OpenIntro Statistics

Fourth Edition

David Diez
Data Scientist
OpenIntro

Mine Çetinkaya-Rundel
Associate Professor of the Practice, Duke University
Professional Educator, RStudio

Christopher D Barr
Investment Analyst
Varadero Capital

Table of Contents

Preface

OpenIntro Statistics covers a first course in statistics, providing a rigorous introduction to applied statistics that is clear, concise, and accessible. This book was written with the undergraduate level in mind, but it's also popular in high schools and graduate courses.

We hope readers will take away three ideas from this book in addition to forming a foundation of statistical thinking and methods.

- Statistics is an applied field with a wide range of practical applications.
- You don't have to be a math guru to learn from real, interesting data.
- Data are messy, and statistical tools are imperfect. But, when you understand the strengths and weaknesses of these tools, you can use them to learn about the world.

Textbook overview

The chapters of this book are as follows:

1. **Introduction to data.** Data structures, variables, and basic data collection techniques.
2. **Summarizing data.** Data summaries, graphics, and a teaser of inference using randomization.
3. **Probability.** Basic principles of probability.
4. **Distributions of random variables.** The normal model and other key distributions.
5. **Foundations for inference.** General ideas for statistical inference in the context of estimating the population proportion.
6. **Inference for categorical data.** Inference for proportions and tables using the normal and chi-square distributions.
7. **Inference for numerical data.** Inference for one or two sample means using the t-distribution, statistical power for comparing two groups, and also comparisons of many means using ANOVA.
8. **Introduction to linear regression.** Regression for a numerical outcome with one predictor variable. Most of this chapter could be covered after Chapter 1.
9. **Multiple and logistic regression.** Regression for numerical and categorical data using many predictors.

OpenIntro Statistics supports flexibility in choosing and ordering topics. If the main goal is to reach multiple regression (Chapter 9) as quickly as possible, then the following are the ideal prerequisites:

- Chapter 1, Sections 2.1, and Section 2.2 for a solid introduction to data structures and statistical summaries that are used throughout the book.
- Section 4.1 for a solid understanding of the normal distribution.
- Chapter 5 to establish the core set of inference tools.
- Section 7.1 to give a foundation for the t-distribution
- Chapter 8 for establishing ideas and principles for single predictor regression.

Examples and exercises

Examples are provided to establish an understanding of how to apply methods

EXAMPLE 0.1

(E) This is an example. When a question is asked here, where can the answer be found?

The answer can be found here, in the solution section of the example!

When we think the reader should be ready to try determining the solution to an example, we frame it as Guided Practice.

GUIDED PRACTICE 0.2

(G) The reader may check or learn the answer to any Guided Practice problem by reviewing the full solution in a footnote.[1]

Exercises are also provided at the end of each section as well as review exercises at the end of each chapter. Solutions are given for odd-numbered exercises in Appendix A.

Additional resources

Video overviews, slides, statistical software labs, data sets used in the textbook, and much more are readily available at

openintro.org/os

We also have improved the ability to access data in this book through the addition of Appendix B, which provides additional information for each of the data sets used in the main text and is new in the Fourth Edition. Online guides to each of these data sets are also provided at **openintro.org/data** and through a companion R package.

We appreciate all feedback as well as reports of any typos through the website. A short-link to report a new typo or review known typos is **openintro.org/os/typos**.

For those focused on statistics at the high school level, consider *Advanced High School Statistics*, which is a version of *OpenIntro Statistics* that has been heavily customized by Leah Dorazio for high school courses and AP® Statistics.

Acknowledgements

This project would not be possible without the passion and dedication of many more people beyond those on the author list. The authors would like to thank the OpenIntro Staff for their involvement and ongoing contributions. We are also very grateful to the hundreds of students and instructors who have provided us with valuable feedback since we first started posting book content in 2009.

We also want to thank the many teachers who helped review this edition, including Laura Acion, Matthew E. Aiello-Lammens, Jonathan Akin, Stacey C. Behrensmeyer, Juan Gomez, Jo Hardin, Nicholas Horton, Danish Khan, Peter H.M. Klaren, Jesse Mostipak, Jon C. New, Mario Orsi, Steve Phelps, and David Rockoff. We appreciate all of their feedback, which helped us tune the text in significant ways and greatly improved this book.

[1] Guided Practice problems are intended to stretch your thinking, and you can check yourself by reviewing the footnote solution for any Guided Practice.