DISCOVERING STATISTICS USING IBM SPSS STATISTICS

CATISFIED CUSTOMERS

DISCOVERING STATISTICS USING IBM SPSS STATISTICS

AND SEX AND DRUGS AND ROCK 'N' ROLL

4TH EDITION

ANDY FIELD

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

Los Angeles | London | New Delhi
Singapore | Washington DC

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

First edition published November 2002
Second edition published 2005
Third edition published 2009

Editor: Michael Carmichael
Development editor: Robin Lupton
Production editor: Ian Antcliff
Copyeditor: Richard Leigh
Proofreaders: Louise Harnby; Kate Harrison; Anna Gilding; Jennifer Hinchliffe
Indexer: David Rudeforth
Marketing manager: Ben Griffin-Sherwood
Cover design: Wendy Scott
Typeset by:
Printed by:

Library of Congress Control Number: 2012944559

British Library Cataloguing in Publication data

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

ISBN 978-1-4462-4917-8
ISBN 978-1-4462-4918-5 (pbk)

CONTENTS

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 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 really 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 good 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 better books: Howell (2012), Stevens (2002) and Tabachnick and Fidell (2012) have taught (and continue to teach) me more about statistics than you could possibly imagine. (I have an ambition to be cited in one of these books, but I don't think that will ever happen.) 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. It's a book that I hope will become your friend from first year at university right through to your professorship. The start of the book is aimed at first-year undergraduates (Chapters 1–9), and then we move onto second-year undergraduate level material (Chapters 5, 8 and 10–15) before a dramatic climax that should keep postgraduates tickled

(Chapters 16–20). There should be something for everyone in each chapter also, and to help you gauge the difficulty of material, I flag the level of each section within each chapter (more on that in a moment).

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 4Q SUBJECT Mathematics

EXAM	
ATTAINMENT	
EFFORT	

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.

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

The catalyst of this remarkable change was having 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. He got the 'good teaching' genes in the family, but wasted them by not becoming a teacher; however, they're a little less wasted because his approach inspires mine. I strongly believe that people appreciate the human touch, and so I try to 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 calling me a pile of gibbon excrement, but you have to take the rough with the smooth.) The second edition of this book also won the British Psychological Society book award in 2007. However, with every new edition, I fear that the changes I make will ruin all of my previous hard work. Let's see what those changes are.

What do you get for your money?

This book takes you on a journey (and 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 stupid examples, bad jokes, smut and filth. Aside from the smut, I have been forced reluctantly to include some academic content. In essence 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 on multilevel modelling (Chapter 20). Of course no book that it's physically possible to lift will 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 stupid faces, some of them my own. You can find out more about the pedagogic function of these 'characters' in the next section, but even without any useful function they're nice to look at.
- **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 hopefully it provides some light relief between chapters.
- **SPSS tips**: SPSS does weird 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 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 to them on the companion website so that you can check on your progress.
- **Companion website**: The companion website contains an absolutely insane amount of additional material, all of which is described in the section about the companion website.
- **Digital stimulation**: No, not the aforementioned type of digital stimulation, but brain stimulation. Many of the features on the companion 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 you write 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. However, just for you, 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?

My publishers are obsessed with me telling you how this book differs from the last edition. 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 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 4 year 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 versions 20 and 21 of IBM SPSS Statistics. IBM bring out a new SPSS each year and this book gets rewritten about every 4 years, so, depending on when you buy the book, it may not reflect the latest version. This shouldn't bother you because one edition of SPSS is much the same as another (see Section 3.2).
- **New! Mediation and Moderation**: Even since the first edition I have been meaning to do a chapter on mediation and moderation, because they are two very widely used techniques. With each new edition I have run out of energy. Not this time though: I wrote it in the middle of the update before I managed to completely burn myself out. Chapter 10 is brand spanking new and all about mediation and moderation.
- **New! Structure**: My publishers soiled their underwear at the thought of me changing the structure because they think lecturers who use the book don't like this sort of change. They might have a point, but I changed it anyway. So, logistic regression (a complex topic) has moved towards the end of the book, and non-parametric tests (a relatively straightforward topic) have moved towards the beginning. In my opinion this change enables the book's story to flow better.
- **New! Focus**: Statistical times are a-changing, and people are starting to appreciate the limitations of significance testing, so I have discussed this more in Chapter 2, and the points made there permeate the rest of the book. The theme of 'everything being the same model' has run through all editions of the book, but I have made this theme even more explicit this time.
- **New! Tasks**: There are 111 more Smart Alex tasks, and 8 more Labcoat Leni tasks. This, of course, means there are quite a lot more pages of answers to these tasks on the companion website.
- **New! Bootstrapping**: The SPSS bootstrapping procedure is covered in every chapter where it is relevant.
- **New! Process diagrams**: Every chapter has a diagrammatic summary of the key steps that you go through for a particular analysis.
- **New! Love story**: Every chapter has a diagrammatic summary at the end (*Brian's attempt to woo Jane*). More interesting, though, Brian Haemorrhage has fallen in love with Jane Superbrain (see next section) and these diagrams follow Brian's attempts to convince Jane to go on a date with him.
- **New! Characters**: I enjoy coming up with new characters, and this edition has a crazy hippy called Oditi, and a deranged philosopher called Confusius (see the next section).
- **New-ish! Assumptions**: I've never really liked the way I dealt with assumptions, so I completely rewrote Chapter 5 to try to give more of a sense of when assumptions actually matter.

Every chapter had a serious edit/rewrite, but here is a chapter-by-chapter run-down of the more substantial changes:

- **Chapter 1 (Doing research)**: I added some more material on reporting data. I added stuff about variance and standard deviations, and expanded the discussion of *p*-values.
- **Chapter 2 (Statistical theory)**: I added material on estimating parameters, significance testing and its limitations, problems with one-tailed tests, running multiple tests (i.e., familywise error), confidence intervals and significance, sample size and significance, effect sizes (including Cohen's *d* and meta-analysis), and reporting basic statistics. It's changed a lot.
- **Chapter 3 (IBM SPSS)**: No dramatic changes.
- **Chapter 4 (Graphs)**: I moved the discussion of outliers into Chapter 5, which meant I had to rewrite one of the examples. I now include population pyramids also.
- **Chapter 5 (Assumptions)**: I completely rewrote this chapter. It's still about assumptions, but I try to explain when they matter and what they bias. Rather than dealing with assumptions separately in every chapter, because everything in the book is a linear model, I deal with the assumptions of linear models here. Therefore, this chapter acts as a single reference point for all subsequent chapters. I also cover other sources of bias such as outliers (which used to be scattered about in different chapters).
- **Chapter 6 (Non-parametric models)**: This is a fully updated and rewritten chapter on non-parametric statistics. It used to be later in the book, but now flows gracefully on from the discussion of assumptions.
- **Chapter 7 (Correlation)**: No dramatic changes.
- **Chapter 8 (Regression)**: I restructured this chapter so that most of the theory is now at the beginning and most of the SPSS is at the end. I did a fair bit of editing, too, moved categorical predictors into Chapter 10, and integrated simple and multiple regression more.
- **Chapter 9 (*t*-tests)**: The old version of this chapter used spider examples, but someone emailed me to say that this freaked them out, so I changed the example to be about cloaks of invisibility. Hopefully that won't freak anyone out. I restructured a bit, too, so that the theory is in one place and the SPSS in another.
- **Chapter 10 (Mediation and moderation)**: This chapter is completely new.
- **Chapter 11 (GLM 1)**: I gave more prominence to ANOVA as a general linear model because this makes it easier to think about assumptions and bias. I moved some of the more technical bits of the SPSS interpretation into boxes so that you can ignore them if you wish.
- **Chapter 12 (GLM 2)**: Again some restructuring and a bit more discussion on whether the covariate and predictor need to be independent.
- **Chapters 13–15 (GLM 3–5)**: These haven't changed much. I restructured each one a bit, edited down/rewrote a lot and gave more prominence to the GLM way of thinking.
- **Chapter 16 (MANOVA)**: I gave the writing a bit of a polish, but no real content changes.
- **Chapter 17 (Factor analysis)**: I added some stuff to the theory to make the distinction between principal component analysis (PCA) and factor analysis (FA) clearer. The chapter used to focus on PCA, but I changed it so that the focus is on FA. I edited out 3000 words of my tedious, repetitive, superfluous drivel.

- **Chapters 18 and 19 (Categorical data and logistic regression)**: Because these chapters both deal with categorical outcomes, I rewrote them and put them together. The basic content is the same as before.
- **Chapter 20 (Multilevel models)**: I polished the writing a bit and updated, but there are no changes that will upset anyone.

Goodbye

The first edition of this book was the result of two years (give or take a few weeks to write up my Ph.D.) 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 you look back at your past work and think you can do things better). This fourth edition is the culmination of about 6 years of full-time work (on top of my actual job). This book has literally consumed the last 15 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 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.

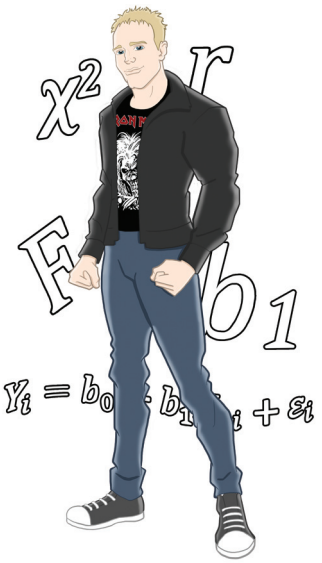
Andy

www.facebook.com/profandyfield

@ProfAndyField

www.youtube.com/user/ProfAndyField

www.methodspace.com/profile/ProfessorAndyField



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, so I really don't assume much).

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

Yes, more or less: Chapters 1–9 are first-year degree level, Chapters 8–15 move into second-year degree level, and Chapters 16–20 discuss more technical topics. However, my main aim is to tell a statistical story rather than worrying 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 linear models): Chapters 1–3.
- Part 2 (Exploring data): Chapters 4–6.
- Part 3 (Linear models with continuous predictors): Chapters 7 and 8.
- Part 4 (Linear models with continuous or categorical predictors): Chapters 9–15.
- Part 5 (Linear models with multiple outcomes): Chapter 16 and 17.
- Part 6 (Linear models with categorical outcomes): Chapters 18–19.
- Part 7 (Linear models with hierarchical data structures): Chapter 20.

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 you can skip the sections (but see Smart Alex in the next section), 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 'I':

① *Introductory*, which I hope means that everyone should be able to understand these sections. These are for people just starting their undergraduate courses.

② *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.

③ *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.

④ *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 stupid 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, in front of their loving family and friends. 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 embark on a journey that will take him from statistically challenged to a genius, in 900 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 she'll go on a date with him.

New! 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.

Cramming Sam: Samantha thinks statistics is a boring waste of time and 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.

Curious Cat: He also pops up and asks questions (because he's curious). The only reason he's here is because I wanted a cat in the book … and preferably one that looks like mine. Of course the educational specialists think he needs a specific role, and so his role is to look cute and make bad cat-related jokes.

Jane Superbrain: Jane is the cleverest person in the whole universe. A mistress of osmosis, she acquired vast statistical knowledge by stealing the brains of statisticians and eating them. Apparently they taste of sweaty tank tops. Having devoured some top statistics brains and absorbed their knowledge, she knows all of the really hard stuff. She appears in boxes to tell you advanced things that are a bit tangential to the main text. Her friends tell her that a half-whit called Brian is in love with her, but she doesn't know who he is.

Labcoat Leni: Leni is a budding young scientist and he's fascinated by real research. He says, 'Andy, man, 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, dude!' So off Leni went: he walked the globe, a lone data warrior in a thankless quest for real data. He turned up at universities, cornered academics, kidnapped their families and threatened to put them in a bath of crayfish unless he was given real data. The generous ones relented, but others? Well, let's just say their families are sore. So, when you see Leni you know that you will get some real data, from a real research study to analyse. Keep it real.

New! 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 re-create 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 there is a video tutorial to guide you.

Oliver Twisted: With apologies to Charles Dickens, Oliver, like the more famous fictional London urchin, is always asking 'Please, Sir, can I have some more?' Unlike Master Twist though, our young Master Twisted wants more statistics information. Of course he does, 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 to be found 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 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 is a very important character because he appears when things get particularly difficult. He's basically a bit of a smart alec, and so whenever you see his face you know that something scary is about to be explained. When the hard stuff is over he reappears to let you know that it's safe to continue. You'll also find that Alex gives you tasks to do at the end of each chapter to see whether you're as smart as he is.

Why do I keep seeing QR codes?

Study on the Go: QR stands for 'quantum reality', and if you scan one of these funny little barcode things into your mobile device (smartphone etc.) it will transport you and your device into a quantum reality in which left is right, time runs backwards, drinks pour themselves out of your mouth into bottles, and statistics is interesting. Scanning these codes will be your gateway to mobile-friendly versions of the materials on the companion website. It means you can have stats anytime, anywhere, even on the toilet.

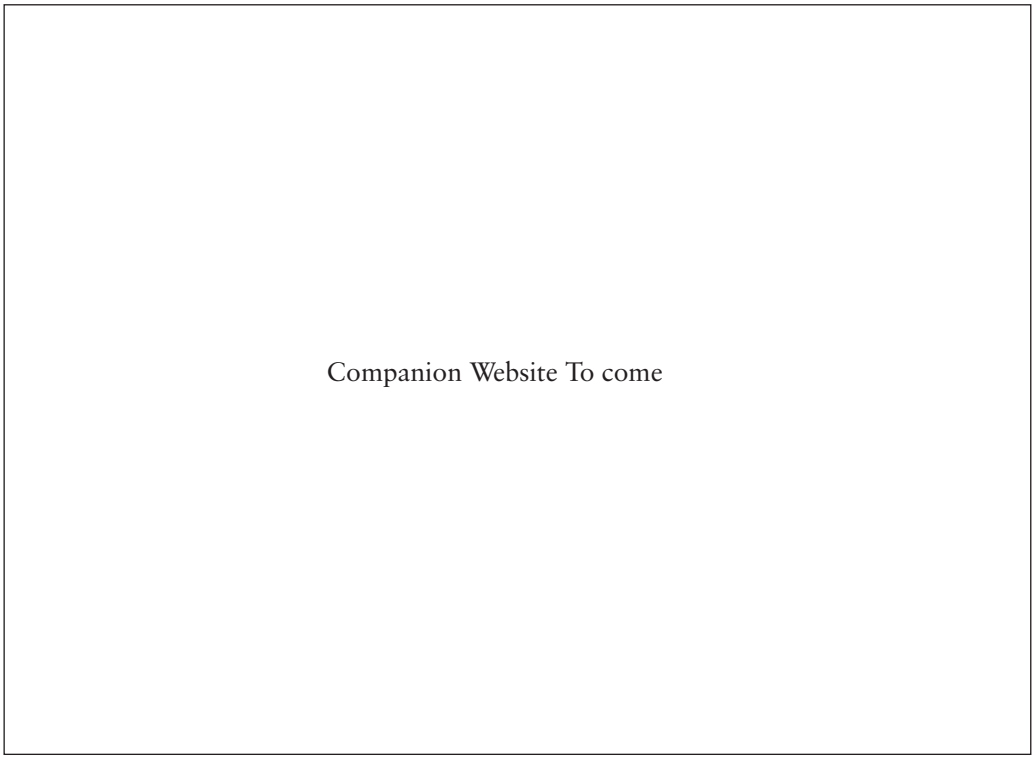
What is on the companion website?

In this age of downloading, CD-ROMs are for losers (at least that's what the 'kids' tell me), so I've put my cornucopia of additional funk on that worldwide interweb thing. To enter my world of delights, go to www.sagepub.co.uk/field4e. The website contains resources for students and lecturers alike, organized by chapter and character (unlike last time):

- **Testbank**: There is a (hopefully) comprehensive testbank of multiple-choice and numeracy-based/algorithmic questions for instructors to use. It comes in two flavours: (1) a file that you can upload into your online teaching system (with answers separated out), or (2) preloaded into an online homework and grading system called WebAssign® (www.webassign.net). WebAssign enables you to assign questions for exams and assignments. Students can receive instant feedback on correct/incorrect answers, including pointers to areas in the book where the right answer can be found.
- **Data files**: You need data files to work through the examples in the book and they are all on the companion website. We did this so that you're forced to go there and once you're there Sage will flash up subliminal messages to make you buy more of their books.
- **Resources for different 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. I realize that not everyone is as psychologically oriented as me, so 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.
- **Webcasts**: Whenever you see Oditi in the book it means that there is a webcast to accompany the chapter. These are hosted on my YouTube channel (www.youtube.com/user/ProfAndyField), which I have amusingly called μ-Tube (see what I did there?). You can also get to them via the companion website.
- **Self-assessment multiple-choice questions**: Organized by chapter, these will 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 technical 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 off of your bookshelf onto your head.
- **Labcoat Leni solutions**: For all of the Labcoat Leni tasks in the book there are full and detailed answers on the companion website.
- **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 around 300 pages (that's a different 300 pages to the 300 above) of 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 superintelligent 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 a picture of an anus on their slides, then remember that's probably my fault.
- **Links**: Every website has to have links to other useful websites, and the companion website is no exception.
- **Cyberworms of knowledge**: I have used nanotechnology to create cyberworms that crawl down your broadband connection, pop out of the USB port of your computer and fly through space into your brain. They rearrange your neurons so that you understand statistics. You don't believe me? Well, you'll never know for sure unless you visit the companion website ….

Happy reading, and don't get distracted by Facebook and Twitter.



ACKNOWLEDGEMENTS

This book (in all its SPSS, SAS and R versions) wouldn't have happened if not for Dan Wright's unwarranted faith in the ability of 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 are due to to Dan (again), David Hitchin, Laura Murray, Gareth Williams, Lynne Slocombe and Kate Lester, who gave me significant amounts of feedback on various incarnations of this text. For this edition, lots of people sent me very helpful emails, but in particular Maria de Ridder (for a very helpful document of suggestions) and Thom Baguley (for feedback on Chapter 5 at very short notice).

Special thanks to Jeremy Miles for his help with various versions of this book over the years. 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 …).

Thanks to the following for sending me 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 Schüetzwohl, 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 contributes so enthusiastically to my Facebook page: I hit some motivational dark times during this edition, but feeling the positive vibes from readers always got me back on track (especially the photos of cats, dogs, parrots and lizards with this book☺). I continue to be amazed and bowled over by the nice things that people say about the book.

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 whole approach to teaching has been to steal all of 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.

This book wouldn't exist without the generous support of IBM who allow me to beta-test SPSS Statistics and keep me up to date with the software (www-01.ibm.com/software/analytics/spss). I wrote this edition on a Mac but used Windows for the screen shots. 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. Thanks to Jess Knott at TechSmith (www.techsmith.com) who provided support for Camtasia (which I use to record and edit my webcasts) and Snagit (which I used for screen shots) for Mac; she also breathed new life into some of my old webcasts. I

created most diagrams and flowcharts in this book using OmniGraffle (www.omnigroup.com). Although it is unhealthy to love a piece of software, I love OmniGraffle and if I ever meet the authors I will buy them beer, and plenty of it.

My publishers, Sage, are rare in being a large, successful company that manages to maintain a family feel. Like any family, I don't always see eye-to-eye with them, and sometimes we drive each other mad, but we wouldn't want to be without each other. They generously co-funded my wife to help update the web materials for this book. My editor Mike takes his fair share of crap from me (what does he expect, he supports Tottenham?), but I always appreciate his unfaltering enthusiasm, support and willingness to make things happen. Thanks to everyone at Sage, but especially Ian, Karen (who supports a proper football team) and Ziyad (who doesn't know what football is) for their help and support over many years.

You'll notice the book is riddled with nicely drawn characters. I didn't draw them. The ones that survived from the previous edition are thanks to Alex Lee. Special thanks to Laura-Jane at Anelina Illustrations (www.anelinaillustrations.com) for my caricature, Oditi and Confusius. It was great working with you.

I always write listening to music. For this edition I predominantly enjoyed (my neighbours less so): Absu, Anathema, Anthrax, Animals as Leaders, Audrey Horne, The Beyond, Black Breath, Black Tusk, Black Sabbath, Blue Öyster Cult, Blut Aus Nord, Deathspell Omega, Deep Purple, Foo Fighters, Genesis, Graveyard, Ihsahn, Iron Maiden, Jethro Tull, Kiss, Manowar, Marillion, Meshuggah (a lot), Metallica, Mastodon, Motörhead, Opeth (a lot), Primal Rock Rebellion, Rainbow, Rush, Secrets of the Moon, Status Quo, Steve Wilson, Storm Corrosion, Sylosis, Torche, Uriah Heep, Watain, and Wolves in the Throne Room.

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. Peter Muris, Birgit Mayer, and especially Leonora Wilkinson gave me a lot of support while writing previous editions of this book. For this edition, and in most cases all of the previous ones, my eternal gratitude goes to Graham Davey, Ben Dyson, Martin Watts, Sam Cartwright-Hatton, Mark Franklin and their lovely families for reminding me that there is more to life than work. You'd also be hard pushed to find more supportive, loving and proud parents than mine. I also throw out a robust set of horns to my brothers of metal Doug Martin and Rob Mepham for letting me deafen them with my drumming (www.myspace.com/fracturepattern). I also became an uncle while writing this update, so thanks to Melody for being small and cute.

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). As if that wasn't enough, she also did the lion's share of the updating of the accompanying web materials and SPSS screen shots for this edition. I never forget, not even for a nanosecond, how lucky I am.

With thanks to the following people for their feedback on the fourth edition

Mahmood Ali, University of Greenwich
Rory Allen, Goldsmiths, University of London
Derek Ashford, Manchester Metropolitan University
Chris Askew, Kingston University
Mirjam Baars, University of Amsterdam
Theodoros Bampouras, University of Cumbria
Kirsten Bartlett, Sheffield Hallam University
Roger Bennett, London Metropolitan University
Kanishka Bhattacharya, Oxford University
Kate Black, University of Chester
Ian Boardley, University of Birmingham
Nicola Brown, St Mary's University College
Douglas Bryson, ESC Rennes School of Business
Hannah Buchanan-Smith, University of Stirling
Louise Bunce, London Metropolitan University
Scott Burnet, Southampton Solent University
Robert Busching, University of Potsdam
Katherine Cagney, Waterford Institute of Technology
Ian Charity, Newcastle Business School
Sam Chenery-Morris, University Campus Suffolk
I-Chant Chiang, Quest University Canada
Nikki Coghill, University of Bristol
Graham Cookson, King's College London
Julie Davies, Bangor University
Matthew Davis, University of Leeds
Rutger de Graaf, University of Amsterdam
Peter de Waal, Utrecht University
Pedro Dias, Catholic University of Portugal
Gerlof Donga, Hogeschool van Amsterdam
Michelle Ellefson, University of Cambridge
Chris Fife-Schaw, University of Surrey
Liam Foster, University of Sheffield
Pauline Fox, University of West London
Robert Francis, King's College London
Benjamin Gardner, University College London
Lise Georgeson, St Mary's University College
Elisabeth Götze, Vienna University of Economics and Business
Martijn Goudbeek, Tilburg University
Hannah Greatwood, Leeds Metropolitan University
Suzanne Hacking, University of Central Lancashire
Sebastian Hagen, University of Leipzig
Carol Haigh, Manchester Metropolitan University
Julie Hall, Birmingham City University
Karen Hambly, University of Kent
Karolina Hansen, Friedrich Schiller University Jena
Azmi Hassali, Universiti Sains Malaysia
Erica Hepper, University of Southampton
Beryl Hilberink-Schulpen, Radboud University Nijmegen
Henriette Hogh, University of Surrey
Lee Hulbert-Williams, University of Wolverhampton
Gerry Humphris, University of St Andrews
Martyn Jarvis, University of Glamorgan
Domhnall Jennings, Newcastle University
Paul Jepson, University of Birmingham
Stephen Johnston, Swansea University
Marie Juanchich, Kingston University
Serdar Karabati, Bilgi University
Roshni Khatri, University of Northampton
Paul Kiff, University of East London
Thomas Koch, LMU Munich
Wolfgang Kotowski, University of Zurich
Wander Lowie, University of Groningen
Ruth Lowry, University of Chichester
John Mallett, University of Ulster
Chris Mamo, Cork Institute of Technology
Irene Manaras, University of Hertfordshire
Anne Manyande, University of West London
Jesse Martin, Bangor University
Carla Martins, University of Minho
Paul McCarthy, Glasgow Caledonian University
Philip McDonald, Leeds Trinity University College
Margaret McGrath, National University of Ireland, Galway
Gary Mckenna, University of the West of Scotland
Alistair McMillan, University of Sheffield
Juanjo Medina, University of Manchester
Remo Mombarg, Hanze University of Applied Sciences, Groningen
Maria Mos, Tilburg University
Eyob Mulat-Weldemeskel, London Metropolitan University
Drew Munn, University of Northampton
Tony Myers, Newman University College
Rich Neil, Cardiff Metropolitan University
Dennis Nigbur, Canterbury Christ Church University
Louis Passfield, University of Kent
Monique Pollmann, Tilburg University
Michelle Pyer, University of Northampton
Shahid Qureshi, Institute of Business Administration, Karachi
Eric Rietzschel, University of Groningen
Claire-Marie Roberts, University of Worcester
Janet Robertson, Lancaster University
Petros Roussos, University of Athens
Tom Scherndl, University of Salzburg
Kim Schildkamp, University of Twente
Shivani Sharma, University of Hertfordshire
Luke Sloan, Cardiff University
Sandrino Smeets, Radboud University Nijmegen
Andy Smith, Leeds Metropolitan University
Martin Southam, Canterbury Christ Church University
Jayne Tidd, Teesside University
Alastair Tomlinson, Cardiff Metropolitan University
Alison Tresidder, University of Bedfordshire
Ivo van der Lans, Wageningen University
Frans Van der Slik, Radboud University Nijmegen
Marcel van Egmond, University of Amsterdam
Nel Verhoeven, University College Roosevelt Academy
Paresh Wankhade, Liverpool Hope University
Maggie Whittaker, University of Essex
Bob Williams, Durham University

Dedication

Like the previous editions, this book is dedicated to my brother Paul and my cat Fuzzy, because one of them is a constant source of intellectual inspiration and the other wakes me up in the morning by sitting on me and purring in my face until I give him cat food: mornings will be considerably more pleasant when my brother gets over his love of cat food for breakfast.☺

SYMBOLS USED IN THIS BOOK

Mathematical operators

Σ	This symbol (called sigma) means 'add everything up'. So, if you see something like Σx_i it just means 'add up all of the scores you've collected'.
Π	This symbol means 'multiply everything'. So, if you see something like Πx_i it just means 'multiply all of the scores you've collected'.
$\sqrt{x}$	This means 'take the square root of x'.

Greek symbols

α	The probability of making a Type I error
β	The probability of making a Type II error
β_i	Standardized regression coefficient
ε	Usually stands for 'error'
η^2	Eta squared
μ	The mean of a population of scores
ρ	The correlation in the population
σ	The standard deviation in a population of data
σ^2	The variance in a population of data
$\sigma_{\bar{x}}$	The standard error of the mean
τ	Kendall's tau (non-parametric correlation coefficient)
χ^2	Chi-square test statistic
χ^2_F	Friedman's ANOVA test statistic
ω^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.

English 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 ith person
F	F-ratio (test statistic used in ANOVA)
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: 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	Test statistic for Student's t-test
T	Test statistic for Wilcoxon's matched-pairs signed-rank test
U	Test statistic for the Mann–Whitney test
W_s	Test statistic for Wilcoxon's rank-sum test
$\bar{X}$	The mean of a sample of scores
z	A data point expressed in standard deviation units

SOME MATHS REVISION

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 makes 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: This is an acronym for the order in which mathematical operations are performed: Brackets, Order, Division, Multiplication, Addition, and Subtraction. Mostly these operations are self-explanatory (e.g., always calculate things within brackets first) except for 'order', which refers 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 (also, if we used 'power', we'd end up with BPDMAS, which doesn't roll off the tongue quite so nicely). Let's look at an example of BODMAS: 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 term: 5^2 is 25, so the equation becomes $1 + 3 \times 25$. There is no division, so we can move on to multiplication: 3×25, which gives us 75. BODMAS tells us to deal with addition next: $1 + 75$, which gives us 76 and the equation is solved. If I'd written the original equation 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 equation becomes 4×5^2. We then deal with the order term, so the equation becomes $4 \times 25 = 100$.

http://www.bbc.co.uk/schools/gcsebitesize/maths/ is a good site for revising basic maths.