INTRODUCTION TO THE NEW STATISTICS

ESTIMATION, OPEN SCIENCE, & BEYOND

GEOFF CUMMING &
ROBERT CALIN-JAGEMAN

COMPANION WEBSITE

ROUTLEDGE

BOOK WEBSITE

- For lots of goodies, make this a favorite or bookmark: www.routledge.com/cw/cumming

ABBREVIATIONS

ANOVA	Analysis of variance
APA	American Psychological Association
CI	confidence interval
df	degrees of freedom
COS	Center for Open Science (cos.io)
DFY	don't fool yourself
DV	dependent variable
ES	effect size
ESCI	Exploratory Software for Confidence Intervals
H_0	null hypothesis
H_1	alternative hypothesis
IQR	inter-quartile range
IV	independent variable
LL	lower limit of a CI
MoE	margin of error (length of one arm of a CI)
MoE_{av}	average of the two arm lengths of an asymmetric CI
NHST	null hypothesis significance testing
NOIR	levels of measurement: nominal, ordinal, interval, ratio
OS	Open Science
OSF	Open Science Framework (osf.io)
Q1	first quartile
Q2	second quartile, median
Q3	third quartile
RCT	randomized control trial
SD	standard deviation
SE	standard error
UL	upper limit of a CI

SYMBOLS

α	Type I error rate (alpha)
β	Type II error rate (beta)
β	slope of a standardized regression line, standardized regression weight
δ	population ES, Cohen's δ (delta)
η^2	proportion of variance, in ANOVA (eta squared)
μ	population mean (mu)
μ_0	μ specified in the null hypothesis
μ_1	μ specified in the alternative hypothesis
Π	population proportion (upper case pi)
ρ	population Pearson's correlation (rho)
Σ	addition (upper case sigma)

σ	population SD (sigma)
σ^2	population variance
ϕ	phi coefficient
χ^2	chi-square
ω^2	proportion of variance, in ANOVA (omega squared)
A	first IV in a factorial design
a	intercept of a regression line
B	second IV in a factorial design
b	slope of a regression line, regression weight
C	level of confidence
d	sample ES, Cohen's d
d_{unbiased}	unbiased estimate of δ
$F(df_1, df_2)$	test statistic used in ANOVA
i	integer, used as an index
k	number of levels of an IV
M	sample mean
M_t	trimmed mean
N, n	sample size
p	p value, in NHST
P	proportion
r	Pearson's correlation
s	sample SD
s^2	sample variance
s_{av}	standardizer for Cohen's d, paired design
s_p	pooled SD
s_t	SD for trimmed data set
t	variable, often with t distribution
$t_{.95}(df)$	.95 critical value of t for stated df
V	variance
X	dependent variable
X	integer, numerator of a proportion
$\hat{Y}$	regression prediction of Y
z	variable, often with normal distribution
$z_{.95}$	z for central .95 area under normal distribution
?	$.05 < p < .10$
*	$.01 < p < .05$
**	$.001 < p < .01$
***	$p < .001$

Introduction to the New Statistics

This is the first introductory statistics text to use an estimation approach from the start to help readers understand effect sizes, confidence intervals (CIs), and meta-analysis ("the new statistics"). It is also the first text to explain the new and exciting Open Science practices, which encourage replication and enhance the trustworthiness of research. In addition, the book explains null hypothesis significance testing (NHST) fully so students can understand published research. Numerous real research examples are used throughout. The book uses today's most effective learning strategies and promotes critical thinking, comprehension, and retention, to deepen users' understanding of statistics and modern research methods. The free ESCI (Exploratory Software for Confidence Intervals) software makes concepts visually vivid, and provides calculation and graphing facilities. The book can be used with or without ESCI. Other highlights include:

- Both estimation and NHST approaches are covered, and full guidance given on how to easily translate between the two.
- Some exercises use ESCI to analyze data and create graphs including CIs, for best understanding of estimation methods.
- Videos of the authors describing key concepts and demonstrating use of ESCI provide an engaging learning tool for traditional or flipped classrooms.
- In-chapter exercises and quizzes with related commentary allow students to learn by doing, and to monitor their progress.
- End-of-chapter exercises and commentary, many using real data, give practice for analyzing data, as well as for applying research judgment in realistic contexts.
- *Don't fool yourself* tips help students avoid common errors.
- *Red Flags* highlight the meaning of "significance" and what *p* values actually mean.
- Chapter outlines, defined key terms, sidebars of key points, and summarized take-home messages provide study tools at exam time.
- www.routledge.com/cw/cumming offers for students: ESCI downloads; data sets; key term flashcards; guides; tips for using IBM's SPSS and R for analyzing data; and videos. For instructors it offers: tips for teaching the new statistics and Open Science; additional assessment exercises; answer keys for homework and assessment items; question bank for quizzes and exams; downloadable slides with text images; and PowerPoint lecture slides.

Designed for introduction to statistics, data analysis, or quantitative methods courses in psychology, education, and other social and health sciences, researchers interested in understanding the new statistics will also appreciate this book. No familiarity with introductory statistics is assumed.

Geoff Cumming is professor emeritus of La Trobe University and has been teaching statistics for over 40 years.

Robert Calin-Jageman is a professor of psychology and the neuroscience program director at Dominican University and has been teaching and mentoring undergraduate students for nine years.

Introduction to the New Statistics

Estimation, Open Science, and Beyond

Geoff Cumming

Robert Calin-Jageman

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

Front cover: Artwork is from *Treasure* by Claire Layman (www.clairelayman.com), and is used with her generous permission.

First published 2017
by Routledge
711 Third Avenue, New York, NY 11107

and by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

Library of Congress Cataloging in Publication Data
Names: Cumming, Geoff. | Calin-Jageman, Robert.
Title: Introduction to the new statistics : estimation, open science, and beyond / Geoffrey Cumming and Robert Calin-Jageman.
Description: New York : Routledge, 2017. | Includes bibliographical references.
Identifiers: LCCN 2016007814 (print) | LCCN 2016008456 (ebook) |
ISBN 9781138825529 (pbk.) | ISBN 9781138825512 (hardback) |
ISBN 9781315708607 (eBook)
Subjects: LCSH: Estimation theory–Textbooks. | Mathematical statistics–Textbooks. | Confidence intervals–Textbooks.
Classification: LCC QA276.8 .C86 2017 (print) |
LCC QA276.8 (ebook) | DDC 519.5–dc23
LC record available at http://lccn.loc.gov/2016007814

ISBN: 978-1-138-82551-2 (hbk)
ISBN: 978-1-138-82552-9 (pbk)
ISBN: 978-1-315-70860-7 (ebk)

Typeset in Meridien LT Std
by Out of House Publishing

GC: For my grandchildren: Tom, Lachie, Odin, Pippa, Lucy, Erin, Zoe, and any who may follow.

RC-J: For Irina, Tavi, and Emilia, and for the many students who’ve made teaching these topics such a pleasure.

Brief Contents

Detailed Contents

Preface

This book is about how you can use limited data to draw reasonable conclusions about how the world works. Put more formally, this book is about *inferential statistics,* the art of using information from a *sample* to estimate what might be true about the world as a whole.

Inferential statistics is an exciting and powerful field! It's how physicians can test a new drug on a limited number of patients and then estimate how well the drug might work for the general public. It's how psychologists can test a new therapy on a limited number of clients and then estimate how well the therapy is likely to work for all patients with the same disorder. It's how pollsters can survey a limited number of likely voters and then estimate how much support there is for a candidate in an upcoming election. All this and so much more: It's no exaggeration to say that inferential statistics is at the very heart of our civilization's expanding ability to understand, predict, and control the world around us. This book will help you learn this amazing set of skills for yourself. With some work, you'll soon be able to make sound estimates from limited data, and you'll also be able to understand and critically assess the attempts of others to do so.

We hope this sounds enticing, but you may have heard that inferential statistics is dull, impenetrable, and confusing. Well—it doesn't have to be. This book teaches what we call the *new statistics,* an approach that we believe is natural and easy to understand. Here's an example. Suppose you read in the news that "Support for the President is 68%, in a poll with a margin of error of 3%." Does that seem particularly confusing? Hopefully not. You can immediately understand that the poll was conducted with a sample of voters, not by surveying everyone in the whole country. Then the pollsters applied inferential statistics to the results from the sample to determine that 68% is our best *estimate*, and that we can be reasonably confident that support in the whole population is within 68% ± 3%, which is the 95% *confidence interval* (CI). That, in a nutshell, is the *estimation approach* to inferential statistics, a key component of the new statistics. Of course, there's a lot to understand to be able to use estimation for yourself. We'll discuss issues like how to select the sample, how big the sample should be, and how to calculate and understand the *margin of error*. We'll also emphasize combining results from multiple studies, an approach called *meta-analysis*, which is a second key component of the new statistics. The important point for now is that the new statistics is not something you need be afraid of—learning from this book will take effort (see *Making the Most of This Book*, below), but we believe it will be easier and more intuitive than the way inferential statistics was taught in the past.

Although inferential statistics is very powerful, it can only lead to sound estimates if the data are collected and analyzed without *bias*. For example, you obviously couldn't trust poll data if certain types of voters were excluded or if the poll asked leading questions. Therefore, this book teaches not only inferential statistics, but also some approaches for minimizing bias while conducting research. Specifically, we emphasize *Open Science,* an evolving set of practices intended to reduce bias by increasing the openness of research and thus ensuring that research results are accurate, and worthy of our trust. Open Science emphasizes the stating of research plans and predictions in advance. Then, after you conduct the study, it emphasizes sharing materials, posting data publicly for others to analyze and use, and conducting replications to double-check your own work and the work of others. It's basically the old scientific method updated for the internet age—it's an exciting development that's leading researchers in many disciplines to change the ways they have traditionally worked. We introduce Open Science in Chapter 1, then throughout the book we discuss Open Science and other ways to limit bias.

Before we begin, you may be wondering: If this book teaches the new statistics, then what was the "old statistics"? In many fields, a more traditional type of inferential statistics, known as null hypothesis significance testing (NHST), has dominated. In Chapter 6, we'll explain this approach in detail. And throughout the book, we'll help you understand how to translate back and forth between estimation and NHST, so that when you read research conducted using NHST you'll easily be able to understand it using the estimation approach we take in this book. As you'll see, the estimation and NHST approaches are built on the same foundations and often lead to similar conclusions. We believe, however, that the estimation approach is not only easier to learn but also helps researchers make better judgments from their data. And this isn't just our opinion. An increasing number of journals and professional organizations are urging researchers to avoid problems with NHST by using the new statistics. This textbook is the very first of its kind to help beginning students learn the new statistics and Open Science practices. We hope you'll be excited to know, then, that working your way through this book will help put you right at the forefront of best research practices.

If You Are a Student

Especially if you are starting your first statistics course, welcome, and we hope you find it rewarding. As we've said, we hope you find estimation a natural way to think about research and data. We also hope that you'll find the section *Making the Most of This Book* helpful.

We hope you come to feel at least some of the passion we have for statistics. It's great to see a beautiful picture that makes clear what some data are telling us! Statistics is not really about mathematics, but about what data reveal, and examining pictures of data is usually the best approach. Perhaps this gives us new insights into the world, or how people think and behave. Welcome to the world of research, statistics, and informative pictures.

If You Are a Researcher or Instructor

You are probably very familiar with NHST and appreciate how well established it is. Between the two of us, we've taught NHST for almost 50 years and understand the challenges of changing. We believe, however, that all of us should carefully consider statistical reform issues, and decide how best to proceed in our own research areas and with our own students. Perhaps the new statistician's greeting will appeal: "May all your confidence intervals be short!"

Although adjusting the way you've been teaching statistics may seem daunting, we believe the work you put in will benefit your students tremendously. Not only should the new statistics be easier for your students to learn and use, but making the change should better prepare your students for a research world that's rapidly adopting Open Science practices. As an example of evolving standards, consider the new guidelines for authors introduced in 2014 by the leading journal *Psychological Science*. The editorial explaining the changes is at tiny.cc/eicheditorial and the new guidelines include this statement:

> *Psychological Science* recommends the use of the "the new statistics"—effect sizes, confidence intervals, and meta-analysis—to avoid problems associated with null-hypothesis significance testing (NHST). Authors are encouraged to consult this *Psychological Science* tutorial [Cumming, 2014, available from tiny.cc/tnswhyhow] by Geoff Cumming, which shows why estimation and meta-analysis are more informative than NHST and how they foster development of a cumulative, quantitative discipline. Cumming has also prepared a video workshop on the new statistics [available from tiny.cc/apsworkshop]. (From: tiny.cc/pssubguide accessed 1 July 2016.)

Psychological Science also encourages researchers to adopt Open Science practices, and offers badges to recognize preregistration of research plans, open materials, and open data (tiny.cc/badges). An editorial published in December 2015 (tiny.cc/lindsayeditorial) drew on Geoff Cumming's work to help justify further steps the journal was taking to increase the reproducibility of research it would accept for publication. Other journals and professional associations are making similar moves.

We are excited to work with you and your students to prepare for a future in which estimation and Open Science are the norm in our fields. We invite you also to consider the following section, *Making the Most of This Book*.

Intended Audience

This book assumes no previous statistical knowledge. It is designed for use in any discipline, especially those that have used NHST, including psychology, education, economics, management, sociology, criminology and other behavioral and social sciences; medicine, nursing and other health sciences; and biology and other biosciences. If you are teaching or studying in any such discipline, then this book is intended for you. We hope it serves you well.

KEY FEATURES OF THIS BOOK

An Estimation Approach Based On Effect Sizes and Confidence Intervals: The New Statistics

We're convinced that the new statistics, meaning estimation based on confidence intervals (CIs), is a better approach to data analysis. We believe it's easier for students to understand and more informative for researchers. Moreover, it's becoming widely used so it's vital that students and researchers understand it and can use it with their own data. We assume no previous statistical knowledge and focus on estimation from the very start, explaining it in simple terms, with many figures and examples. We also explain the traditional approach (NHST, null hypothesis significance testing) in Chapter 6 and use it alongside estimation in the subsequent chapters—with ample guidance for easy conversion back and forth between the two approaches.

Meta-Analysis, From the Very Start

Meta-analysis combines results from several studies and is a key component of the new statistics. It allows us to draw quantitative conclusions from a research literature, and these are what we need for evidence-based practice. We introduce meta-analysis in Chapter 1, then in Chapter 9 we explain it in a highly accessible way using the simple forest plot, without any formulas. This is the first introductory textbook to do so.

Open Science, From the Very Start

This is the first introductory textbook to integrate Open Science all through. The new statistics and Open Science are closely linked, and together are the way of the future.

Open Science promotes openness and replicability. Journals, funding bodies, and professional associations are revising their policies in accord with new Open Science standards. The basic ideas, including preregistration and open data, are easy for students to grasp. We discuss them throughout the book, with many examples—including examples of student research projects, which are often part of a highly valuable world-wide replication effort.

Promotion of Effective Learning and Studying Techniques

Recent research on how students study has identified how learning can be strikingly more efficient; for example, by having students work with meaningful examples and express things in their own words, and by asking them to keep retrieving earlier material. We've used these findings to guide the design of the book and the way we use numerous real research examples. We explain the effective learning techniques in the section *Making the Most of This Book*.

Compatible with Traditional or Flipped Classrooms

This book and all the materials provided at the website are designed to support effective learning, whether the course is organized along traditional lines or is based on a flipped classroom, in which students undertake assigned work with the book and its materials before coming to class.

Promotes Critical Thinking and Statistical Judgment

We emphasize careful critical thought about every stage of conducting research, rather than focusing on calculations. We provide essential formulas and many examples of data analysis, but our discussion of the numerous real research examples aims to develop students' deep understanding and confidence in making their own statistical judgments. The use of ESCI (Exploratory Software for Confidence Intervals) simulations, and guidance from the instructional videos, help students develop a deeper understanding and greater confidence in their own statistical judgment.

SUPPORTIVE PEDAGOGY

- Each chapter starts with pointers to what it contains, and closes with summary *take-home messages*. These summarize key points of the chapter, provide an overview, and serve as a study tool.
- Often in the text the student is asked to *pause, reflect, and discuss* intriguing issues. Research shows this is an effective learning technique, so we often ask students to write about a topic or discuss it with another student, to encourage critical thinking. These are also useful as prompts for class discussion or activities.
- Definitions of *key terms* are set off from the text. Many terms and expressions are also defined in the *Glossary* near the end of the book, which provides students with a quick reference and study tool. Lists of abbreviations and symbols appear at the very start of the book, and a list of selected formulas at the very end.
- *Exercises* and *quizzes* are placed throughout each chapter. Answers and our commentary, including much discussion of conceptual issues, are at the end of the chapter to allow students to test their understanding and quickly obtain feedback about their progress.
- Sidebars in the margins are visual markers highlighting key issues and points. This makes it easier for the reader to gain an overview and to find key points when reviewing for exams.
- Some of the exercises use the *ESCI software* (see below), for interactive learning and a visual grasp of concepts.
- We highlight common pitfalls, or things to watch out for. We call these *Don't fool yourself* (DFY) points, in recognition of Richard Feynman's sage advice that "The first principle is that you must not fool yourself". We hope these will help students avoid making such errors.
- In considering the NHST approach to data analysis, we explain important cautions that students always need to keep in mind, including troubles with the meaning of "significance" and what *p* values can and cannot tell us. These are the five *Red Flags*.
- There are *end-of-chapter exercises*, which often use real data sets and allow students to analyze real data as well as practice research judgment in realistic contexts. Our answers and commentary for these exercises are at the end of the book.

SOFTWARE SUPPORT

ESCI (Exploratory Software for Confidence Intervals) is the free software that goes with the book and is available for download on the book's website. You can readily use the book without ESCI, but it's designed to help by presenting statistical concepts vividly and interactively. Watch out for the dance of the means, dance of the CIs, and dance of the p values. You can use ESCI to analyze your own data, especially to calculate confidence intervals and create graphical depictions of your results. See the *Appendix* for more about ESCI. At the book website (see below) there are also guides for using the book with other software, in particular R and IBM's SPSS.

SUPPLEMENTAL RESOURCES ON THE BOOK WEBSITE

The book's website is: www.routledge.com/cw/cumming For easy typing, use tiny.cc/itns. The website is an integral part of the learning package we offer.

For Students:

- Reading guides that provide chapter-by-chapter guidance for making best use of the book and materials.
- Free download of ESCI, which runs under Microsoft Excel.
- Downloadable data sets, including those used in end-of-chapter exercises.
- Model manuscripts showing how to report your research in APA format.
- Glossary flashcards for practice and exam preparation.
- Guides for using other statistical software—in particular R and SPSS—for analyzing your own data and example data discussed in the text, and for answering end-of-chapter exercises.
- Videos that explain important concepts. Many of the videos show how to use ESCI to see concepts and analyze data.

For Instructors:

- An Instructor's Manual, which includes guidance for instructors teaching the new statistics, including additional reading suggestions and sample syllabi.
- Additional homework exercises (with solutions for instructors).
- Complete Powerpoints for each chapter, plus in-class activities with answer keys.
- Quiz and test bank questions.
- Downloadable images from the text.

CONTENTS

Here's a brief outline of what each chapter contains. The sequence is what we feel is best, but chapters can easily be used in a different order, in accord with the preferences of different instructors.

Chapter 1 introduces the process of asking research questions and using data to provide answers. It mentions Open Science and introduces many research and statistical concepts informally and intuitively.

Chapter 2 introduces further fundamental research ideas, says more about Open Science, and explains many terms.

Chapter 3 describes basic descriptive statistics, introduces the ESCI software, and uses ESCI to illustrate a number of ways to picture data.

Chapter 4 discusses the normal distribution and explains the basics of sampling. It uses ESCI simulations to explore sampling variability.

Chapter 5 explains CIs and effect sizes, and describes four ways to think about and interpret CIs. It also introduces the *t* distribution.

Chapter 6 discusses *p* values, NHST, and their close links with estimation.

Chapter 7 discusses the independent groups design for comparing two treatments. It describes both estimation and NHST approaches, including the *t* test for independent groups. It also introduces the standardized effect size measure, Cohen's *d*.

Chapter 8 describes the paired design, also taking both estimation and NHST approaches, including the paired *t* test. It discusses Cohen's *d* for the paired design.

Chapter 9 introduces meta-analysis using a visual approach based on forest plots, and provides many examples to illustrate its importance.

Chapter 10 has more on Open Science, then takes two approaches to planning studies: first, by finding *N* to achieve a desired precision of estimation and, second, by using statistical power.

Chapter 11 discusses Pearson correlation, *r*, and describes applications, including its value for meta-analysis.

Chapter 12 discusses linear regression, and explains how regression relates to correlation.

Chapter 13 uses proportions to analyze frequencies and discuss risk, and also introduces chi-square.

Chapter 14 takes a contrasts approach to analyzing one-way designs, and introduces one-way analysis of variance (ANOVA).

Chapter 15 continues the contrasts approach with two-way factorial designs, including discussion of interactions, and introduces two-way ANOVA.

Chapter 16 brings together earlier discussions of Open Science, and sketches a number of future directions, including longitudinal studies and big data.

The *Appendix* explains how to download and use ESCI, with numerous hints for getting the most out of the software. Look here to find which ESCI page you need to explore a concept, or to carry out calculations on your own data.

ACKNOWLEDGMENTS

GC: Numerous colleagues and students have contributed to my learning that has led to this book. I thank them all. Mark Burgman, Fiona Fidler, and Tim van Gelder are valued colleagues who have assisted in many ways. Eric Eich and Steve Lindsay, successive editors-in-chief of *Psychological Science*, and Alan Kraut and other leaders of the *Association for Psychological Science*, have guided and encouraged my work, and have also provided outstanding leadership toward better research practices. My treasured colleague Neil Thomason continues to provide inspiration, wonderful ideas, and an unswerving commitment to sound argument and clear communication. This book is way better because of his efforts. I'm grateful for the sustained dedication, intellectual rigor, and good humor of all members of the group Neil led to critique drafts, especially John Campbell, Keith Hutchison, Larry Lengbeyer, Michael Lew, Erik Nyberg, and Yanna Rider. At home I'm forever grateful to Lindy for her enduring support and encouragement.

RC-J: My contributions to this book would not have been possible had I not been fortunate enough to co-teach methods and statistics for the past eight years with outstanding and dedicated colleagues: Tracy Caldwell, Tina Taylor-Ritzler, Kathleen O'Connor, and most especially Rebecca Pliske, who mentored us all. Thanks for the many discussions, beers, and tears spent together pushing each other to do even better for our students.

Together we warmly thank Gabrielle Lehmann and Sarah Rostron for their painstaking work. We also thank the many researchers whose data we use, both those who provided their data openly to the world and those who kindly provided data in response to our request. We thank Gideon Polya for the artistry of his drawings and for his generosity in creating such a broad range to fit the book. We are grateful to the reviewers of draft chapters, who gave us valuable guidance that improved the book in numerous ways. Some remain anonymous; reviewers who agreed to be named were: Dale Berger, Claremont Graduate University; Bruce Blaine, Saint John Fisher College; Karen Brakke, Spelman College; Stephanie DeLuca, Virginia Tech; Thomas Faulkenberry, Tarleton State University; Franklin Foote, University of Miami; Catherine Fritz, University of Northampton (U.K.); Jon Grahe, Pacific Lutheran University; Rink Hoekstra, Groningen University (The Netherlands); S. Jeanne Horst, James Madison University; Fred Li, National Chai-yi University of Taiwan; Kevin Matlock, Humboldt State University; Christopher Nave, Rutgers University-Camden; Carrol Perrino, Morgan State University; Ed Purssell, King's College London; Robert Rieg, Aalen University (Germany); William Roweton, Chadron State University; Christopher Sink, Old Dominion University; Susan Troncoso Skidmore, Sam Houston State University; Jorge Tendeiro, Groningen University (The Netherlands); Patrizio Tressoldi, Università di Padova (Italy); Meg Upchurch, Transylvania University; Michael Van Duuren, University of Winchester (U.K.); Jordan Wagge, Avila University; Meg Waraczynski, University of Wisconsin-Whitewater. We warmly thank Toby Cumming, David Erceg-Hurn, Chuck Huber, Ryne Sherman, and others named at the book website, who developed valuable materials to go with this book—all are free downloads from that website. We thank David Erceg-Hurn again, for assistance with Chapter 16. We are greatly indebted to Debra Riegert, senior editor, and Rebecca Pearce, our development editor, both at Routledge, and Tom Moss Gamblin, our copy editor, for their outstanding support and professionalism. After that long list of wise advisors we must say that any remaining errors and weaknesses are all our own work. Please let us know of any you discover. Finally, thank you for joining us on this exciting journey of helping to shape how research will be done. May this book serve you well on that journey.

ABOUT THE AUTHORS

As you can see on the front cover, there are two of us, but we have decided, starting with the next section *Making the Most of This Book*, to write as if we were one, and to use "I" often. A particular "I" may refer to either of us, but usually it's both. We hope this gives a more informal and personal tone, which is how we both like to discuss ideas with students.

Geoff Cumming is professor emeritus at La Trobe University, Melbourne, and the author of *Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*, published by Routledge in 2012. He has taught statistics for more than 40 years at every level, from introductory to advanced. His statistics tutorial articles have been downloaded more than 370,000 times. See: tiny.cc/errorbars101 and tiny.cc/tnswhyhow. The *Association for Psychological Science* has published six videos of his highly successful workshop on the new statistics (see: tiny.cc/apsworkshop). His main research interests are the investigation of statistical understanding and promotion of improved statistical practices. A Rhodes Scholar, he received his Doctorate degree in experimental psychology from Oxford University.

Robert Calin-Jageman is a professor of psychology and the neuroscience program director at Dominican University. He has taught statistics and mentored students in psychological science for nine years, publishing with 16 undergraduate co-authors (so far). His research focuses on how memories are formed and forgotten. He has also been active in exploring the replicability of psychological science and promoting Open Science. He received his PhD in biological psychology from Wayne State University.

Making the Most of This Book

While writing this book I've been fascinated by recent research on practical ways that people can learn more efficiently. Before saying more about that, I invite you to consider the common learning techniques listed in Table 0.1 and record how often you use each strategy and how effective you judge each to be. If you like, have a guess at the third column—what research tells us.

> Throughout the book, I use this little picture when I'm asking you to pause and do something. It really is worth giving it a try before reading on.

Table 0.1 Learning Techniques

Technique	How often you use it	How effective you think it is	How effective research finds it to be
1. Reread the textbook.	_____	_____	_____
2. Highlight key points.	_____	_____	_____
3. Write summaries.	_____	_____	_____
4. Study one topic thoroughly before moving on.	_____	_____	_____
5. Use tests, including self-tests. Ask and answer questions, alone or in a group.	_____	_____	_____
6. Retrieve material from memory, even when not fully mastered and retrieval is difficult. Correct any errors.	_____	_____	_____
7. Move on to the next topic before mastering the current topic. Try to figure out the next thing for yourself.	_____	_____	_____
8. Practice reflection: Identify important ideas, invent examples, and make links with earlier material.	_____	_____	_____
9. Distribute study activities over time. Retrieve material later, then again after a delay. Then again.	_____	_____	_____
10. Interleave study activities. Mix things up. Study a range of different topics; use a variety of activities.	_____	_____	_____

If you skipped to here, I urge you to go back to the table and think about all the items on the left, and how useful you think they are.

Research on how students study and learn has in recent years found startling results. There's good evidence that most of us can do way better. Before I go on, here's an analogy: I recently read an article (tiny.cc/runfast) about a long-distance runner who learned from a coach how to breathe differently: more from the diaphragm and in a different rhythm. It took effort and practice, but with the new technique he shattered his previous personal best for the marathon.

I found that story intriguing and instructive, because it just goes to show that even something you already think you are good at (like breathing!) can be drastically improved through science. Well, according to a range of research from psychology, it's the same for study skills.

I'll summarize a few of what I see as important lessons from the research, then I'll revisit Table 0.1, and suggest some first steps you can take.

Get Motivated and Engaged

Of course, anyone learns better if they feel engaged, and see the material as relevant for themselves. I'll do my best to explain why I believe statistics are so important. There will be numerous real world examples that will help you see that statistics are part of all our lives, and persuade you that they really matter—to you, me, and all of us. That's one reason I find them so fascinating. I'll also try to provide lots of interesting, even enjoyable activities. I invite you to seek reasons relevant to you for getting engaged and I hope you find that statistical understanding helps not only in your studies, but also in the wider world. Unfortunately, reports in the media and discussions of current affairs often invoke research findings, but draw unjustified conclusions. Distortions and misleading claims can be tricky to spot, but basic statistical understanding can be a great help—you can enjoy using your data detective skills.

Thinking of a different approach to motivation, you no doubt appreciate that finding the right friends to work with can help—in person or in cyberspace. I'll often invite you to pause and reflect on some issue; you may find that discussion with others is a good strategy.

I've seen many students at first skeptical about statistics and their own abilities who become absorbed by the challenges of understanding research and drawing conclusions from data. I hope you also can become absorbed by these challenges.

Spend Time on Task

Seek ways to keep motivated and engaged, to help you put in the time.

Who would have guessed it? We need to put in the hours. Motivation and engagement help us find the time, and keep up the concentration. Working with others may help. If rewards work for you, then allow yourself coffee, chocolate, or a walk on the beach when you finish a chapter, or master a tricky idea—whatever it takes to keep up the motivation and put in the time.

Build Your Statistics Confidence

For years I asked students at the start of my introductory course to rate their attitude towards statistics on a scale from "no worries" to "blind panic". Then I'd invite especially the blind panic students to extra lunchtime meetings where we discussed any statistical questions they cared to ask. They were usually reassured to find others who shared their concerns, and also that working through the basics, with many pictures, led them to feel increasingly confident. If you are initially anxious, I hope the many examples and pictures in this book, and the interactive simulations we'll use in ESCI, will similarly reassure you and help you build your confidence. Maybe find some others with initial doubts, and work at it together.

Use the Most Effective Learning Techniques

Before reading on, if you haven't written down your responses for the blanks in Table 0.1, please do so now. One of the main messages of this section is that it's highly valuable to think and do, as well as read. My request that you think about your response for all the blanks in the table is a first go at that.

Many students rely on rereading and highlighting, but these strategies may give only superficial learning that won't last.

Surveys suggest that enormous numbers of students, perhaps the majority, rely on 1–4 in the table. However, research indicates that these techniques are generally not the most effective use of time. They often give the illusion of understanding—you seem to be working hard, focusing on the material, and grasping it. However, the learning is often superficial and won't persist because it's not sufficiently elaborated and integrated with prior knowledge, or linked with examples and practical application.

By contrast, 5–10 have been found effective for achieving learning that's deep and enduring. One key idea is *retrieval*, which means closing the book and trying to bring to mind the main points, and maybe writing some bullet points in your own words. You could practice now for this section. Then open the book and check. It's fine if you miss lots and make mistakes—the great value is retrieval itself, even when you only partly grasp something. Come back to it, retrieve again, and enjoy doing way better!

Work at a challenging retrieval to change your brain and get smarter.

In other words, a valuable learning activity is to work at retrieving something, even if it's only half-learned, half-understood. Persist, do your best, compare with the correct answer, then come back later and retrieve again. It can be difficult working with not-quite-understood material, but it's effective, even if it doesn't seem so at the time. Researchers who study retrieval suggest that achieving a difficult retrieval actually changes your brain and makes you smarter. In summary, the slogan is: "Don't read again, retrieve again".

If you are a runner, maybe think of retrieval as the studying equivalent of diaphragm breathing—a great way to do better that, with a bit of effort, anyone can learn, but which most people don't appreciate.

- Don't read again, retrieve again.
- Make it your own.

I summarize 5–10 in the table as "Make it your own". Take any new idea and express it in your own words, make a picture, link it back to things you know already, think up an example, then a crazy example, try explaining it to someone else—do whatever helps to make it your own. Then later test yourself—do your best to retrieve it. Then tomorrow retrieve it again.

Change a Fixed Mindset to a Growth Mindset

Fixed mindset: The belief that my capabilities are more or less fixed, whatever I do.
Growth mindset: The belief that effort, persistence, and using good techniques can help me learn more successfully and become more capable.

A further key idea is the distinction between a *fixed mindset* and a *growth mindset*. Carol Dweck and colleagues have demonstrated that helping students adopt a growth mindset can be a highly effective way to help them learn better and achieve more. Here's how Dweck describes the two mindsets:

> In a fixed mindset students believe their basic abilities, their intelligence, their talents, are just fixed traits. They have a certain amount and that's that.... In a growth mindset students understand that their talents and abilities can be developed through effort, good teaching and persistence. They don't necessarily think everyone's the same or anyone can be Einstein, but they believe everyone can get smarter if they work at it. (Carol Dweck, tiny.cc/dwecktalk)

I've mentioned three important ideas about learning.
...before reading on, you may care to close the book and practice retrieval...

- Retrieval is valuable, even when it is difficult, even when you don't fully grasp the material.
- To "make it your own" by elaboration, discussion, or in any other way can be highly effective.
- Adopting a growth mindset can motivate effective learning efforts.

Reflect on how the three relate, and how you might make use of them. Explain your thinking to someone else.

Make It Stick

Make it stick: The science of successful learning is a great book by Brown, Roediger, and McDaniel (2014). It describes the research findings on effective learning, and uses real stories to make the main recommendations intuitive and vivid. You may find reading the book helpful. For a start, you could try one or more of the following:

- Browse the book's website, at makeitstick.net At the "About" tab, find a paragraph that summarizes the main message. At the "Contents" tab go to a page for each chapter with a one-paragraph summary and a box with a brief version of that chapter's story. Which is your favorite story? (Mine is about Michael Young, the Georgia medical student.)
- Watch this video: tiny.cc/misvideo

Writing Take-Home Messages

Each chapter in this book ends with take-home messages, and towards the end of each chapter I'll encourage you to write your own, before reading mine. Make that part of your doing, not just reading.

Here's a first chance to write your own take-home messages. Think (or look) back over this *Making the Most of This Book* section, and choose what, for you, are the main points. I've written four, but you can write as few or as many as you wish.

Pause, write, discuss, before reading on…

It really is worth closing the book and bringing to mind what you think are the main messages.

No, don't read on yet…

Take-Home Messages

- **Find ways to engage.** Find whatever strategies work for you to find motivation, to relate statistical ideas to your own interests, and to keep engaged—so you can keep putting in the time. Work with others if it helps.
- **Make it your own.** Use a mix of activities—asking and answering questions, discussing, writing in your own words, using the software, applying the ideas—as you seek to make sense of it all, and to make the material your own.
- **Retrieve and retrieve again.** Retrieve again, rather than read again. Retrieval that's challenging can give good learning, change your brain, and make you smarter. Then retrieve again later, then again later.
- **Adopt a growth mindset.** Use good learning techniques, seek guidance, and persist, and you will learn and become more capable.

1

Asking and Answering Research Questions

A large part of science is asking questions, then trying to find data that can help answer them. In this chapter I'll use an everyday example to illustrate the general idea of asking and answering questions. I'm hoping you'll find the example pretty intuitive—you may discover that you already have a good idea of how data can show us how the world works.

This chapter introduces:

- A simple opinion poll that illustrates how data can help answer a research question
- The scientific research process, from asking questions to interpreting answers
- Pictures that help us understand data
- Basic ideas of *population* and *sample*, and of *estimate* and *margin of error*
- The idea of a *confidence interval*, a vital part of the answer to our research question
- *Open Science*: An approach to research that tackles some of the ways that data can mislead, and emphasizes the need to think carefully about every stage of the research process
- The value of *replication* studies that repeat research to check its accuracy, and of *meta-analysis* to combine results from a number of similar studies

Words in italics, like *population*, are terms I'll define later. For the moment, read them as normal English words, although you could, if you wished, consult the Index or Glossary at the back of this book. Also, be sure to explore the book's website, which has lots of goodies, including videos. Make it a favorite or bookmark: www.routledge.com/cw/cumming or, for easy typing: tiny.cc/itns

A SIMPLE OPINION POLL

Here's the example—a simple opinion poll. You read this in the news:

> *Public support for Proposition A is 53%, in a poll with a 2% margin of error.*

Let's say Proposition A proposes a law requiring serious action on climate change by reducing the use of fossil fuels and switching to renewable energy. Soon there will be a state-wide vote to determine whether the proposition becomes law. You and your friends have set up a website explaining why the proposition is a great idea, and are eager to know the extent of support for it among likely voters. Therefore, our question is:

What's the support for Proposition A in the population of people likely to vote?

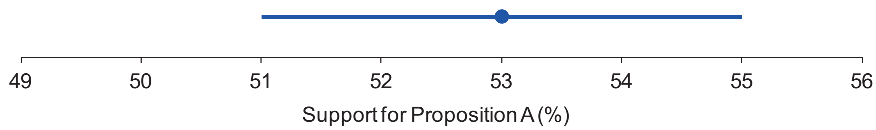

Figure 1.1. Support for Proposition A, in percent, as reported by the poll. The dot marks the point estimate, and the two lines display the margin of error (2%) either side of the dot. The full interval, from 51% to 55%, is the 95% confidence interval.

The poll's answer is:

Estimated support in the population of likely voters is 53±2%.

This result from the poll is displayed in Figure 1.1. Does this make you happy? Probably yes, because estimated support is greater than 50%, so the proposition is likely to pass, although perhaps only by a small margin.

A Thumbnail Sketch of Research

Here's a slightly fuller account of the poll example, which illustrates a common way research proceeds.

1. Ask a research question. *What's the support for Proposition A in the population of people likely to vote?*
2. Design a study to collect data that can answer the question. *Design a poll that will ask a sample of intending voters about their support for Proposition A.*
3. Carry out the study and collect the data. *Choose a sample of intending voters and ask them about their support for the proposition.*
4. Apply statistical analysis to picture and describe the data, and provide a basis for drawing conclusions. *Calculate that 53% of people in the sample say they support the proposition. Use knowledge of the poll design, especially the size of the sample, to calculate from the data that the margin of error is 2%, and therefore the confidence interval extends from 51% to 55%. Prepare Figure 1.1.*
5. Draw conclusions about what the data tell us in answer to our original question. *We take the 53% as the best estimate the data can give us of support in the population of likely voters, and the 2% margin of error as indicating the uncertainty in that estimate. In the figure, the dot marks the best estimate, and the confidence interval indicates the range of uncertainty.*
6. Interpret the results, give a critical discussion of the whole study, and prepare a report. Think about the next study. *Most likely, the true level of support among intending voters is within the interval from 51% to 55%, therefore the proposition is likely to be approved—although it may not be.*

Of course, that's a mere sketch of the research process. You may have many questions: "How do we choose the sample?", "How large a sample should we use?", "How do we calculate the margin of error?", "How should we interpret the 95% confidence interval in Figure 1.1?" We'll discuss answers to these and many other relevant questions throughout this book.

Where in the process do you need to know about statistics? Most obviously at Step 4, to calculate the confidence interval. However we need statistical understanding at every single one of the steps, from formulating the question and designing a study, to interpreting the results and making a critical evaluation of the whole study. Throughout the book, whatever statistical idea we're discussing, always bear in mind the whole research process. Statistical ideas are needed at every stage.

Perhaps the most amazing thing about statistics-based research is that the process sketched above permits us to study just a relatively small sample of people, and yet draw conclusions that might apply broadly, in some cases to the whole world! Statistical techniques give us a sound basis for analyzing sample data and making inferences—drawing conclusions—that sometimes apply very broadly. Yes, there's always uncertainty, but our analysis can tell us *how much* uncertainty. That's the magic of statistics.

Scientists have used more fully developed versions of this framework—my thumbnail sketch above—and statistical understanding to discover much of what we know about people and the world. Among a vast number of examples, such research has told us about

- how effective cognitive-behavior therapy can be for depression;
- how much ice mass the Greenland icecap is likely to lose in the next two decades; and
- the extent that having more friends can lead to improved learning in elementary school.

You may not wish to be a researcher, although you may have the chance to participate in worthwhile research as part of your course. In any case, to appreciate how such knowledge was gained requires statistical understanding. Beyond that, to be a critically aware citizen means being able to understand data reported about society and our immediate world, and to know what searching questions to ask. Statistical understanding is essential for that.

Conducting research properly can be tricky—Chapter 2 is about lots of ways we can fool ourselves. We'll see examples where wrong statistical choices cost lives, and poor research practices cause widespread misconceptions about what's true in the world. It's essential to use the best research and statistical practices we can, and to use them correctly. And always to think carefully about what any data really tell us.

Intuitions About the Poll

I invite you now to think informally and intuitively about the poll example. Here are some points worth thinking about:

- Our question is about the whole *population*, meaning all people likely to vote on the proposition.
- The poll couldn't ask everyone, or even most people, in the population, so it took a *sample* from the population, and asked people in the sample whether they supported the proposition.
- If the sample was chosen in a fair and unbiased way, it's probably representative of the population, so we can take the sample results as a reasonable *estimate* of support in the population.
- There is some unknown *true* level of support in the population. Our best *point estimate* of that is 53%, the support the poll found in the sample.

We use results from a sample to *estimate* something about a population.

The *point estimate* is the best single value the data can give us for what we're estimating about the population.

- We calculate the *margin of error* (2%) as the likely greatest error in the point estimate. In other words, 53% is unlikely to be more than 2% away from the true value.
- Most likely the true value of support in the population lies in the range 53±2%, or [51, 55]. That's the full extent of the interval displayed in Figure 1.1.

If at least some of those points match your intuitions, well done! You are well on the way to appreciating the basic logic of research that asks and seeks to answer questions of this kind.

The *95% confidence interval* (CI) is a range of values calculated from our data that, most likely, includes the true value of what we're estimating about the population.

We call that range of values, [51, 55], our *95% confidence interval*, abbreviated as "CI." It's an interval inside which the true value is likely to lie, which means we can say:

> *We are 95% confident the interval [51, 55] includes the true level of support in the population.*

The *margin of error* is half the length of the 95% CI, and the likely greatest error in the point estimate.

The margin of error is our measure of the *precision* of estimation. A small margin of error means a short CI and a precise estimate.

The 95% CI extends from 51% to 55%, so the *margin of error* (2%) is half the length of the CI, as Figure 1.1 illustrates. The "95%" means we are not guaranteed that the CI includes the true value. However, most likely it does, assuming that the poll was carried out well—later there's much more on what it means to carry out studies well. You might be dissatisfied with "most likely"—we would prefer to be certain. However, research studies rarely, if ever, give definitive answers to our questions, so we must be willing to think about uncertainty and not fool ourselves by looking for certainty. The great value of a CI is that it *quantifies* uncertainty—its length is a measure of the extent of uncertainty in our point estimate.

We can also say that the CI tells us how *precise* our estimate is likely to be, and the margin of error is our measure of precision. A short CI means a small margin of error and that we have a relatively precise estimate—the 53% is likely to be close to the population value. A long CI means a large margin of error and that we have low precision—the 53% may be further from the true value.

Figure 1.2 illustrates how values near the center of a CI are most *plausible* for the true population value, and how plausibility decreases toward the ends of the CI and then beyond the CI.

The curve in Figure 1.2 illustrates how *plausibility* or *likelihood* varies across and beyond the interval. Values around the center of the CI, say around 52% to 54%, are the most plausible, the most likely, for the true value in the population. Values toward either end of the CI are progressively less plausible, and values outside the interval even less so. The further a value lies outside the CI the more implausible it is. In other words, values near the point estimate are relatively good bets for where the true value lies, and values progressively

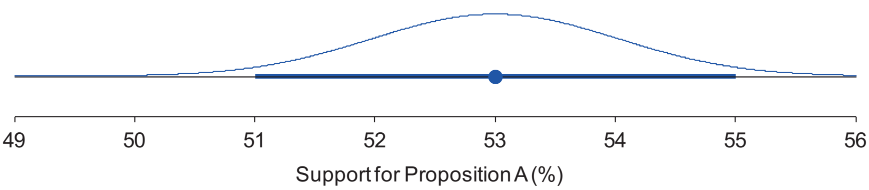

Figure 1.2. Same as Figure 1.1, but with the addition of the smooth curve that pictures how likelihood varies across and beyond the 95% CI. Likelihood, or plausibility, is represented by the height of the curve above the CI and the fine horizontal line.

further away from the point estimate are progressively less good bets. In other words again, the likelihood that a point is the true value is greatest for points near the center of the interval and drops progressively for points further from the center.

Note in particular that there's nothing special about the endpoints of the CI. Always keep in mind the smooth likelihood curve of Figure 1.2, which applies to just about any CI and illustrates how likelihood, or plausibility, varies across and beyond the interval.

Why 95%? Good question. You might occasionally come across other CIs, perhaps 90% or 99% CIs, but 95% CIs are by far the most common. It's almost always best if we agree on which CI to use, so I recommend we follow convention and use 95% CIs, unless there are very strong reasons for using something different. I'll routinely use 95% CIs, so if I mention a CI assume 95% CI unless I say otherwise.

Estimates and Estimation

We can refer to a CI as an *interval estimate* because it's an interval containing the most plausible values for the population value, as illustrated in Figures 1.1 and 1.2. The main approach to data analysis in this book is based on point and interval estimates, and you won't be surprised to hear that this general approach is referred to as *estimation*. It's a highly informative way to analyze data. I hope you'll find it a natural and easily understood way to report results and draw conclusions from data.

The CI is our *interval estimate* of the population value of interest.

To use estimation, what type of research questions should we ask? We could ask "Is Proposition A likely to pass?" but this suggests a yes-or-no way of thinking about the world, and that a yes-or-no answer would be sufficient. However, we're much less likely to fool ourselves if we think about the world in a *quantitative* way, and therefore ask quantitative questions, such as "What's the extent of support?" or "How great is the support?" Such questions call for quantitative answers, in terms of percent support, which are more informative and therefore preferable. Using estimation, we should always express research questions in quantitative rather than yes-or-no terms. We should ask "To what extent…?", "How much…?", or similar questions, then appreciate the quantitative, informative answers.

Express research questions in estimation terms. Ask, for example, "How much...?", or "To what extent...?"

Making Your Interpretation

A couple of paragraphs back I said that, after calculating point and interval estimates, we need to "draw conclusions from data". After reporting a CI, you should give us your interpretation—what do the values mean, in the context of the study? In our example, what might you say about the poll result? We can summarize the result as "53% support, 95% CI [51, 55]". What do those values imply, considering the impending vote on Proposition A?

I'll often ask you questions like that. You can read straight on and see my answer, but it's much better to look away and think of your own. Write it down! Even better—call a friend for a chat before you write. I'll use the pause and think logo, as below, to suggest a good moment for you to pause, think, discuss, and write. But be encouraged to pause, chat, and write whenever you like. Often is good.

So, have you written down your answer?

As I explained earlier, in "Making the Most of This Book", research tells us that learning is much better if you write things in your own words, even if you feel you are making a wild guess.

The *limits* of a CI are its two endpoints.

Use judgment to interpret the point estimate and CI, in the particular context.

Think of the campaigns for and against the proposition. Think of what 51% and 55%, the *limits* of the CI, might mean—"limits" is what we call those two endpoints of a CI.

You might suggest that a 2% margin of error is not bad—we'd always like a smaller margin of error, meaning higher precision, but the result we have is useful. You might say the CI indicates that all plausible values for the true level of support in the population are greater than 50%, so we can feel confident the proposition will pass. However, you might also worry that a strong "No" campaign has been running, and there's enough time for a few percent of voters to be persuaded to change their minds—the poll suggests that such a small change could tip the result. You'd therefore encourage your friends to make a final effort to keep support high, perhaps by stepping up your social media campaign. The important point is that how you interpret the result requires you to think about the context and implications. You need to consider both the point estimate and the CI, then go beyond those mere numbers and give your judgment of what they mean in the particular situation. One aim of this book is to help you build your confidence to acknowledge uncertainty and make interpretations based on judgment.

FURTHER INTUITIONS

Here are some questions to test your intuitions further:

If we ran the poll again, with a new sample, but using the same procedure and as close as possible at the same time, what's the likely result?

Pause... think... call... chat... write...

Instead of my answer, here's another question:

Suppose we ran the poll again, with a much larger sample, what do you think is likely to happen to the margin of error? With a much smaller sample? Which result is most useful?

Hint: Think of an enormous sample. A tiny sample. It's often a good strategy to think of extreme cases.

Sampling variability is variability in results caused by using different samples.

Larger sample, shorter CI; smaller sample, longer CI—all else remaining the same. A sample four times as large gives a CI about half as long.

For the first question you probably quickly appreciated that a second poll would be very unlikely to give exactly the same point estimate. However, it's likely to give a similar estimate, not too far from 53%. Most likely, it will give a value in the interval [51, 55], which is our 95% CI from the original poll. *Sampling variability* is the name we give to the variation caused by using different samples. It's the variation from poll to poll—when we assume they are all carried out at the same time and in the same way, but using different samples. The CI gives us a good idea of the extent of sampling variability.

For the second question, a much larger sample is likely to give a result that's closer to the true value in the population, meaning its CI will be shorter, its estimate more precise. In fact, if we used a sample four times as large, the CI would probably be about half the length. On the other hand, a smaller sample is likely to give us a longer CI.

Do we prefer our 95% CIs to be long or short?

If you like, reward yourself (chocolate? coffee?) for taking a break to think about the question.

That's an easy one: short, of course. A short CI means our sample estimate is most likely very close to the true value—the margin of error is smaller, the precision is greater. That's good news. That's why we go to the expense and trouble of running a poll with a large sample—to get a smaller margin of error, meaning a short CI.

From now on I'm going to refer to the margin of error as MoE, which I pronounce as "MO-ee", although you can say it as you wish. So MoE is half the length of a 95% CI, and MoE is our measure of precision.

MoE stands for margin of error.

Quiz 1.1

1. A company is interested in how satisfied its customers are. To help find out, 50 customers are randomly selected to take part in a survey. Which of the following is true?
   a. The 50 customers surveyed are the sample, all the company's customers are the population.
   b. Whatever result is found in the sample will be exactly the same in the population.
   c. The company would be better off sampling only 10 customers, as this would produce less uncertainty about overall customer satisfaction.
   d. All of the above.
2. A confidence interval (CI) expresses
   a. a range of plausible values for what is most likely true in the population.
   b. our uncertainty about what is true in the population.
   c. the fact that results from a sample may not perfectly reflect the population, due to sampling variability.
   d. all of the above.
3. You read a poll result that says "62±4% of likely voters support the referendum". What is the ±4% part?
   a. This is the point estimate for referendum support.
   b. This is the population for referendum support.
   c. This is the margin of error (MoE).
   d. This is the sample size.
4. If the poll in Question 3 was conducted well, which of these results would be most *un*likely?
   a. The referendum passes with 66% support.
   b. The referendum passes with 63% support.
   c. The referendum passes with 61% support.
   d. The referendum passes with 55% support.
5. We calculate a CI from the sample / population and use it to tell us about the sample / population. Half the length of the CI is called the ____________________, with abbreviation ______.
6. If *N*, the sample size, is made four times as large, the CI length will be about __________ what it was before, the precision will be lower / higher, and the researcher is likely to be more / less happy.
7. Make for yourself at least three further quiz questions, then give your answers. Swap with a friend.

Next, some exercises. It's so important to be thinking and doing, not just reading, that I've included exercises throughout the text. These in-chapter exercises often introduce new ways of thinking about what we've been discussing, or even new concepts. They are not just there for practice, but often play an important part in the main discussion,

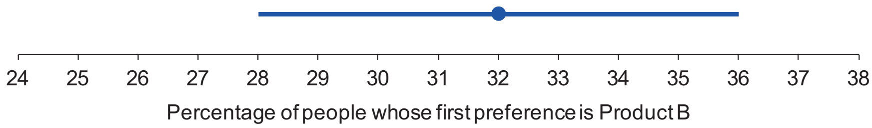

Figure 1.3. First preference for your product (Product B), marked by the dot, with 95% CI.

so please be encouraged to read and think about them all. You'll find my commentary and the answers at the end of each chapter.

1.1 Your company has decided to branch out into beauty products and has produced Invisible Blemish Cream. (I didn't invent that name—who cares about blemishes you can't see?!) A survey assessed people's first preference when given a choice of your company's cream and three competing products. For the test, the products were given the neutral labels A, B (your cream), C, and D. Figure 1.3 displays the results.

   a. What is the point estimate of the first preference for your product? The interval estimate? The margin of error?
   b. What is the population? Who would you like to have in the sample?
   c. Make two statements about the level of first preference for your product in the population.

1.2 If people chose randomly, you would expect 25% first preference for your product. Is your product more strongly preferred than this? Why or why not?

1.3 How could you achieve a CI about half as long as that shown in Figure 1.3?

CAREFUL THINKING ABOUT UNCERTAINTY

In later chapters we'll discuss important ideas raised by this poll example, including sampling, point estimates, and CIs, and how to use sample data to make conclusions about a population. We'll see definitions and formulas, and discover how to calculate 95% CIs. But for now I want to continue our informal discussion.

It's vital when reading a result like "53% with a 2% margin of error", or seeing a picture like Figure 1.1, to appreciate immediately that the result—the percentage support in the sample—could easily have been different. The CI gives us an idea of how different it might have been, if all details of the poll remained the same but we'd happened to choose a different sample. Sampling variability is one source of uncertainty with our results, and statistical procedures—calculating the CI—quantifies that for us.

However, beyond sampling variability there's virtually always additional uncertainty, which is much harder to pin down. It can have different causes in different situations, and usually there's no statistical formula to quantify it. We need careful critical thought to identify problems that might be contributing additional uncertainty.

> Beyond sampling variability there may be uncertainty arising from incomplete or biased reporting, or other causes.

Thinking of the poll example, here's one problem that could be contributing additional uncertainty. The news website where we read the poll result reported only this single poll, but perhaps there were other polls taken at the same time that it didn't report? If so, did it report the largest or best, or were the results we saw *selected*—by which I mean chosen to reflect some preference or bias? Other polls may have given different results, and our news source may have chosen to report just this particular poll because it liked its message. If a news source *selects* what to report from a number of results, then we can't draw confident conclusions from what it does report. There's an unknown amount of extra uncertainty and we can no longer be 95% confident the CI based on the poll results includes the true value. We need to seek out the most trustworthy news sources, seek out any other poll results, and note any signs of bias in a particular news source. In general, we need to think carefully and critically about any results, especially by asking:

> *Do we have the full story, or were these results selected in some way that might give a misleading message?*

If we suspect such selection, we can't draw confident conclusions.

You might also ask how the sample of people to be polled was obtained—we need to have confidence that it's likely to be reasonably representative of the whole population of intending voters. You could also be thinking that a poll result can be influenced by the wording of the question people are asked, by the communication channel used—phone or email or face-to-face—and especially by the proportion of people in the sample who cannot be contacted or refuse to respond. These are all good thoughts. Reputable polling companies have refined their procedures to minimize all these problems, but we still need to be alert to such additional ways that poll results may be uncertain. To help us assess the results, we need to have full details of the poll, including information about how it was conducted, what questions were asked, how the sample was chosen, and how many of the people in the sample answered the questions. In general, to have confidence in any data we need to ask:

> *Do we have full details about how the data were collected?*

THE REPLICABILITY CRISIS AND OPEN SCIENCE

Those two questions (*Were the results selected in a way that might mislead? Do we have full information?*) mark our first encounter with *Open Science*, a central idea that we'll meet often in this book. We can only have full confidence in conclusions from research when a number of Open Science requirements are met, and these questions express two of those.

The first two requirements for *Open Science* are: (1) avoid misleading selection of what's reported, and (2) report research in full detail.

"Open Science" comprises a number of practices designed to improve research. It has emerged only in the last few years and is still developing. It has largely been prompted by the *replicability crisis*—the alarming discovery that a number of widely known and accepted research findings cannot be replicated. In other words, when researchers repeat the earlier studies that reported the findings in question, they get clearly different results. In one dramatic example, a company wanting to develop new cancer therapies examined 53 findings from cancer research that looked promising. The company first attempted to confirm each finding by running a *replication* study, meaning a study designed to be as similar as possible to the original study that reported the promising result. In only 6 of the 53 cases (Begley & Ellis, 2012) could they confirm the main findings. That's terrible! It seems that some well-accepted research findings are simply wrong.

A *replication* is a repeat of an original study, similar to the original but with a new sample.