Statistics 201B:
Introduction of Statistics at an Advanced Level

Haiyan Huang <hyh0110@berkeley.edu>
Department of Statistics

UC Berkeley, Fall 2025

Syllabus

Instructor: Dr. Haiyan Huang
Tu/Th 11:00am-12:29pm Lecture, 106 Stanley
Office: 317 Evans

GSI: Karissa Huang (krhuang@berkeley.edu)
W 12:00pm-1:59pm (101 Discussion Section), 334 Evans
W 2:00pm-3:59pm (102 Discussion Section), 334 Evans

GSI: Drew Thanh Nguyen (drew.t.nguyen@berkeley.edu)
W 4:00pm-5:59pm (103 Discussion Section), 344 Evans

About This Course:

- We will cover the fundamentals of statistical inference, testing, and modeling, including point estimation, confidence intervals, hypothesis testing, linear models, large sample theory, categorical models, decision theory, and classification.

- The course is primarily a survey of the theoretical basis for the statistical methods frequently encountered, but students will also do some implementation of the techniques on data.

- Prerequisites: This course expects students to have either taken Stat201A or to be taking it concurrently. Also assumed is multivariable calculus (at the level of MATH 53) and linear algebra (at the level of MATH 54).

Textbook:

1. We will be following the content of "All of Statistics" by Larry Wasserman, Section II and parts of Section III.

2. This book is available online for free for UC Berkeley: https://link-springer-com.libproxy.berkeley.edu/book/10.1007/978-0-387-21736-9#about

3. However, at times, the content of the book is not detailed enough, so the material in lectures will fill this out. Homework and exam questions will be based on the lectures!

Additional useful references:

1. "Probability and Statistics" by Morris H. DeGroot. It is detailed but still misses some modern compoenents and depth.

2. "Statistical Inference" by Casella and Berger. It is detailed and present "traditional" topics in more depth. But it lacks materials on modern topics.

3. "Theoretical Statistics: Topics for a Core Course" by Robert Keener. This book is detailed and has relatively more modern topics than the Casella and Berger book.

Note that "An Intermediate Course in Probability" by Allan Gut is a good resource if you are looking for probability background (you can also refer to Section I of Wasserman).

Online tools

We will use several different online services in the class that are probably familiar to you:

1. **Bcourses.** We will post Lecture notes, HWs, Solutions, Practice midterms, etc on bCourses. Grades, announcements, updates, and the like will also be through bCourses.

2. **Ed Discussion.** You should have received a message from your GSI Karissa about using Ed Discussion for class Q&A. It is the best place to ask questions about the course, whether academic or administrative. If you have trouble accessing it, please contact your GSIs.

3. **Gradescope.** HW Assignments will be submitted to Gradescope. Your GSIs will provide more instructions.

Grading

Your final grade will be a weighted average of your average homework score (30%), midterm (25%), and the final exam (45%).

- **Homeworks:** Problem sets will be assigned roughly each Wednesday, for a total of 9 assignments. You should download the assignments from Bcourses. Each problem set is to be turned in on Friday a week later. No late assignments will be accepted. The homework with lowest score will not be included in the final homework grade. Some problems may not be graded, and you should review the solutions carefully for those problems. Students can discuss homework assignments. Each student must write up his/her own solutions individually. Any evidence of cheating will be subject to disciplinary action.

- **Midterm:** The in-class midterm is scheduled on Thursday October 16. You are allowed to bring one double sided A4 page of handwritten notes to the exam.

- **Final:** The final exam will be cumulative. You are allowed to bring two double sided A4 pages of handwritten notes to the exam. Based on the Fall 2025 Final Exam Groups Calendar, the final exam will happen on Wednesday Dec 17 8-11am.

Other logistics

Computing: The assignments will involve small amount of data analysis or computing. We will use R, which is available for free at http://cran.r-project.org/, along with the RStudio which is a (free) very useful interface for R. The GSI will cover how to implement the necessary computations in R.
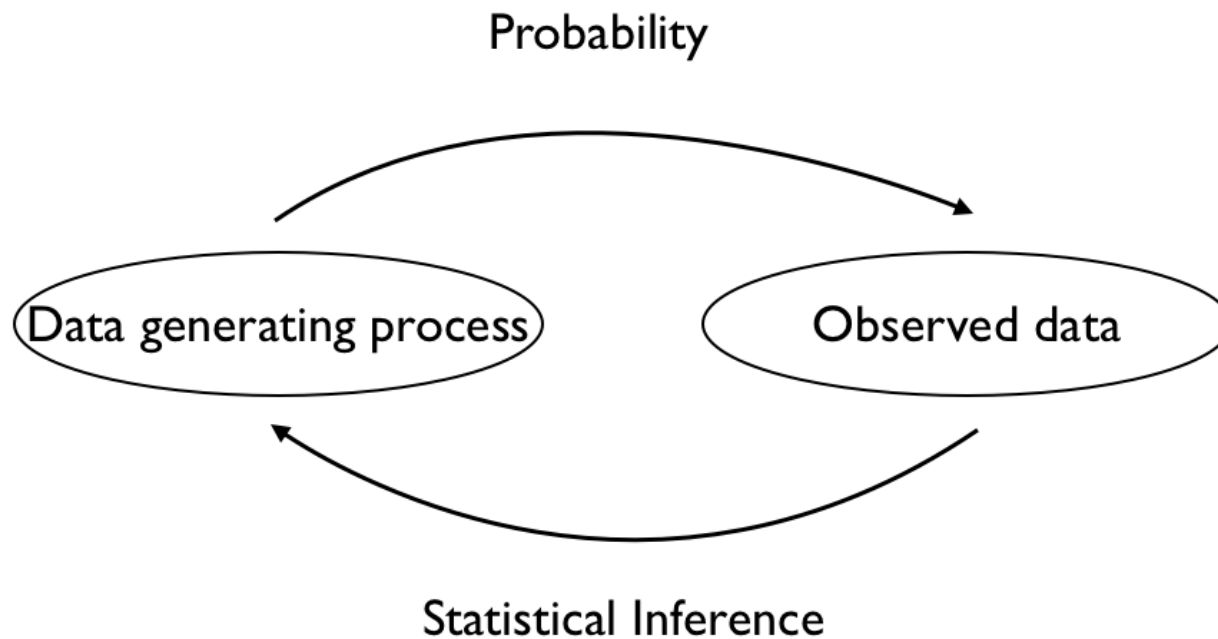
Email: Detailed homework questions will not be answered via email. Instead, please use office hours or Ed Discussion.

Students with disabilities: If you need accommodations for any disabilities, please speak to me after class or during office hours so that we can make the necessary arrangements.

Academic Integrity: Any homework, test or report submitted by you and that bears your name is presumed to be your own original work. In all of your assignments, you may use words or ideas written by other individuals in publications, web sites, or other sources, but only with proper attribution.
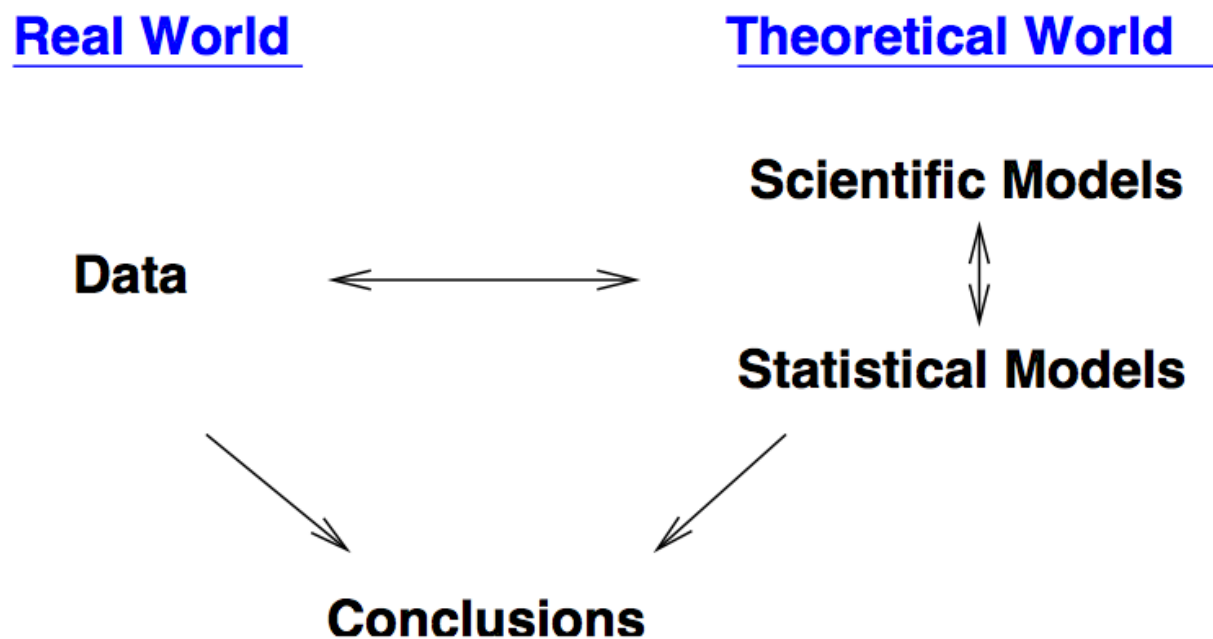
Copyright: Course material is copyrighted and reposting to third party sites or any other form of redistribution is prohibited.

The Big Picture

adapted from Wasserman, 2004

A Slightly Bigger Picture

Real World

Theoretical World

Scientific Models

Data

Statistical Models

Conclusions

from Kass, 2009

THE DATA SCIENCE LIFE CYCLE

1. Define and understand the problem
2. Data collection
3. Data cleaning, and preparation
4. Exploratory data analysis
5. Model building and deployment

By Madison Hunter