

Readings

Stats 1101 Fall 2023

The primary text is Freedman et al. 2007. What makes a text primary? I won't contradict it and it is a source of homework problems. It is good to stay close to the primary text because intro conventions are not always consistent. Some authors will use convention X for the whiskers in a box-and-whisker plot and some will use convention Y . Some apply the term selection bias more broadly than others. Some will pool a standard deviation and some will not. These details are not always apparent in an intro text.

Below, \bullet indicates a required reading and \circ indicates an optional reading. I briefly explain the significance of some readings underneath their citation. The semester includes 26 class sessions, meaning some sections are covered over several lectures. The midterm is after the probability lectures.

1 Intro

- \bullet Hand 2008, Chapter 1

What is data? What is a statistic? Statistics is useful despite what the haters say. Statistics can be misused.

2 Types of Data

- \bullet Diez et al. 2019, Chapter 1.2

This establishes vocabulary for tabular data and variables. These readings are interchangeable.

3 Experiments and Observational Studies

- \bullet Freedman et al. 2007, Chapters 1-2

4 Natural Experiments

- \bullet Rosenbaum 2017, Chapter 6

A perfect natural experiment involves “ignorable treatment assignment,” meaning the process determining treatment and control is unrelated to the outcomes of interest and as good as random.

5 Summarizing Data

● Freedman et al. 2007, Chapters 3-4

● Diez et al. 2019, Chapter 2

This fills out some of the gaps in Freedman et al. 2007.

○ Van Belle 2011, Chapter 9

This gives a few rules to help you be more thoughtful in using tables and graphs.

○ Schwabish 2021

This is a brief starting point for anyone interested in the craft of data storytelling, covering things like choosing color-blind-friendly color palettes and reducing clutter.

○ Schwabish 2023

This reference will help you make graphs, even fancy ones, in Excel.

6 Normal Distribution

● Freedman et al. 2007, Chapter 5

7 Correlation

● Freedman et al. 2007, Chapters 8-9

8 Simple Linear Regression

● Freedman et al. 2007, Chapters 10-12

9 Probability

● Freedman et al. 2007, Chapters 13-18

● Diez et al. 2019, Chapter 3.2.7-3.2.8

This will supplement for tree diagrams, Bayes’ theorem, and some formalities.

● Strogatz 2010

This covers Bayes’ theorem and discusses the usefulness of natural frequencies as a solution strategy.

10 Sampling

- 🔊 Freedman et al. 2007, Chapters 19-20

11 Confidence Intervals

- 🔊 Freedman et al. 2007, Chapters 21, 23

12 One-Sample Hypothesis Testing

- 🔊 Freedman et al. 2007, Chapter 26

13 Two-Sample Hypothesis Testing

- 🔊 Freedman et al. 2007, Chapter 27

14 χ^2 Hypothesis Testing for Categorical Data

- 🔊 Freedman et al. 2007, Chapter 28

15 Computer Skills

- 🔊 [My notes](#)

These notes will prepare you for using Python in Google Colab.

16 A Closer Look at Significance

- 🔊 Freedman et al. 2007, Chapter 29
- 🔊 TBD

Additional reading will cover Type I error, Type II error, statistical power, and the false discovery rate.

17 Simple Linear Regression

- 🔊 Diez et al. 2019, Chapter 8

18 Multiple Linear Regression

- 🔊 Diez et al. 2019, Chapter 9

References

- Diez, D. M., Barr, C. D., & Cetinkaya-Rundel, M. (2019). *Openintro statistics*. OpenIntro. <https://www.openintro.org/book/os/>
- Freedman, D., Pisani, R., & Purves, R. (2007). *Statistics* (4th ed.). WW Norton & Company. <https://clio.columbia.edu/catalog/6285515>
- Hand, D. J. (2008). *Statistics: A very short introduction*. Oxford University Press, USA. <https://clio.columbia.edu/catalog/14180088>
- Rosenbaum, P. (2017). *Observation and experiment: An introduction to causal inference*. Harvard University Press. <https://clio.columbia.edu/catalog/15176619>
- Schwabish, J. (2021). The practice of visual data communication: What works. *Psychological Science in the Public Interest*, 22(3), 97–109. <https://doi-org.ezproxy.cul.columbia.edu/10.1177/15291006211057899>
- Schwabish, J. (2023). *Data visualization in excel: A guide for beginners, intermediates, and wonks*. CRC Press. <https://clio.columbia.edu/catalog/17325179>
- Strogatz, S. (2010). Chances are. *New York Times*. <https://archive.nytimes.com/opinionator.blogs.nytimes.com/2010/04/25/chances-are/>
- Van Belle, G. (2011). *Statistical rules of thumb* (Vol. 699). John Wiley & Sons. <https://clio.columbia.edu/catalog/6883875>