Introduction to Statistical Thought

Stat 250: Introduction to Statistical Inference

Click here for PDF version

Three prongs of statistics

1. **Design**
   The design of surveys/experiments and collection of data to more efficiently/correctly address scientific questions

2. **Exploration**
   Understand the major features of and detect patterns in data

3. **Inference**
   Account for randomness, variability, and bias in a sample in order to draw reasonable and correct conclusions about a population

Statistics vs. probability

Probability (Math 240)

We learned how to calculate the probability of seeing a result (data) given a specific probability model (e.g., a specific distribution)

Statistics (Stat 250)

We will learn how to make statements about the underlying probability models given the data we see

Example: Spies vs. Statisticians

During WWII, the Allies wanted to determine production rates of tanks (and airplanes, missiles, etc.)

Spies

Gathered intelligence (intercepted messages, interrogated of prisoners, etc.) and made the following estimates:

- June 1940: 1000
- June 1941: 1550
- August 1942: 1550

Example: Spies vs. Statisticians

Statisticians

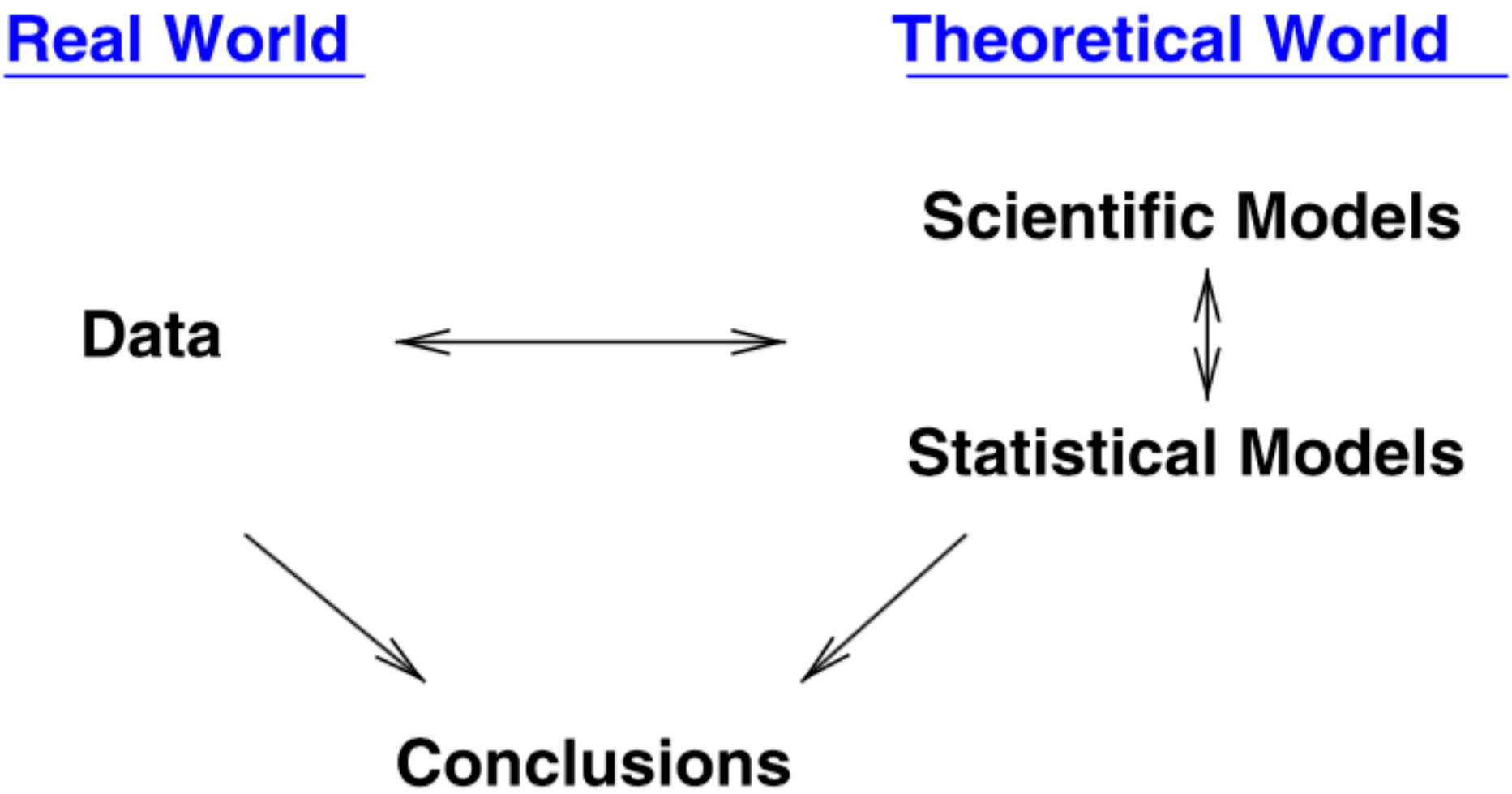
- The Allies had a sample of serial numbers (via capture, photography, etc.), $X_1, X_2, \ldots, X_n$, and there were N produced.
- Allied statisticians needed to devise an **estimator** to obtain N
- Ultimately, they used $\widehat{N} = X_{max} + \frac{X_{max}}{n} - 1$
  - June 1940: 169
  - June 1941: 244
  - August 1942: 327

Example: Spies vs. Statisticians

After the war, the Allies discovered documents revealing the true number of tanks produced:

Month	Truth	Statisticians	Spies
June 1940	122	169	1000
June 1941	271	244	1550
August 1942	342	327	1550

The big picture



Statistical models

A statistical model consists of

- a collection of random variables to describe observable data,
- the possible joint distribution(s) of the random variables,
- and the parameters, $\boldsymbol{\theta}$, that define those distributions

(Morris and DeGroot, 377)

Types of Inference in Stat 250

Nonparametric

> The basic idea of nonparametric inference is to use data to infer an unknown quantity while making as few assumptions as possible. Usually, this means using statistical models that are infinite-dimensional. (Wasserman, 2006)

Parametric

> A parametric inference uses models that consist of a set of distributions/densities that can be parameterized by a finite number of parameters.

Tentative schedule

Topic	Chapters	Approx. Duration
Exploratory Data Analysis	1-2	1 week
Nonparametric inference	3-5	3 weeks
Parametric inference	6-10	6 weeks

Types of Studies

Experimental

Investigators apply treatments to experimental units (people, animals, plots of land, etc.) and then proceed to observe the effect of the treatments on the experimental units

Observational

Investigators observe subjects and measure variables of interest without assigning treatments to the subjects. The “treatment” that each subject receives is determined beyond the control of the investigator.

Your turn

- Introduce yourself to your neighbors
- Discuss the questions on the worksheet
- Be prepared to report back

Syllabus highlights

Where to find materials

Announcements	→ Ed
Syllabus, slides, handouts, etc.	→ Moodle
A place to ask questions	→ Ed
Homework submission portal	→ Gradescope

Drop-in hours

Regular drop-in hours are held in CMC 307, just stop by if you want to ask a question or chat

Day	Time
Monday	3:15-4:15 pm
Tuesday	1:00-2:00 pm
Wednesday	11:00-12:00 (noon)
Friday	11:00-12:00 (noon)

Assignments

- After each class a few problems will be assigned
- Problems are collected on Fridays via Gradescope
- You get 2 no-questions-asked late days
  (just fill out the Google form)

For all the details

- Read the syllabus
- Complete HW0 before next class
- Come to class on Wednesday with any questions