entities allocated to each of the groups. The latter factor, in this instance, is likely to create considerable random variation both within each condition and between them. When we look at the effect of our experimental manipulation, it is always against a background of 'noise' created by random, uncontrollable differences between our conditions. In a repeated-measures design this 'noise' is kept to a minimum and so the effect of the experiment is more likely to show up. This means that, other things being equal, repeated-measures designs are more sensitive to detect effects than independent designs.

### 1.7.5 Randomization

In both repeated-measures and independent designs it is important to try to keep the unsystematic variation to a minimum. By keeping the unsystematic variation as small as possible we get a more

<sup>14</sup> Actually, this isn't the only difference because, by condition 2, they have had some practice (in condition 1) at running the economy; however, we will see shortly that these practice effects are easily eradicated.

<sup>15</sup> Obviously I mean that they receive a banana as a reward for their correct response and not that the bananas develop little banana mouths that sing them a little congratulatory song.

sensitive measure of the experimental manipulation. Generally, scientists use the **randomization** of entities to treatment conditions to achieve this goal. Many statistical tests work by identifying the systematic and unsystematic sources of variation and then comparing them. This comparison allows us to see whether the experiment has generated considerably more variation than we would have got had we just tested participants without the experimental manipulation. Randomization is important because it eliminates most other sources of systematic variation, which allows us to be sure that any systematic variation between experimental conditions is due to the manipulation of the independent variable. We can use randomization in two different ways depending on whether we have an independent or repeated-measures design.

Let's look at a repeated-measures design first. I mentioned earlier (in a footnote) that when the same entities participate in more than one experimental condition they are naive during the first experimental condition but they come to the second experimental condition with prior experience of what is expected of them. At the very least they will be familiar with the dependent measure (e.g., the task they're performing). The two most important sources of systematic variation in this type of design are:

- **Practice effects:** Participants may perform differently in the second condition because of familiarity with the experimental situation and/or the measures being used.
- **Boredom effects:** Participants may perform differently in the second condition because they are tired or bored from having completed the first condition.

Although these effects are impossible to eliminate completely, we can ensure that they produce no systematic variation between our conditions by **counterbalancing** the order in which a person participates in a condition.

We can use randomization to determine in which order the conditions are completed. That is, we randomly determine whether a participant completes condition 1 before condition 2, or condition 2 before condition 1. Let's look at the teaching method example and imagine that there were just two conditions: no feedback and harsh feedback. If the same participants were used in all conditions, then we might find that statistical ability was higher after the harsh feedback. However, if every student experienced the harsh feedback after the no feedback seminars then they would enter the harsh condition already having a better knowledge of statistics than when they began the no feedback condition. So, the apparent improvement after harsh feedback would not be due to the experimental manipulation (i.e., it's not because harsh feedback works), but because participants had attended more statistics seminars by the end of the harsh feedback condition compared to the no feedback one. We can use randomization to ensure that the number of statistics seminars does not introduce a systematic bias by randomly assigning students to have the harsh feedback seminars first or the no feedback seminars first.

If we turn our attention to independent designs, a similar argument can be applied. We know that participants in different experimental conditions will differ in many respects (their IQ, attention span, etc.). Although we know that these confounding variables contribute to the variation between conditions, we need to make sure that these variables contribute to the unsystematic variation and *not* to the systematic variation. A good example is the effects of alcohol on behaviour. You might give one group of people 5 pints of beer, and keep a second group sober, and then count how many times you can persuade them to do a fish impersonation. The effect that alcohol has varies because people differ in their tolerance: teetotal people can become drunk on a small amount, while alcoholics need to consume vast quantities before the alcohol affects them. If you allocated a bunch of hardened drinkers to the condition that consumed alcohol, and teetotal people to the no alcohol condition, then you might find that alcohol doesn't increase the number of fish impersonations you get. However, this finding could be because (1) alcohol does not make people engage in frivolous activities, or (2) the hardened drinkers were unaffected by the dose of alcohol. You have no way to dissociate these explanations



because the groups varied not just on dose of alcohol but also on their tolerance of alcohol (the systematic variation created by their past experience with alcohol cannot be separated from the effect of the experimental manipulation). The best way to reduce this eventuality is to randomly allocate participants to conditions: by doing so you minimize the risk that groups differ on variables other than the one you want to manipulate.

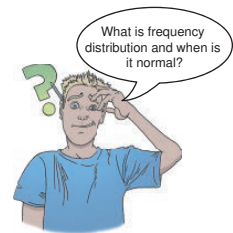


## 1.8 Analysing data ■■■

The final stage of the research process is to analyse the data you have collected. When the data are quantitative this involves both looking at your data graphically (Chapter 5) to see what the general trends in the data are, and also fitting statistical models to the data (all other chapters). Given that the rest of the book is dedicated to this process, we'll begin here by looking at a few fairly basic ways to look at and summarize the data you have collected.

### 1.8.1 Frequency distributions ■■■

Once you've collected some data a very useful thing to do is to plot a graph of how many times each score occurs. This is known as a **frequency distribution**, or **histogram**, which is a graph plotting values of observations on the horizontal axis, with a bar showing how many times each value occurred in the data set. Frequency distributions can be very useful for assessing properties of the distribution of scores. We will find out how to create these types of charts in Chapter 5.



Frequency distributions come in many different shapes and sizes. It is quite important, therefore, to have some general descriptions for common types of distributions. In an ideal world our data would be distributed symmetrically around the centre of all scores. As such, if we drew a vertical line through the centre of the distribution then it should look the same on both sides. This is known as a **normal distribution** and is characterized by the bell-shaped curve with which you might already be familiar. This shape implies that the majority of scores lie around the centre of the distribution (so the largest bars on the histogram are around the central value). Also, as we get further away from the centre, the bars get smaller, implying that as scores start to deviate from the centre their frequency is decreasing. As we move still further away from the centre our scores become very infrequent (the bars are very short). Many naturally occurring things have this shape of distribution. For example, most men in the UK are around 175 cm tall;<sup>16</sup> some are a bit taller or shorter, but most cluster around this value. There will be very few men who are really tall (i.e., above 205 cm) or really short (i.e., under 145 cm). An example of a normal distribution is shown in Figure 1.3.

<sup>16</sup> I am exactly 180 cm tall. In my home country this makes me smugly above average. However, I often visit the Netherlands, where the average male height is 185 cm (a little over 6ft, and a massive 10 cm higher than the UK), and where I feel like a bit of a dwarf.