Introduction to the New Statistics: Estimation, Open Science, and Beyond

Book · January 2017

DOI: 10.4324/9781315708607

2 authors:

Geoff Cumming
La Trobe University

Robert Calin-Jageman
Dominican University

Second Edition

Preview copy of Table of Contents and 1 Chapter.
For instructor and student resources
and to request an inspection copy, please see:

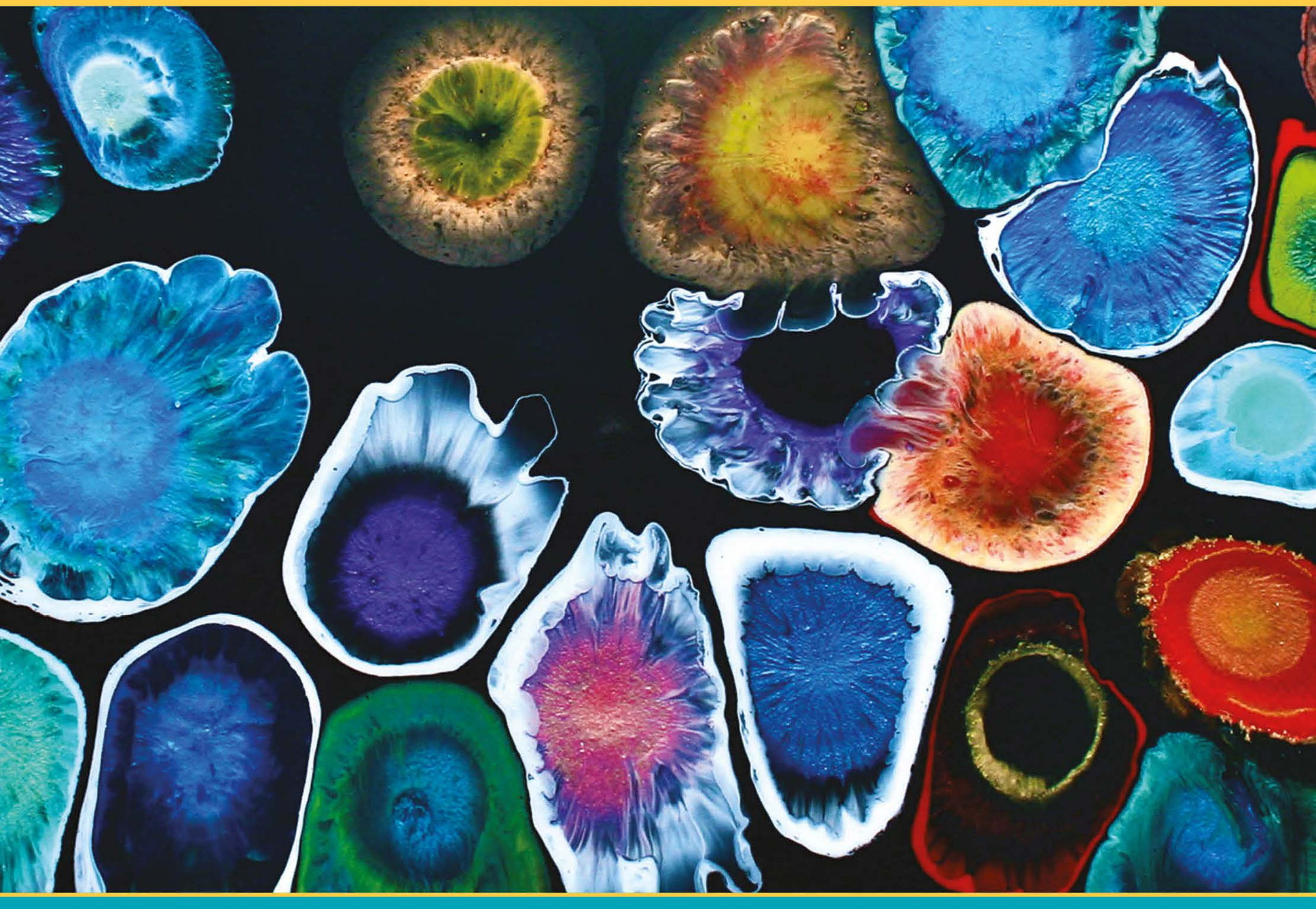
https://www.routledge.com/Introduction-to-the-New-Statistics-Estimation-Open-Science-and-Beyond/Cumming-Calin-Jageman/p/book/9780367531508

FREE INSTRUCTOR & STUDENT RESOURCES

ROUTLEDGE

Introduction to the New Statistics

ESTIMATION, OPEN SCIENCE, AND BEYOND



GEOFF CUMMING AND
ROBERT CALIN-JAGEMAN

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
COS	Center for Open Science (cos.io)
df	degrees of freedom
DFY	don't fool yourself
DR	diamond ratio
DV	dependent variable
ES	effect size
esci	Estimation Statistics with Confidence Intervals
H_0	null hypothesis
H_1	alternative hypothesis
HE	hypothesis evaluation
IPD	individual participant data, from studies in a meta-analysis
IQR	interquartile range
IV	independent variable
LL	lower limit of a CI
MoE	margin of error (length of one arm of a 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
UI	user interface
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	sum of (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
Mdn	median
N	size of a single group, or grand total if more than one group
n	group size when more than one group
P	proportion
p	p value, in NHST
r	Pearson's correlation
s	sample SD
s^2	sample variance
s_{av}	standardizer for Cohen's d, paired design
s_{diff}	SD of the differences, paired design
s_{p}	pooled SD
t	variable, often with t distribution
$t_{.95}(df)$	.95 critical value of t for stated df
V	variance
X	dependent variable, predictor variable in regression
X	integer, numerator of a proportion
Y	second DV in correlation, predicted or criterion variable in regression
$\hat{Y}$	regression prediction of Y
z	variable, often with normal distribution
$z_{.95}$	z for central .95 area under normal distribution
?	$.05 < p < .10$
*	$.01 < p < .05$
**	$.001 < p < .01$
***	$p < .001$

"Cumming and Calin-Jageman are psychology's New Statistics evangelists, and with this text they demonstrate how to train our field's newest scholars. This book explains the statistical estimation process with patience and clarity. Just as importantly, each section keeps students in mind. The authors anticipate learners' misconceptions, build quantitative reasoning with 'eyeballing' tips, and offer more practice just when students need it. It's a great text for students and for anyone who wants to deepen their understanding."

Beth Morling, *Professor of Psychological and Brain Sciences, University of Delaware, author of* Research Methods in Psychology, *and winner, 2023 Brewer Distinguished Teaching Award*

"If I were teaching introductory statistics to undergraduates, this is the textbook I'd use. The things that make it distinctive are first, the focus on estimation rather than *p*-values (though the latter are covered), second, the link with free open-source software that allows users to explore analyses and visualisations, and third an emphasis on Open Science practices, coupled with red flags and examples of DFY (Don't Fool Yourself!). There are plenty of exercises, quizzes, and take-home messages, which will bring the material alive even for the most maths-phobic students."

Dorothy Bishop, *Emeritus Professor of Developmental Neuropsychology, University of Oxford*

"*Introduction to the New Statistics* is a next generation statistics textbook. Doing statistics is not the rote application of formulas and reporting answers. Statistics is a tool to support reasoning about evidence. Cumming and Calin-Jageman provide an accessible introduction to using statistics to improve reasoning. *New Statistics* integrates two features that are absent from other texts: meta-analysis and Open Science. No single study or statistical outcome provides the answer to a research question. *New Statistics* teaches data analysis in the context of combining evidence across many studies to gain confidence in conclusions. Also, the best data analysts will plan and show how they made their decisions to enable others to assess their reasoning. *New Statistics* deftly integrates Open Science in every chapter to illustrate how transparency and rigor are fundamental to doing statistics well."

Brian Nosek, *Executive Director, Center for Open Science, Professor, University of Virginia*

"A clear and accessible introduction to statistics, perfect for beginners. This book covers both the old and the new – giving students the fundamentals they need to understand their field, while equipping them with a more sophisticated understanding of the pros and cons of those established practices. The focus on Open Science and integration with statistical tools (e.g., **jamovi**) makes the book particularly useful for training future researchers."

Simine Vazire, *Professor, Melbourne School of Psychological Sciences, University of Melbourne, Editor-in-Chief,* Psychological Science

Introduction to the New Statistics

This fully revised and updated second edition is an essential introduction to inferential statistics. It is the first introductory statistics text to use an estimation approach from the start and also to explain the new and exciting Open Science practices, which encourage replication and enhance the trustworthiness of research. The estimation approach, with meta-analysis ("the new statistics"), is exactly what's needed for Open Science.

Key features of this new edition include:

- Even greater prominence for Open Science throughout the book. Students easily understand basic Open Science practices and are guided to use them in their own work. There is discussion of the latest developments now being widely adopted across science and medicine.
- Integration of new open-source **esci** (Estimation Statistics with Confidence Intervals) software, running in **jamovi**. This is ideal for the book and extends seamlessly to what's required for more advanced courses, and also by researchers. See www.thenewstatistics.com/itns/esci/jesci/.
- Colorful interactive simulations, including the famous *dances*, to help make key statistical ideas intuitive. These are now freely available through any browser. See www.esci.thenewstatistics.com/.
- Coverage of both estimation and null hypothesis significance testing (NHST) approaches, with full guidance on how to translate between the two.
- Effective learning strategies and pedagogical features to promote critical thinking, comprehension, and retention.

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

A comprehensive website offers data sets, key term flashcards, learning guides, and videos describing key concepts and demonstrating the use of **esci**. For instructors, there are guides for teaching the new statistics and Open Science, assessment exercises, question banks, downloadable slides, and more. Altogether, the website provides engaging learning resources for traditional or flipped classrooms. See www.routledge.com/cw/cumming.

Geoff Cumming is a professor emeritus of La Trobe University, Melbourne, Australia, and has been teaching statistics for over 50 years.

Robert Calin-Jageman is a professor of psychology and the neuroscience program director at Dominican University, River Forest, IL, USA, and has been teaching and mentoring undergraduate students since 2007.

Introduction to the New Statistics

Estimation, Open Science, and Beyond

Second Edition

Geoff Cumming and

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.

Second edition published 2024
by Routledge
605 Third Avenue, New York, NY 10158

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

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

First edition published by Routledge 2017

Library of Congress Cataloging-in-Publication Data
Names: Cumming, Geoff, author. | Calin-Jageman, Robert, author.
Title: Introduction to the new statistics : estimation, open science, and beyond / Geoff Cumming, Robert Calin-Jageman.
Description: Second edition. | New York : Routledge, Taylor & Francis Group, [2024] | Revised edition of the authors' Introduction to the new statistics, 2017. | Includes bibliographical references and index.
Identifiers: LCCN 2023041453 (print) | LCCN 2023041454 (ebook) | ISBN 9780367531508 (paperback) | ISBN 9780367531492 (hardback) | ISBN 9781032689470 (ebook)
Subjects: LCSH: Estimation theory—Textbooks. | Mathematical statistics—Textbooks. | Confidence intervals—Textbooks. | AMS: Statistics—Sampling theory, sample surveys—Sampling theory, sample surveys. | Statistics—Instructional exposition (textbooks, tutorial papers, etc.). | Statistics—Applications—Applications to psychology. | Mathematics education—Combinatorics, graph theory, probability theory, statistics—Descriptive statistics. | Mathematics education—Combinatorics, graph theory, probability theory, statistics—Foundations and methodology of statistics.
Classification: LCC QA276.8 .C86 2024 (print) | LCC QA276.8 (ebook) | DDC 519.5—dc23/eng/20231130
LC record available at https://lccn.loc.gov/2023041453
LC ebook record available at https://lccn.loc.gov/2023041454

ISBN: 978-0-367-53149-2 (hbk)
ISBN: 978-0-367-53150-8 (pbk)
ISBN: 978-1-032-68947-0 (ebk)

DOI: 10.4324/9781032689470

Typeset in Meridien LT Std
by Apex CoVantage, LLC

Access the Instructor and Student Resources: www.routledge.com/cw/cumming

GC: For my grandchildren: Tom, Lachie, Odin, Pippa, Lucy, Erin, and Zoe.

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 NEW EDITION

The most important advance is the new **esci** software that's designed to make statistical ideas vivid and to provide data analysis facilities tailored to the needs of beginning students and this book. It's very easy to learn and use and provides a seamless transition to what's needed for more advanced statistics courses and by the practicing researcher. In addition, the whole text has been thoroughly revised and updated with many new examples that illustrate Open Science in action. The latest Open Science developments are described and explained. It is the ideal introductory statistics textbook for the twenty-first century.

INTRODUCTION TO THE NEW STATISTICS

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 practice.

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 more than 70 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 revised 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 included a strong recommendation that researchers use the new statistics. Almost a decade later, the guidelines still include a very similar 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 argues that estimation and meta-analysis are more informative than NHST and that 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 20 December 2023)

Psychological Science also encourages researchers to adopt Open Science practices, and offers badges to recognize preregistration of research plans, open analysis code, 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 section "Making the Most of This Book," below.

INTENDED AUDIENCE

This book assumes no previous statistical knowledge. It's 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) 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. In fact, Open Science is fast becoming mainstream practice.

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 worldwide replication effort.

In January 2023 the U.S. President announced 2023 as the *Year of Open Science* “to advance national open science policies across the federal government”. See tiny.cc/OpenScience2023.

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,” below.

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** (Estimation Statistics with 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. They make it easier for the reader to gain an overview and to find key points when reviewing for exams.
- Many of the exercises use the *esci software* (see below) for interactive learning and a visual grasp of concepts—and for data analysis.
- 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 six *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

For this second edition, the Microsoft Excel-based **ESCI** (Exploratory Software for Confidence Intervals) software used throughout the first edition has been totally replaced. The interactive demonstrations of statistical concepts have all been moved online (www.esci.thenewstatistics.com), thanks to the dedicated work of Gordon Moore. This means you can follow along with the explorations in the book on almost any device, with no need to install software. This new online component is engaging, vivid, easy to use, and includes the famous *dances*—dance of the means, dance of the confidence intervals, dance of the *p* values, and more.

For analyzing data, we now provide **esci** (Estimation Statistics with Confidence Intervals). This free, open-source software provides functions for analyzing both simple and advanced research designs and for synthesizing results using meta-analysis. It provides excellent figures with an emphasis on confidence intervals. For free installation go to thenewstatistics.com/itns/esci/jesci.

Users can interact with the new version of **esci** for data analysis in several ways. For students, **esci** is available as a module in **jamovi**, providing an easy-to-use graphical interface for data analysis. The **jamovi** software is rapidly gaining popularity around the world as a free, open-source alternative to SPSS. We are excited to focus on **esci** within **jamovi** throughout this book.

For advanced users, **esci** is also available as a package in the statistical programing language R. Use of **esci** within R allows scripting of analyses and exposes additional options and features that are not available within **jamovi**. Finally, we are continuing to develop additional interfaces for **esci**. Our plans include releasing **esci** as a plugin for JASP, which is another fantastic option for data analysis for both students and experienced researchers.

See the *appendix* for more about the online simulations and **esci**, and for information about many helpful resources for students and teachers. For news on the latest **esci** developments go to thenewstatistics.com/itns.

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 access to **esci** and to an online suite of interactive demonstrations. See the preceding section.
- 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 installing and using **esci** as a package in the statistical programming language R and soon, we hope, in JASP as well.
- 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, including all figures.

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 and includes consideration of interactions. It discusses the two-way independent groups design and the mixed design, also known as the design for a randomized controlled trial. It introduces two-way ANOVA.

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

The *Appendix* explains how to access and use **esci**, with numerous hints for getting the most out of the software, and links to further sources of information and advice.

ACKNOWLEDGMENTS

GC: Numerous colleagues and students have contributed to my learning that has led to this book. I thank them all. Fiona Fidler and Neil Thomason 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 encouraged my work. They and more recent editors-in-chief have also provided outstanding leadership toward better research practices. 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 with outstanding and dedicated colleagues. These include Tracy Caldwell, Tina Taylor-Ritzler, Persis Driver, Sophia Duffy, TJ Krafnick, and most especially Rebecca Pliske. Thanks for the many discussions, beers, and tears spent together pushing each other to do even better for our students.

Together we warmly thank Jonathon Love, Damian Dropmann, Ravi Selker, and the **jamovi** community, creators of the wonderful **jamovi** data analysis program. This is beautifully designed to be easy to use, while also being fully extensible—simply click to have **esci**, or other modules of your choice, appear as additional menus. The **jamovi** team have patiently provided extensive help with making **esci** work well within **jamovi**. They have been integral to making this second edition possible. Gordon Moore, creator of **esci web**, has also made an enormous contribution. He came to us with ideas about web-based versions of our simulations. He finished up building the full range of simulations and tools that make up **esci web**. These are ideal for this book, as well as being a valuable resource for any students or teachers of statistics. We also owe a huge debt of gratitude to Doug Bonett. Doug's foundational and extensive work in estimation has been essential for us: **esci** is built entirely upon his authoritative **statpsych** package in R. In addition, the figures in **esci** would not look so amazing were it not for the **ggdist** package and extensive help from Matthew Kay. We include in the appendix further acknowledgement of software and techniques essential for **esci**, with citations.

We warmly thank Eli Lehmann and Petra Vaiglova for their painstaking work on the text and exercises. Craig Wendorf, Jeremy Wilmer, and Jason Pych provided extensive feedback that was very useful for shaping this edition. 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 very grateful to the Routledge editorial team, including Georgette Enrique, Ceri McLardy, Leah Burton, and Tori Sharpe; and to the production team led by Marie Louise Roberts. 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 (Cumming, 2012). He has taught statistics for more than 50 years at every level, from introductory to advanced. His statistics tutorial articles have been downloaded more than 500,000 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 been teaching statistics and mentoring students in psychological science since 2007. 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 web**, will similarly reassure you and help you build your confidence. Maybe you can 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.

Surveys suggest that enormous numbers of students, perhaps the majority, rely on Items 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.

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

By contrast, Items 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!

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".

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

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.

I summarize Points 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.

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

Change a Fixed Mindset to a Growth Mindset

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)

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.

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 et al. (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.

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 wish, 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 in the news:

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

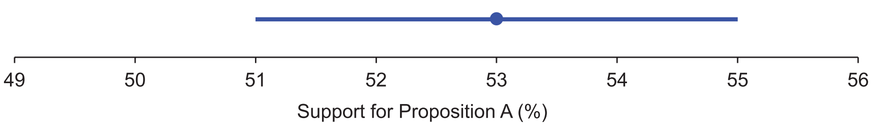
Let's say Proposition A proposes a law requiring serious action on climate change by reducing the use of fossil fuels and switching to renewable energy. Soon there will be a statewide vote to determine whether the proposition becomes law. You and your friends have set up a website explaining why the 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?

DOI: 10.4324/9781032689470-1

Figure 1.1

Support for Proposition A as Reported by the Poll



Note. Support is expressed in percent. The dot marks the point estimate, and the two lines display the margin of error (2%) on 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 behavioral therapy can be for depression;
- how much ice mass the Greenland icecap is likely to lose in the next two decades; and
- the extent to which 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 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 that are 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.

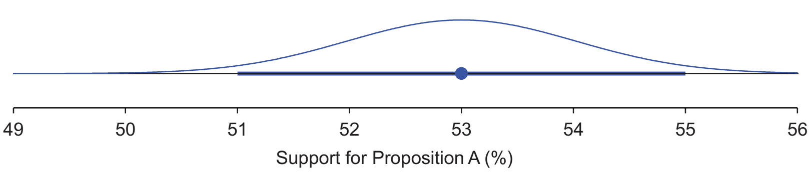
Now for a caution: Unfortunately, ±2 can sometimes refer to something other than margin of error. Perhaps *standard deviation* or *standard error*—don't worry, we'll get to such techie stuff later—but I recommend using ±2 to refer to half the length of a 95% CI, not for anything else. That's what I'll do in this book.

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.

Figure 1.2 displays the *plausibility picture* of a CI; the curve is the *plausibility curve* of a 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 further away

Figure 1.2
Same as Figure 1.1, With Addition of the Plausibility Curve



Note. The smooth curve pictures how likelihood, or plausibility, 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.

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.

We can say that Figure 1.2 displays the *plausibility picture* of the CI, and refer to the curve as the *plausibility curve* on that CI.

Note in particular that there's nothing special about the endpoints of the CI. Always keep in mind the smooth plausibility curve, 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. In Chapter 6 we'll see why, but for now I recommend we follow convention and my preference 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 a 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 point estimate is most likely very close to the true value—the margin of error is smaller, and 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 CI, and MoE is our measure of precision.

MoE stands for margin of error.

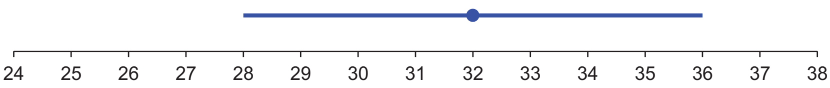
Quiz 1.1

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. The point estimate for referendum support.
   b. The population for referendum support.
   c. The margin of error (MoE).
   d. 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.

Check your quiz answers, maybe discuss with a friend, before you look at my answers at the end of the chapter.

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

Figure 1.3
First Preference for your Product, Product B, with 95% CI



not there just for practice, but often play an important part in the main discussion, 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, but 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. What if the news website where we read the poll result reported only this single poll, but there are 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 like these are met.

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

Open Science comprises a number of practices designed to improve research. It has emerged only in the last decade or so 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!

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

The *replicability* crisis is the realization that some published findings can't be replicated and therefore are almost certainly wrong.

Similar, although happily less extreme, results have been reported in psychology and other disciplines. Here's an example from psychology. Gorn (1982) published evidence that even unobtrusive music, perhaps in a supermarket, can markedly influence consumer preference. The finding has been cited hundreds of times and is routinely included even in recent textbooks on consumer psychology. The first large and careful attempt at replication was by Vermeulen et al. (2014), who reported three studies that together suggested the effect was either zero or very much smaller than reported by Gorn, and was of little, if any, practical import. Students, teachers, and psychologists working in marketing have been misled by the original results for several decades.

Open Science

Open Science addresses the crisis by aiming to reduce the chance that incorrect research results are obtained and reported. I've mentioned two of its requirements—to avoid selection in what's reported, and to report research in full detail. We'll meet further requirements in later chapters, but another aim of Open Science is to encourage replication. Rarely, if ever, can a single study give a definitive answer to a research question, and so we should look for a replication study that found similar results before starting to have confidence in any finding.

Why "open"? Good question. The idea is that, as much as possible, full information about every stage of research should be openly available. In particular, the data and full details of the data analysis should be available, so other researchers can repeat the original data analysis as a check, or analyze the data in a different way. Sometimes, for privacy or other reasons, data can't be openly available, but where possible they should be. Having full information also allows other researchers to conduct a replication study, confident that it's as similar as possible to the original.

Two Open Science slogans are:

- Do we have the full story?
- Seek replication.

A good summary of a number of Open Science requirements is the question I asked earlier: "Do we have the full story?" We can now add "Seek replication" as another Open Science guideline.

Suppose we have results from an original study and a number of replication studies. The results look broadly similar but of course are not exactly the same—for a start, sampling variability will cause them to vary. Fortunately, there are statistical tools that allow us to combine the results and provide a basis for overall conclusions. I'm referring to meta-analysis, our next topic.

META-ANALYSIS

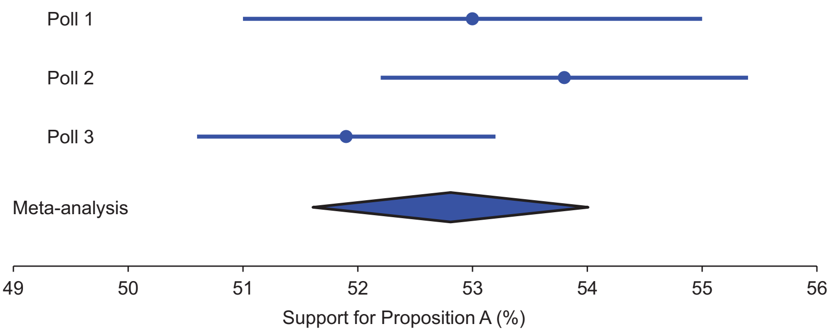
Use *meta-analysis* to combine results from two or more studies on the same or similar questions.

One great thing about estimation is that it can be extended beyond a single study. If we have results from two or more similar studies, we can use *meta-analysis* to combine the results. Usually, meta-analysis gives an overall CI that's shorter than the CI for any of the single studies. That's reasonable, because adding further information from additional studies should reduce our uncertainty about where the population average lies, and reduced uncertainty corresponds to a shorter CI.

We'll discuss meta-analysis more fully in Chapter 9, but here I'll report a meta-analysis of our original poll (Poll 1) and two further polls that I'm supposing were taken at a similar time, asking the same or a similar question. Figure 1.4 shows the result of Poll 1, the same as in Figures 1.1 and 1.2, and also the results of Polls 2 and 3. For each poll we have the point estimate and 95% CI. I applied meta-analysis to combine the three results and obtained

Figure 1.4

A Forest Plot Showing Meta-Analysis of the Original Poll and Two Further Polls



Note. For each poll, the point estimate and 95% CI are displayed. The diamond is the 95% CI that is the result of the meta-analysis.

52.8% as the overall point estimate, with CI of [51.6, 54.0]. That result is pictured as the diamond in Figure 1.4. That figure is our first example of a *forest plot*, which displays individual study results and uses a diamond to show the result of a meta-analysis. We'll see later that the diamond shape is a revealing way to picture a CI, but for the moment just note that the diamond indicates a special CI, a CI that is the result of a meta-analysis.

A *forest plot*, such as Figure 1.4, shows point and interval estimates for individual studies, and displays the meta-analysis result as a diamond.

Just to be sure, say in words what those numbers found by the meta-analysis represent. How are they pictured by the diamond?

Hint: Point and interval estimates?

Yes, 52.8%, the overall point estimate, is our best single estimate of support in the population of intending voters. We are combining, or integrating, evidence from the three studies, and so it makes sense that the overall estimate (52.8%) lies well within the range of the three separate estimates, as marked by the three large dots in Figure 1.4. The 95% CI, the interval estimate given by the meta-analysis, is [51.6, 54.0], meaning that after considering all three polls we can be 95% confident this interval includes the true level of support in the population. This interval is shorter than the other CIs, as we expect for the result of a meta-analysis, which, after all, is based on the results from all three studies. As usual, a short CI is good news.

Meta-Analytic Thinking

Partly because of the rise of Open Science, researchers now increasingly appreciate the crucial importance of replication. Well-conducted replications make a vital contribution to building a research literature that includes fewer wrong findings, and therefore they deserve our trust. Once we have an original study and at least one replication study, we can use meta-analysis to integrate the results.

Replication and meta-analysis are so important that we should adopt *meta-analytic thinking*. That's thinking that prompts us to watch out for any opportunity for meta-analysis to help. It prompts us to think of any study as

Meta-analytic thinking is the consideration of any study in the context of similar studies already conducted, or to be conducted in the future.

one contribution, to be combined where possible with earlier studies, replications, and other related studies yet to come. We should adopt meta-analytic thinking whenever we review, plan, or interpret research.

What Are Your Thoughts?

Are you thinking that needing the full story before feeling confident in a result is a simple and obvious requirement? That a forest plot combining replication results simply makes sense? In my experience, beginning students often express such thoughts after we've been discussing Open Science. Of course, students saying that they understand, and even that something is simple, is music to the ears of any teacher! If you feel these ideas are clear and appealing, then give yourself a big pat on the back: You are on the way to becoming an open scientist.

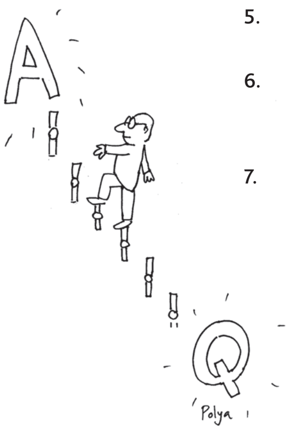
For the decade or so that Open Science practices have been spreading across science, students and early career researchers have been prominent in driving the advances. I encourage you to hold on to any thoughts about the simplicity of Open Science. You and your friends might soon be participating in making it mainstream.

It's time to step back and look in a general way at the estimation approach we've been using.

A STEP-BY-STEP PLAN FOR ESTIMATION

Here's my suggested list of the important steps in an estimation approach to asking and answering research questions. It's a slight expansion of my thumbnail sketch earlier in this chapter.

1. State the research question. Express it as a "how much" or "to what extent" question.
   What's the support for Proposition A in the population of people likely to vote?
2. Identify the measure that's most appropriate for answering that question.
   The percentage of likely voters who express support.
3. Design a study that uses that measure and gives us good point and interval estimates to answer our question.
   Choose a sample of intending voters and ask them about their support for the proposition.
4. After running the study, examine the data, calculate point and interval estimates, and make a figure.
   See Figures 1.1 and 1.2. In answer to our question, the poll found a point estimate of 53%, with CI of [51, 55].
5. Interpret these, using judgment in the research context.
   See the section "Making Your Interpretation".
6. Report the study, making sure to state there was no selective reporting of just some of the results, and giving full details of every aspect of the study.
   These are two Open Science requirements.
7. Adopt meta-analytic thinking throughout. Seek other similar studies and, if appropriate, conduct a meta-analysis. Consider conducting a replication.
   Figure 1.4 shows results from two further similar polls and meta-analysis of the three polls.



This list is not meant to be a comprehensive guide to conducting good research, but it does express many of the important steps. We'll expand and refine the list in later chapters.

In the course of discussing the poll example we've encountered a number of important ideas in an informal way. I've introduced many terms, again informally. If you feel you are beginning to understand the whole estimation approach to finding out about the world, then give yourself another pat on the back: You are well on the way to understanding the statistics needed for Open Science.

Quiz 1.2

1. You see the result of a poll. What are three additional things you need to know before you can understand the result?
2. A poll found 66% support for a proposition, with margin of error of 4%. What is the CI? If the poll were repeated, with a second sample of the same size, what result would you expect?
3. A replication study
   a. uses the same sample as the original study.
   b. most likely gives exactly the same result as the original study.
   c. is usually a waste of time.
   d. is similar to the original study but with a new sample.
4. A forest plot displays for each study a dot that marks ______________ and a line that marks ______________. The diamond represents ______________.
5. What are two important slogans for Open Science? ________________________________ ____________
6. It is important not only to replicate results but also to combine the data across replications to determine the overall pattern of results. Which is the correct term for this?
   a. Repli-analysis
   b. Meta-analysis
   c. Under-analysis
   d. Repeato-analysis
7. Make further quiz questions, some on material from near the start of the chapter. Swap with a friend.

Looking Ahead

Our poll was perhaps the simplest possible estimation study—it estimated just a single percentage. In future chapters we'll discuss many more studies, which address various types of research questions. For all of them we'll take the estimation approach that I've introduced in this chapter, and our step-by-step plan will be a useful guide. Here are a few words about several of the studies that we'll discuss:

Two separate groups of participants. To what extent is retrieval practice, such as taking a quiz, a better study method than simply rereading, as I claimed in *Making the Most of This Book*? A study by Alice Latimier and colleagues (Latimier et al., 2019) compared final test performance for one group of learners who completed a quiz after each learning module, with that for another group who instead read the learning material a second time. We'll discuss this study briefly in Chapter 2, then in more detail in later chapters.

A single group of participants, measured twice. What can psychology tell us about fake news? Daniel Effron and Medha Raj (Effron & Raj, 2020) provided some possibly troubling evidence on this question. They compared ratings of headlines seen previously (Old headlines) with headlines seen for the first time (New headlines). There was only one group of participants, with each rating both Old and New headlines. There's more on this in Chapter 8.

The correlation between two variables. To what extent do people in large cities tend to walk faster than people in towns or small villages? Walking speeds were observed for people in cities, towns, and villages in various countries around the world, and then used to assess the correlation, or relationship, between observed walking speed and the population size of the city, town, or village. What do you guess the researchers found? The answer is in Chapter 11, where we discuss correlation.

In the next five chapters we'll discuss some further fundamentals of research (Chapter 2), graphs to help us explore our data and statistics that summarize data (Chapter 3), and a number of steps that lead to the calculation of confidence intervals (Chapters 4 and 5), with a look in Chapter 6 at an alternative to estimation—a second approach to analyzing data and drawing conclusions about a population. Then we'll have the tools we need to discuss and analyze studies of the types I described just above.

I now invite you to revisit this chapter, and discuss with other learners anything you find particularly interesting, or puzzling, or surprising.

I'll close the chapter with take-home messages. You can look ahead to see what I've written, but it's much more useful if you write yours first. What are the major points so far? What's worth remembering? I suggest you try writing a list, then sleep on it, discuss with others, then revise your list. Then you can compare it with my list—and give yourself extra points for including any important items I've left out.

After your first go at writing some take-home messages, try these exercises—which may suggest more items for your list.

1.4 Draw a picture of a CI and use it to explain (a) point estimate, (b) interval estimate, (c) MoE.

1.5 Make up two interesting exercises involving CIs. Swap with a friend, or post to your discussion group. Discuss everyone's answers. How about a prize for the most interesting exercise?

1.6 Search the web for examples of different ways that CIs are pictured in figures reporting data. Do you have a preference?

1.7 Revise again your list of take-home messages. Scan through the chapter again, looking for anything further that's important enough to include. I have 10 items on my list, but your list can have whatever number you think best.

Finally, a quick reminder that it's worth visiting the book website—have you made a favorite or bookmark? It's at www.routledge.com/cw/cumming, but for easy typing you can use tiny.cc/itns. Incidentally, I use *tiny.cc* shortenings for some links in later chapters, for easy typing and so that I can, if necessary, update where they take you if a site later moves, after the book is published. At the book website there's a list of those shortenings and the full links.

Reporting Your Work

An essential part of the research process is reporting your work—presenting or publishing your results so that others can learn from the data you have collected. Because this is so important, you'll find a section like this one at the end of most chapters with pointers for this vital step. You'll also find example manuscripts on the book website that you can use as a model.

To get us started, here are four basic principles to follow in reporting your work. They may seem a bit vague for the moment, but you'll see how they work in practice as you move through the book. In the last chapter of the book, Chapter 16, there's a recap of these principles plus, for each, bullet points that we've encountered along the way.

Tell the full story is our main guideline for reporting research.

Tell the full story. Give a complete account of your research process and the data you collected. Don't selectively report results.

Provide sufficient detail. Include all the details necessary for someone else to replicate your work. Include all the data necessary for someone else to incorporate your results into a meta-analysis. Share your data online, if possible.

Show the data. Whenever possible, provide figures that show your key findings and almost certainly include CIs. Prefer figures that show all the data rather than just summary statistics.

Interpret the point estimate and CI. Focus your conclusions and interpretation on the point estimates and confidence intervals.

Choose a style and format that is appropriate for your audience. To help with this, these sections provide tips for following the APA style laid out in the *Publication Manual of the American Psychological Association* (APA, 2020). APA style is the most commonly used style for writing manuscripts in a wide range of disciplines. In fact, you'll notice that I've tried to make this book consistent with APA style, so examples you encounter in the text and the end-of-chapter exercises will, in many cases, provide further examples of APA style in action. Keep in mind, though, that these pointers are not a comprehensive treatment of APA style; you'll still need to consult the *Manual* when writing your own reports. Better still, explore the helpful resources at apastyle.apa.org.

Even though APA style is our focus, that doesn't mean you should always follow its conventions. If you are using this book for a class, be sure to follow the assignment guidelines carefully. In this book I don't always follow exact APA referencing style—I might write Chaix et al. rather than stating the date of publication, as specified by APA, when there's no ambiguity and to avoid cluttering the text, especially when we're discussing that study for several pages.

If you'd like to work seriously on your writing, or just read about how to be a great writer, try *The Sense of Style* by Steven Pinker (2014)—would you believe, it's beautifully written.

Take-Home Messages

- Think in a quantitative, not a yes-or-no way. State research questions in estimation terms: "How much . . .?", "To what extent . . .?"
- An outline of the research process is provided by the seven-step plan for estimation. Refer to the listing of the seven steps.
- Point estimates and interval estimates provide quantitative, informative answers to research questions.
- A confidence interval (CI) is our interval estimate, and is a range of values calculated from data that most likely includes the true population value we are estimating.
- We can say we're 95% confident our 95% CI includes the true population value. Values near the center of the CI are most plausible for the true value, and plausibility decreases with increasing distance from the center of the CI.
- Use judgment and knowledge of the research context to interpret the results—the point and interval estimates. A figure is usually helpful.
- Open Science requires (a) avoidance of possibly misleading selection of results to be reported, (b) fully detailed reporting of research, and if possible (c) replication with an eye to meta-analysis.
- Meta-analysis combines results from two or more studies on the same or similar research questions, and is likely to give more precise estimates.
- Meta-analytic thinking considers any study in the context of past similar studies and with the expectation that replications are likely to be valuable.
- Using simple examples, pictures, and estimation suggests that learning about research methods and statistics need not be scary or mysterious but is actually absorbing, highly useful, and even fun. (I didn't say that earlier, and you may not agree, but my aim is to persuade you it's true.)

End-of-Chapter Exercises

Answers to all end-of-chapter exercises are at the back of the book.

1) A study estimated the percentage decrease in pain ratings following 60 minutes of guided relaxation. The result was 34% [19, 49].
   a. What is the point estimate of the decrease? The interval estimate? The 95% CI? MoE?
   b. Make two statements about what the CI tells us. Give your interpretation of the results.
   c. Considering Open Science, what concerns do you have?
2) Suppose the result from the pain study had instead been 13% [−3, 29]. Answer all the same questions, noting especially that negative value.
3) You are considering a replication of the pain study of Exercise 1 above.
   a. Supposing you want MoE from the replication to be about half of MoE from the original study, describe your replication study.
   b. Invent the results of such a replication. Compare with the original results.
   c. Suppose you meta-analyze those two sets of results. Suggest roughly what results the meta-analysis might give, and interpret those.
   d. What's the value of replication and meta-analysis?

Answers to Quizzes

Quiz 1.1

1) a; 2) d; 3) c; 4) d; 5) sample, population, margin of error, MoE; 6) half, higher, more; 7) Questions don't need to be complicated, but it's best to choose things you find somewhat tricky.

Quiz 1.2

1) Are there other similar polls? Information about the sample—how was it chosen, how large? How was the poll conducted—what questions, what procedure? 2) [62, 70], most likely a point estimate within the first CI; 3) d; 4) the point estimate, the CI, the result of the meta-analysis; 5) Do we have the full story? Seek replication; 6) b; 7) It's worth using quizzes to review the whole chapter.

Answers to In-Chapter Exercises

1.1 a. 32%, marked by the solid dot; the 95% CI, which is [28, 36]; 4%, the length of either arm, or half the full CI length; b. All potential purchasers of the cream. You might try to describe that population—mainly older people? Mainly adolescents? Potential purchasers, from that population; c. We are 95% confident the interval from 28% to 36% includes the true first preference in the population for Product B, and values inside the CI are plausible for the true level of first preference, whereas values outside are relatively implausible.

1.2 The whole CI lies above 25%, so all plausible values for the true first preference are greater than 25%. Most likely our product is more strongly preferred than that.

1.3 Use a sample four times as large.

1.4 Refer to Figure 1.1. a. The point estimate is 53%, the solid dot, and is our best single estimate of the true value in the population; b. The interval estimate, the 95% CI, is the interval from 51% to 55%, marked by the line; c. MoE, the margin of error, is 2%, and is the length of either segment of the line, and half the length of the CI.

1.5 You might choose a current media report that includes—or should include—data, then ask how knowing the CI would help us understand what's really going on. Once you are aware of sampling variability, as signaled by a CI, you begin to notice how often people mention only the point estimate, for example an average or a percentage, with no mention that there's often considerable uncertainty in that value.

1.6 I found pictures like those in this chapter, or with little crossbars at the ends, and others that pictured a CI as a narrow stripe, without a dot in the middle or crossbars at the ends. I also saw the diamond, as in Figure 1.4. Often CIs are reported only as numbers in text, with no picture. But pictures with CIs are so informative, so cool. An easy search that presents many pictures is for "pictures of confidence intervals".

References

American Psychological Association . (2020). Publication manual of the American Psychological Association (7th ed.). https://doi.org/10.1037/0000165-000

Antonakis, J. (2017). On doing better science: From thrill of discovery to policy implications. The Leadership Quarterly, 28(1), 5–21. https://doi.org/10.1016/j.leaqua.2017.01.006

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. Nature, 533, 452–454. https://doi.org/10.1038/533452a

Basol, M. , Roozenbeek, J. , & van der Linden, S. (2020). Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. Journal of Cognition, 3(1), 2–9. https://doi.org/10.5334/joc.91

Beek, T. , Dijkhoff, L. , Wagenmakers, E.-J. , & Simons, D. J. (2014). Strack replication. https://osf.io/pkd65

Begley, C. G. , & Ellis, L. M. (2012). Raise standards for preclinical cancer research. Nature, 483, 531–533. https://doi.org/10.1038/483531a

Bem, D. J. (1994). “Anomaly or artifact? Comments on Bem and Honorton”: Response to Hyman. Psychological Bulletin, 115, 25–27. https://doi.org/10.1037/0033-2909.115.1.25

Bem, D. J. , & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. Psychological Bulletin, 115, 4–18. https://doi.org/10.1037/0033-2909.115.1.4

Bennett, C. M. , Baird, A. A. , Miller, M. B. , & Wolford, G. L. (2009). Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction. NeuroImage, 47, S125. https://doi.org/10.1016/S1053-8119(09)71202-9

Bisson, J. I. , Roberts, N. P. , Andrew, M. , Cooper, R. , & Lewis, C. (2013). Psychological therapies for chronic post-traumatic stress disorder (PTSD) in adults. Cochrane Database of Systematic Reviews, 2013(12), Art. No.: CD003388. https://doi.org/10.1002/14651858.CD003388.pub4

Bonett, D. G. (2008). Confidence intervals for standardized linear contrasts of means. Psychological Methods, 13(2), 99–109. https://doi.org/10.1037/1082-989X.13.2.99

Bonett, D. G. (2023). Statpsych: Statistical methods for psychologists (1.4.0) [Computer software]. https://cran.r-project.org/web/packages/statpsych/index.html

Bonett, D. G. , & Price, R. M. (2002). Statistical inference for a linear function of medians: Confidence intervals, hypothesis testing, and sample size requirements. Psychological Methods, 7(3), 370–383. https://doi.org/10.1037/1082-989X.7.3.370

Bonett, D. G. , & Price, R. M. (2012). Adjusted Wald confidence interval for a difference of binomial proportions based on paired data. Journal of Educational and Behavioral Statistics, 37(4), 479–488. https://doi.org/10.3102/1076998611411915

Bonett, D. G. , & Price, R. M. (2020a). Confidence intervals for ratios of means and medians. Journal of Educational and Behavioral Statistics, 45(6), 750–770. https://doi.org/10.3102/1076998620934125

Bonett, D. G. , & Price, R. M. (2020b). Interval estimation for linear functions of medians in withinsubjects and mixed designs. British Journal of Mathematical and Statistical Psychology, 73(2), 333–346. https://doi.org/10.1111/bmsp.12171

Bornstein, M. H. , & Bornstein, H. G. (1976). The pace of life. Nature, 259, 557–559. https://doi.org/10.1038/259557a0

Bosco, F. A. , Aguinis, H. , Singh, K. , Field, J. G. , & Pierce, C. A. (2015). Correlational effect size benchmarks. Journal of Applied Psychology, 100, 431–449. https://doi.org/10.1037/a0038047

Brand, T. (2023). ggh4x: Hacks for ‘ggplot2’ (0.2.5) [Computer software]. https://cran.rstudio.com/web/packages/ggh4x/index.html

Brethel-Haurwitz, K. M. , & Marsh, A. A. (2014). Geographical differences in subjective well-being predict extraordinary altruism. Psychological Science, 25, 762–771. https://doi.org/10.1177/0956797613516148

Brown, P. C. , Roediger III, H. L. , & McDaniel, M. A. (2014). Make it stick: The science of successful learning. Belknap.

Burgmer, P. , & Englich, B. (2012). Bullseye! How power improves motor performance. Social Psychological and Personality Science, 4, 224–232. https://doi.org/10.1177/1948550612452014

Button, K. S. , Ioannidis, J. P. A. , Mokrysz, C. , Nosek, B. A. , Flint, J. , Robinson, E. S. J. , & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. Nature Reviews. Neuroscience, 14, 365–376. https://doi.org/10.1038/nrn3475

Cairns, M. , Cumming, G. , Calin-Jageman, R. , & Prendergast, L. A. (2022). The diamond ratio: A visual indicator of the extent of heterogeneity in meta-analysis. British Journal of Mathematical and Statistical Psychology, 75(2), 201–219. https://doi.org/10.1111/bmsp.12258

Calin-Jageman, R. J. , & Caldwell, T. L. (2014). Replication of the superstition and performance study by Damisch, Stoberock, and Mussweiler (2010). Social Psychology, 45, 239–245. https://doi.org/10.1027/1864-9335/a000190

Carpenter, S. K. , Wilford, M. M. , Kornell, N. , & Mullaney, K. M. (2013). Appearances can be deceiving: Instructor fluency increases perceptions of learning without increasing actual learning. Psychonomic Bulletin & Review, 20, 1350–1356. https://doi.org/10.3758/s13423-013-0442-z

Carter, T. J. , Ferguson, M. J. , & Hassin, R. R. (2011). A single exposure to the American flag shifts support toward Republicanism up to 8 months later. Psychological Science, 22, 1011–1018. https://doi.org/10.1177/0956797611414726

Caruso, E. M. , Vohs, K. D. , Baxter, B. , & Waytz, A. (2013). Mere exposure to money increases endorsement of free-market systems and social inequality. Journal of Experimental Psychology: General, 142, 301–306. https://doi.org/10.1037/a0029288

Cash, T. F. (2000). The multi-dimensional body-self relations questionnaire. Users manual. www.body-images.com/assessments

Castles, A. , Rastle, K. , & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. Psychological Science in the Public Interest, 19(1), 5–51. https://doi.org/10.1177/1529100618772271
Chaix, A. , Zarrinpar, A. , Miu, P. , & Panda, S. (2014). Time-restricted feeding is a preventative and therapeutic intervention against diverse nutritional challenges. Cell Metabolism, 20, 991–1005. https://doi.org/10.1016/j.cmet.2014.11.001
Chan, D. W. (2007). Components of leadership giftedness and multiple intelligences among Chinese gifted students in Hong Kong. High Ability Studies, 18, 155–172. https://doi.org/10.1080/13598130701709749
Chervin, R. D. , Ruzicka, D. L. , Giordani, B. J. , Weatherly, R. A. , Dillon, J. E. , Hodges, E. K. , Marcus, C. L. , & Guire, K. E. (2006). Sleep-disordered breathing, behavior, and cognition in children before and after adenotonsillectomy. Pediatrics, 117, e769–e778. https://doi.org/10.1542/peds.2005-1837
Cian, L. , Longoni, C. , & Krishna, A. (2020). Advertising a desired change: When process simulation fosters (vs. hinders) credibility and persuasion. Journal of Marketing Research, 57, 489–508. https://doi.org/10.1177/0022243720904758
Clarke, E. , Sherrill-Mix, S. , & Dawson, C. (2023). ggbeeswarm: Categorical scatter (Violin Point) plots (0.7.2) [Computer software]. https://cran.r-project.org/web/packages/ggbeeswarm/index.html
Cohen, J. (1977). Statistical power analysis for the behavioral sciences (Revised ed.). Academic Press.
Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Erlbaum.
Cohen, J. (1994). The earth is round ($p < .05$). American Psychologist, 49, 997–1003. https://doi.org/10.1037/0003-066X.49.12.997
Cousineau, D. , & Goulet-Pelletier, J.-C. (2020). A review of five techniques to derive confidence intervals with a special attention to the Cohen's d p in the between-group design. PsyArXiv. https://doi.org/10.31234/osf.io/s2597
Crüwell, S. , Apthorp, D. , Baker, B. J. , Colling, L. , Elson, M. , Geiger, S. J. , Lobentanzer, S. , Monéger, J. , Patterson, A. , Schwarzkopf, D. S. , Zaneva, M. , & Brown, N. J. L. (2023). What's in a badge? A computational reproducibility investigation of the open data badge policy in one issue of *Psychological Science* . Psychological Science, 34(4), 513–522. https://doi.org/10.1177/09567976221140828
Cumming, G. (2012). Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis. Routledge.
Cumming, G. (2014). The new statistics: Why and how. Psychological Science, 25, 7–29. https://doi.org/10.1177/0956797613504966. https://tiny.cc/tnswhyhow
Cumming, G. , & Finch, S. (2005). Inference by eye: Confidence intervals, and how to read pictures of data. American Psychologist, 60, 170–180. https://doi.org/10.1037/0003-066X.60.2.170
Cumming, G. , & Maillardet, R. (2006). Confidence intervals and replication: Where will the next mean fall? Psychological Methods, 11, 217–227. https://doi.org/10.1037/1082-989X.11.3.217
Cusack, M. , Vezenkova, N. , Gottschalk, C. , & Calin-Jageman, R. J. (2015). Direct and conceptual replications of Burgmer & Englich (2012): Power may have little to no effect on motor performance. PLoS ONE, 10(11), e0140806. https://doi.org/10.1371/journal.pone.0140806
Damisch, L. , Stoberock, B. , & Mussweiler, T. (2010). Keep your fingers crossed! How superstition improves performance. Psychological Science, 21, 1014–1020. https://doi.org/10.1177/0956797610372631
Deci, E. L. , Koestner, R. , & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychological Bulletin, 125, 627–668. https://doi.org/10.1037/0033-2909.125.6.627
Delacre, M. , Lakens, D. , Ley, C. , Liu, L. , & Leys, C. (2021). Why Hedges' g * s based on the non-pooled standard deviation should be reported with Welch's *t*-test. PsyArXiv. https://doi.org/10.31234/osf.io/tu6mp
de Lorgeril, M. , Salen, P. , Martin, J.-L. , Monjaud, I. , Boucher, P. , & Mamelle, N. (1998). Mediterranean dietary pattern in a randomized trial. Archives of Internal Medicine, 158, 1181–1187. https://doi.org/10.1001/archinte.158.11.1181
Diekelmann, S. , & Born, J. (2010). The memory function of sleep. Nature Reviews Neuroscience, 11, 114–126. https://doi.org/10.1038/nrn2762
Diener, E. , & Diener, C. (1996). Most people are happy. Psychological Science, 7, 181–186. https://doi.org/10.1111/j.1467-9280.1996.tb00354.x
Diener, E. , Emmons, R. A. , Larsen, R. J. , & Griffin, S. (1985). The satisfaction with life scale. Journal of Personality Assessment, 49, 71–75. https://doi.org/10.1207/s15327752jpa4901_13
Donohue, A. , van Gelder, T. , Cumming, G. , & Bissett, M. (2002). Reason! Project studies, 1999–2002. Reason! Project technical report 2002/1. University of Melbourne.
Effron, D. A. (2018). It could have been true: How counterfactual thoughts reduce condemnation of falsehoods and increase political polarization. Personality and Social Psychology Bulletin, 44, 729–745. https://doi.org/10.1177/0146167217746152
Effron, D. A. , & Raj, M. (2020). Misinformation and morality: Encountering fake-news headlines makes them seem less unethical to publish and share. Psychological Science, 31(1), 75–87. https://doi.org/10.1177/0956797619887896
Ekman, P. , Levenson, R. , & Friesen, W. (1983). Autonomic nervous system activity distinguishes among emotions. Science, 221(4616), 1208–1210. https://doi.org/10.1126/science.6612338
Ericsson, K. A. , Krampe, R. T. , & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. Psychological Review, 100, 363–406. https://doi.org/10.1037/0033-295X.100.3.363
Eskine, K. J. (2013). Wholesome foods and wholesome morals? Organic foods reduce prosocial behavior and harshen moral judgments. Social Psychological and Personality Science, 4, 251–254. https://doi.org/10.1177/1948550612447114

Facione, P. A. , & Facione, N. C. (1992). The California critical thinking skills test and manual. California Academic Press.
Floretta-Schiller, S. , Berent, B. , & Salinas, G. (2015). The effects of labels on the enjoyment of chocolate. Unpublished manuscript, Department of Psychology, Dominican University (Illinois).
Flynn, J. R. (2012). Are we getting smarter? Rising IQ in the twenty-first century. Cambridge University Press.
Franco, A. , Malhotra, N. , & Simonovits, G. (2015). Underreporting in psychology experiments: Evidence from a study registry. Social Psychological and Personality Science, 7, 8–12. https://doi.org/10.1177/1948550615598377
Gelman, A. , & Stern, H. (2006). The difference between “significant” and “not significant” is not itself statistically significant. American Statistician, 60, 328–331. https://doi.org/10.1198/000313006X152649
Gervais, W. M. , & Norenzayan, A. (2012). Analytic thinking promotes religious disbelief. Science, 336, 493–496. https://doi.org/10.1126/science.1215647
Gigerenzer, G. (2014). Risk savvy. How to make good decisions. Allen Lane.
Gilbert, R. , Salanti, G. , Harden, M. , & See, S. (2005). Infant sleeping position and the sudden infant death syndrome: Systematic review of observational studies and historical review of recommendations from 1940 to 2002. International Journal of Epidemiology, 34, 874–887. https://doi.org/10.1093/ije/dyi088
Gladwell, M. (2008). Outliers: The story of success. Little Brown.
Godden, D. R. , & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. British Journal of Psychology, 66(3), 325–331. https://doi.org/10.1111/j.2044-8295.1975.tb01468.x
Gorn, G. J. (1982). The effects of music in advertising on choice behavior: A classical conditioning approach. Journal of Marketing, 46(1), 94–101. https://doi.org/10.2307/1251163
Goulet-Pelletier, J.-C. , & Cousineau, D. (2018). A review of effect sizes and their confidence intervals, Part I: The Cohen’s *d* family. The Quantitative Methods for Psychology, 14(4), 242–265. https://doi.org/10.20982/tqmp.14.4.p242
Halagappa, V. K. M. , Guo, Z. , Pearson, M. , Matsuoka, Y. , Cutler, R. G. , LaFerla, F. M. , & Mattson, M. P. (2007). Intermittent fasting and caloric restriction ameliorate age-related behavioral deficits in the triple-transgenic mouse model of Alzheimer’s disease. Neurobiology of Disease, 26(1), 212–220. https://doi.org/10.1016/j.nbd.2006.12.019
Hattie, J. A. C. (2009). Visible learning. Routledge.
Hattie, J. A. C. (2023). Visible learning: The sequel. Routledge. https://doi.org/10.4324/9781003380542
Hilgard, J. (2015). Game violence, game difficulty, and 2D:4D digit ratio as predictors of aggressive behavior. University of Missouri-Columbia.
Hinkle, D. E. , Wiersma, W. , & Jurs, S. G. (2003). Applied statistics for the behavioral sciences (5th ed.). Houghton Mifflin.
Hölzel, B. K. , Carmody, J. , Vangel, M. , Congleton, C. , Yerramsetti, S. M. , Gard, T. , & Lazar, S. W. (2011). Mindfulness practice leads to increases in regional brain gray matter density. Psychiatry Research: Neuroimaging, 191, 36–43. https://doi.org/10.1016/j.pscychresns.2010.08.006
Huff, D. (1954). How to lie with statistics. Norton.
Hunt, M. (1997). How science takes stock: The story of meta-analysis. Russell Sage Foundation.
Huynh, C.-L. (1989). A unified approach to the estimation of effect size in meta-analysis. NBER Working Paper Series, 58(58), 99–104. https://eric.ed.gov/?id=ED306248
Hyde, J. S. (2007). New directions in the study of gender similarities and differences. Current Directions in Psychological Science, 16, 259–263. https://doi.org/10.1111/j.1467-8721.2007.00516.x
Hyman, R. (1994). Anomaly or artifact? Comments on Bem and Honorton. Psychological Bulletin, 115, 19–24. https://doi.org/10.1037/0033-2909.115.1.19
IgNobel Prize in Neuroscience: The Dead Salmon Study . (2012, September 25). The scicurious brain, Scientific American blog network. https://tiny.cc/deadsalmon
Inzlicht, M. (2016, February 29). Reckoning with the past. Getting better: Random musings about science, psychology, and what-have-yous. https://michaelinzlicht.com/getting-better/2016/2/29/reckoning-with-the-past
Ioannidis, J. P. A. (2005). Why most published research findings are false. PLoS Medicine, 2, e124. https://doi.org/10.1371/journal.pmed.0020124. https://tiny.cc/mostfalse
Jacowitz, K. E. , & Kahneman, D. (1995). Measures of anchoring in estimation tasks. Personality and Social Psychology Bulletin, 21, 1161–1166. https://doi.org/10.1177/01461672952111004
John, L. K. , Loewenstein, G. , & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. Psychological Science, 23(5), 524–532. https://doi.org/10.1177/0956797611430953
Johnson, D. J. , Cheung, F. , & Donnellan, M. B. (2014). Does cleanliness influence moral judgments? A direct replication of Schnall, Benton, and Harvey (2008). Social Psychology, 45, 209–215. https://doi.org/10.1027/1864-9335/a000186
Junk, T. R. , & Lyons, L. (2020). Reproducibility and replication of experimental particle physics results. Harvard Data Science Review, 2(4). https://doi.org/10.1162/99608f92.250f995b
Kang, S. H. K. , & Pashler, H. (2014). Is the benefit of retrieval practice modulated by motivation? Journal of Applied Research in Memory and Cognition, 3, 183–188. https://doi.org/10.1016/j.jarmac.2014.05.006
Kardas, M. , & O'Brien, E. (2018). Easier seen than done: Merely watching others perform can foster an illusion of skill acquisition. Psychological Science, 29(4), 521–536. https://doi.org/10.1177/0956797617740646
Kay, M. (2022). ggdist: Visualizations of distributions and uncertainty [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.6300140
Kay, M. (2024). ggdist: Visualizations of distributions and uncertainty in the grammar of graphics. IEEE Transactions on Visualizations and Computer Graphics (to appear).

Kelley, K. (2007). Confidence intervals for standardized effect sizes: Theory, application, and implementation. Journal of Statistical Software, 20(8), 1–24. https://doi.org/10.18637/jss.v020.i08

Klein, R. A. , Ratliff, K. A. , Vianello, M. , Adams, R. B., Jr. , Bahník, Š. , Bernstein, M. J. , Bocian, K. , Brandt, M. J. , Brooks, B. , Brumbaugh, C. C. , Cemalcilar, Z. , Chandler, J. , Cheong, W. , Davis, W. E. , Devos, T. , Eisner, M. , Frankowska, N. , Furrow, D. , Galliani, E. M. , … Nosek, B. A. (2014a). Data from investigating variation in replicability: A "Many Labs" replication project. Journal of Open Psychology Data, 2(1), e4. https://doi.org/10.5334/jopd.ad

Klein, R. A. , Ratliff, K. A. , Vianello, M. , Adams, R. B., Jr. , Bahník, Š. , Bernstein, M. J. , Bocian, K. , Brandt, M. J. , Brooks, B. , Brumbaugh, C. C. , Cemalcilar, Z. , Chandler, J. , Cheong, W. , Davis, W. E. , Devos, T. , Eisner, M. , Frankowska, N. , Furrow, D. , Galliani, E. M. , … Nosek, B. A. (2014b). Investigating variation in replicability. A "Many Labs" replication project. Social Psychology, 45, 142–152. https://doi.org/10.1027/1864-9335/a000178

Kline, R. B. (2013). Beyond significance testing: Statistics reform in the behavioral sciences (2nd ed.). American Psychological Association. https://doi.org/10.1037/14136-000

Lakens, D. (2013). Using a smartphone to measure heart rate changes during relived happiness and anger. IEEE Transactions on Affective Computing, 4, 238–241. https://doi.org/10.1109/T-AFFC.2013.3

Latimier, A. , Riegert, A. , Peyre, H. , Ly, S. T. , Casati, R. , & Ramus, F. (2019). Does pre-testing promote better retention than post-testing? NPJ Science of Learning, 4(15), 1–7. https://doi.org/10.1038/s41539-019-0053-1

Lecoutre, B. (2007). Another look at the confidence intervals for the noncentral *T* distribution. Journal of Modern Applied Statistical Methods, 6(1), 107–116. https://doi.org/10.22237/jmasm/1177992600

Lovakov, A. , & Agadullina, E. R. (2021). Empirically derived guidelines for effect size interpretation in social psychology. European Journal of Social Psychology, 51(3), 485–504. https://doi.org/10.1002/ejsp.2752

Maccoby, E. E. , & Jacklin, C. N. (1974). The psychology of sex differences. Stanford University Press.

Macnamara, B. N. , Hambrick, D. Z. , & Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: A meta-analysis. Psychological Science, 25, 1608–1618. https://doi.org/10.1177/0956797614535810

McCabe, D. P. , & Castel, A. D. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. Cognition, 107, 343–352. https://doi.org/10.1016/j.cognition.2007.07.017

McCarthy, R. J. (2013). Pre-specified hypotheses and planned analyses. https://osf.io/g3kt6

McCarty, M. K. , & Kelly, J. R. (2015). When door holding harms: Gender and the consequences of non-normative help. Social Influence, 10, 1–10. https://doi.org/10.1080/15534510.2013.869252

McDaniel, M. A. , Howard, D. C. , & Einstein, G. O. (2009). The read-recite-review study strategy: Effective and portable. Psychological Science, 20, 516–522. https://doi.org/10.1111/j.1467-9280.2009.02325.x

McEldoon, K. L. , Durkin, K. L. , & Rittle-Johnson, B. (2013). Is self-explanation worth the time? A comparison to additional practice. The British Journal of Educational Psychology, 83(4), 615–632. https://doi.org/10.1111/j.2044-8279.2012.02083.x

McFerran, B. , & Mukhopadhyay, A. (2013). Lay theories of obesity predict actual body mass. Psychological Science, 24, 1428–1436. https://doi.org/10.1177/0956797612473121

McGuinness, D. (2004). Early reading instruction: What science really tells us about how to teach reading. MIT Press.

Messerli, F. H. (2012). Chocolate consumption, cognitive function, and Nobel Laureates. New England Journal of Medicine, 367, 1562–1564. https://doi.org/10.1056/NEJMon1211064

Michael, R. B. , Newman, E. J. , Vuorre, M. , Cumming, G. , & Garry, M. (2013). On the (non) persuasive power of a brain image. Psychonomic Bulletin & Review, 20, 720–725. https://doi.org/10.3758/s13423-013-0391-6

Moery, E. , & Calin-Jageman, R. J. (2016). Direct and conceptual replications of Eskine (2013): Organic food exposure has little to no effect on moral judgments and prosocial behavior. Social Psychology and Personality Science, 7, 312–319. https://doi.org/10.1177/1948550616639649

Moore, E. R. , Bergman, N. , Anderson, G. C. , & Medley, N. (2016). Early skin-to-skin contact for mothers and their healthy newborn infants. Cochrane Database of Systematic Reviews (11), Art. No.: CD003519. https://doi.org/10.1002/14651858.CD003519.pub4

Moshontz, H. , Campbell, L. , Ebersole, C. R. , Ijzerman, H. , Urry, H. L. , Forscher, P. S. , Grahe, J. E. , McCarthy, R. J. , Musser, E. D. , Antfolk, J. , Castille, C. M. , Evans, T. R. , Fiedler, S. , Flake, J. K. , Forero, D. A. , Janssen, S. M. J. , Keene, J. R. , Protzko, J. , Aczel, B. , … Chartier, C. R. (2018). The psychological science accelerator: Advancing psychology through a distributed collaborative network. Advances in Methods and Practices in Psychological Science, 1(4), 501–515. https://doi.org/10.1177/2515245918797607

Mosley, M. , & Spencer, M. (2013). The FastDiet: Lose weight, stay healthy, and live longer with the simple secret of intermittent fasting. Atria.

National Reading Panel . (2000). Report. National Institute of Child Health and Human Development.

Neville, L. (2012). Do economic equality and generalized trust inhibit academic dishonesty? Evidence from state-level search-engine queries. Psychological Science, 23, 339–345. https://doi.org/10.1177/0956797611435980

Nieuwenhuis, S. , Forstmann, B. , & Wagenmakers, E. J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. Nature Neuroscience, 14, 1105–1107. https://doi.org/10.1038/nn.2886

Nosek, B. A. , Alter, G. , Banks, G. C. , Borsboom, D. , Bowman, S. D. , Breckler, S. J. , Buck, S. , Chambers, C. D. , Chin, G. , Christensen, G. , Contestabile, M. , Dafoe, A. , Eich, E. , Freese, J. , Glennerster, R. , Goroff, D. , Green, D. P. , Hesse, B. , Humphreys, M. , … Yarkoni, T. (2015). Promoting an open research culture: Author guidelines for journals could help to promote transparency, openness, and reproducibility. Science, 348(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Nosek, B. A. , Banaji, M. R. , & Greenwald, A. G. (2002). Math = male, me = female, therefore math not = me. Journal of Personality and Social Psychology, 83, 44–59. https://doi.org/10.1037/0022-3514.83.1.44

Open Science Collaboration . (2015). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716. https://doi.org/10.1126/science.aac4716
O'Reilly, T. , Symons, S. , & MacLatchy-Gaudet, H. (1998). A comparison of self-explanation and elaborative interrogation. Contemporary Educational Psychology, 23, 434–445. https://doi.org/10.1006/ceps.1997.0977
O'Rourke, E. , Haimovitz, K. , Ballwebber, C. , Dweck, C. S. , & Popović, Z. (2014). Brain points: A growth mindset incentive structure boosts persistence in an educational game. CHI 2014 proceedings of the SIGCHI conference on human factors in computing systems. Toronto, April 2014 (pp. 3339–3348). ACM. https://doi.org/10.1145/2556288.2557157
Ortega, F. B. (2013). The intriguing association among chocolate consumption, country's economy and Nobel Laureates. Clinical Nutrition, 32, 874–875. https://doi.org/10.1016/j.clnu.2013.05.011
Paille, F. M. , Guelfi, J. D. , Perkins, A. C. , Royer, R. J. , Steru, L. , & Parot, P. (1995). Double-blind randomized multicentre trial of acamprosate in maintaining abstinence from alcohol. Alcohol and Alcoholism, 30(2), 239–247. https://doi.org/10.1093/oxfordjournals.alcalc.a045720
Park, L. E. , Young, A. F. , Troisi, J. D. , & Pinkus, R. T. (2011). Effects of everyday romantic goal pursuit on women's attitudes toward math and science. Personality & Social Psychology Bulletin, 37(9), 1259–1273. https://doi.org/10.1177/0146167211408436
Parris, K. M. , & McCarthy, M. A. (2001). Identifying effects of toe-clipping on anuran return rates: The importance of statistical power. Amphibia-Reptilia, 22, 275–289. https://doi.org/10.1163/156853801317050070
Pav, S. E. (2020). sadists: Some additional distributions (0.2.4) [Computer software]. https://cran.r-project.org/web/packages/sadists/
Pearl, J. , & Mackenzie, D. (2018). The book of why: The new science of cause and effect. Penguin.
Pearson, H. (2023). Medical-evidence giant Cochrane battles funding cuts and closures. Nature, 621, 13–14. https://doi.org/10.1038/d41586-023-02741-z
Pennington, C. (2023). A student's guide to Open Science: Using the replication crisis to reform psychology. McGraw Hill, Open University Press.
Pinker, S. (2014). The sense of style. Viking.
Price, R. M. , & Bonett, D. G. (2004). An improved confidence interval for a linear function of binomial proportions. Computational Statistics & Data Analysis, 45(3), 449–456. https://doi.org/10.1016/S0167-9473(03)00007-0
Rattan, A. , Good, C. , & Dweck, C. S. (2012). "It's ok—Not everyone can be good at math": Instructors with an entity theory comfort (and demotivate) students. Journal of Experimental Social Psychology, 48, 731–737. https://doi.org/10.1016/j.jesp.2011.12.012
Richardson, M. , Abraham, C. , & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. Psychological Bulletin, 138, 353–387. https://doi.org/10.1037/a0026838
Ritchie, S. J. , & Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. Psychological Science, 24, 1301–1308. https://doi.org/10.1177/0956797612466268
Robinson, T. N. , Borzekowski, D. L. G. , Matheson, D. M. , & Kraemer, H. C. (2007). Effects of fast food branding on young children's taste preferences, Archives of Pediatrics & Adolescent Medicine, 161, 792–797. https://doi.org/10.1001/archpedi.161.8.792
Rohrer, D. , Pashler, H. , & Harris, C. R. (2015). Do subtle reminders of money change people's political views? Journal of Experimental Psychology: General, 144, e73–e85. https://doi.org/10.1037/xge0000058
Rosenthal, R. , & Rosnow, R. L. (1985). Contrast analysis: Focused comparisons in the analysis of variance. Cambridge University Press.
Rozin, P. , Scott, S. E. , Zickgraf, H. F. , Ahn, F. , & Hong, J. (2014). Asymmetrical social Mach bands: Exaggeration of social identities on the more esteemed side of group borders. Psychological Science, 25, 1955–1959. https://doi.org/10.1177/0956797614545131
Ruffman, T. , Perner, J. , Naito, M. , Parkin, L. , & Clements, W. A. (1998). Older (but not younger) siblings facilitate false belief understanding. Developmental Psychology, 34(1), 161–174. https://doi.org/10.1037/0012-1649.34.1.161
Rugg, D. (1941). Experiments in wording questions: II. Public Opinion Quarterly, 5, 91–92. https://doi.org/10.1086/265467
Rutledge, T. , & Loh, C. (2004). Effect sizes and statistical testing in the determination of clinical significance in behavioral medicine research. Annals of Behavioral Medicine, 27, 138–145. https://doi.org/10.1207/s15324796abm2702_9
Salsburg, D. (2001). The lady tasting tea. How statistics revolutionized science in the twentieth century. W. H. Freeman.
Sanchez, C. , Sundermeier, B. , Gray, K. , & Calin-Jageman, R. J. (2017). Direct replications of Gervais & Norenzayan (2012): No evidence that analytic thinking decreases religious belief. PLoS One, 12(2), e0172636. https://doi.org/10.1371/journal.pone.0172636
Schnall, S. , Benton, J. , & Harvey, S. (2008). With a clean conscience: Cleanliness reduces the severity of moral judgments. Psychological Science, 19, 1219–1222. https://doi.org/10.1111/j.1467-9280.2008.02227.x
Sedky, K. , Bennett, D. S. , & Carvalho, K. S. (2014). Attention deficit hyperactivity disorder and sleep disordered breathing in pediatric populations: A meta-analysis. Sleep Medicine Reviews, 18, 349–356. https://doi.org/10.1016/j.smrv.2013.12.003
Sedrakyan, A. , & Shih, C. (2007). Improving depiction of benefits and harms: Analyses of studies of well-known therapeutics and review of high-impact medical journals. Medical Care, 45, S23–S28. https://doi.org/10.1097/MLR.0b013e3180642f69

Simmons, J. P. , & Nelson, L. D. (2020, December 3). Data Replicada #9: Are progression ads more credible? Data Colada. https://datacolada.org/94

Simmons, J. P. , Nelson, L. D. , & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological Science, 22, 1359–1366. https://doi.org/10.1177/0956797611417632

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. Psychological Science, 26, 559–569. https://doi.org/10.1177/0956797614567341

Smith, A. M. , Floerke, V. A. , & Thomas, A. K. (2016). Retrieval practice protects memory against acute stress, Science, 354(6315), 1046–1048. https://doi.org/10.1126/science.aah5067. https://tiny.cc/SmithRetrieval

Soderberg, C. K. , Errington, T. M. , Schiavone, S. R. , Bottesini, J. , Thorn, F. S. , Vazire, S. , Esterling, K. M. , & Nosek, B. A. (2021). Initial evidence of research quality of registered reports compared with the standard publishing model. Nature Human Behaviour, 5, 990–997. https://doi.org/10.1038/s41562-021-01142-4

Spielberger, C. D. , Gorssuch, R. L. , Lushene, P. R. , Vagg, P. R. , & Jacobs, G. A. (1983). Manual for the state-trait anxiety inventory. Consulting Psychologists Press.

Steiger, J. H. (2004). Beyond the *F* test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. Psychological Methods, 9, 164–182. https://doi.org/10.1037/1082-989X.9.2.164

Stickgold, R. , James, L. , & Hobson, J. A. (2000). Visual discrimination learning requires sleep after training. Nature Neuroscience, 3, 1237–1238. https://doi.org/10.1038/81756

Tindle, R. (2021). Improving the global reach of psychological research. Discover Psychology, 1, 5. https://doi.org/10.1007/s44202-021-00004-4

Vassallo, S. , & Sanson, A. (Eds.). (2013). The Australian temperament project: The first 30 years. Australian Institute of Family Studies. https://tiny.cc/atp30

Vermeulen, I. , Batenburg, A. , Beukeboom, C. J. , & Smits, T. (2014). Breakthrough or one-hit wonder? Three attempts to replicate single-exposure musical conditioning effects on choice behavior (Gorn, 1982). Social Psychology, 45, 179–186. https://doi.org/10.1027/1864-9335/a000182

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. Journal of Statistical Software, 36(3), 1–43. https://doi.org/10.18637/jss.v036.i03

Wagenmakers, E.-J. , Beek, T. , Dijkhoff, L. , Gronau, Q. F. , Acosta, A. , Adams, R. B., Jr. , Albohn, D. N. , Allard, E. S. , Benning, S. D. , Blouin-Hudon, E.-M. , Bulnes, L. C. , Caldwell, T. L. , Calin-Jageman, R. J. , Capaldi, C. A. , Carfagno, N. S. , Chasten, K. T. , Cleeremans, A. , Connell, L. , DeCicco, J. M. , … Zwaan, R. A. (2016). Registered replication report: Strack, Martin, & Stepper (1988). Perspectives on Psychological Science, 11, 917–928. https://doi.org/10.1177/1745691616674458

Wagge, J. R. , Baciu, C. , Banas, K. , Nadler, J. T. , Schwarz, S. , Weisberg, Y. , Ijzerman, H. , Legate, N. , & Grahe, J. (2019). A demonstration of the Collaborative Replication and Education Project: Replication attempts of the red-romance effect. Collabra: Psychology, 5(1), 5. https://doi.org/10.1525/collabra.177

Wang, K. , Goldenberg, A. , Dorison, C. A. , Miller, J. K. , Uusberg, A. , Lerner, J. S. , Gross, J. J. , Agesin, B. B. , Bernardo, M. , Campos, O. , Eudave, L. , Grzech, K. , Ozery, D. H. , Jackson, E. A. , Garcia, E. O. L. , Drexler, S. M. , Jurković, A. P. , Rana, K. , Wilson, J. P. , … Moshontz, H. (2021). A multi-country test of brief reappraisal interventions on emotions during the COVID-19 pandemic. Nature Human Behaviour, 5(8), 1089–1110. https://doi.org/10.1038/s41562-021-01173-x

Weisberg, D. S. , Keil, F. C. , Goodstein, J. , Rawson, E. , & Gray, J. R. (2008). The seductive allure of neuroscience explanations. Journal of Cognitive Neuroscience, 20(3), 470–477. https://doi.org/10.1162/jocn.2008.20040

Wickham, H. (2009). ggplot2: Elegant graphics for data analysis. Springer. https://doi.org/10.1007/978-0-387-98141-3

Wilke, C. O. , & Wiernik, B. M. (2022). ggtext: Improved text rendering support for “ggplot2” (0.1.2) [Computer software]. https://cran.r-project.org/web/packages/ggtext/

Wilson, P. H. , & McKenzie, B. E. (1998). Information processing deficits associated with developmental coordination disorder: A meta-analysis of research findings. Journal of Child Psychology and Psychiatry, 39(6), 829–840. https://doi.org/10.1111/1469-7610.00384

Winokur, M. , Holtan, A. , & Batchelder, K. E. (2014). Kinship care for the safety, permanency, and well-being of children removed from the home for maltreatment: A systematic review. Campbell Systematic Reviews, 2014, 2. https://doi.org/10.4073/csr.2014.2

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. Psychological Methods, 12, 399–413. https://doi.org/10.1037/1082-989X.12.4.399