

Welcome

ST551 Lecture 1

Charlotte Wickham

2017-09-20

Introductions

My name is Charlotte. . .

- I'm an Assistant Professor in the Statistics Department (half-time)
- I'm from New Zealand, but I've lived in the US for 12 years
- Last year, I taught only online classes
- I'm a cat person
- In my "free" time I'm renovating a 117 year old house

Introduce yourself to your neighbors:

- Where are you from?
- Why are you taking ST551?
- Two other interesting facts about yourself

What are we doing here?

ST551 Statistical Methods “Properties of t, chi-square and F tests; randomized experiments; sampling distributions and standard errors of estimators, delta method, comparison of several groups of measurements; two-way tables of measurements.”

- Understanding the fundamental concepts that allow statistical inference
- The subtleties of how study setup, choice of methods, and assumptions impact the kind of inferences we can make

Along the way:

- Learning a ton of statistical methods
- Practice communicating findings in a correct and clear way
- Building our R skills to analyse and explore data and use simulation to understand properties of methods.

Class logistics

Two important websites

Class webpage: <http://st551.cwick.co.nz>

- Syllabus
- Lecture slides
- Lab materials
- Homework assignments

canvas:

<https://oregonstate.instructure.com/courses/1653112>

- Announcements
- Discussion
- Homework submission
- Grades

If you have questions. . .

Class content, homework, class logistics, R:

- Come to office hours: mine (Mon or Wed 2-2:50pm Weniger 255) or TAs (TBA)
- Use canvas discussion board!

Email me only for situations of a personal nature.

Highlights from the syllabus

Grade = 40% homework + 25% midterm + 35% final exam

- Lowest homework score dropped
- Midterm: Friday, October 27th in class
- Final: Thursday, December 7th 9:30am-11:20am

Getting started with R and RStudio

Two options for using R and RStudio

1. RStudio Server:

`https://rstudio.cosine.oregonstate.edu`

Log in with your ONID credentials

2. Install R and RStudio on your own machine.

R and RStudio are also installed on Lab computers and in Cordley 3003.

Never used R before? Use the server, and work through Chapter 1 - Hands on Programming with R

Fill out data collection forms

Statistical inference

Components of a dataset

Observational units The units on which measurements are made.

Variable(s) measured for each observational unit. Any characteristic that can be measured/recorded for each observational unit.

- Quantitative (e.g. height, temperature), or
- Qualitative (e.g. hair color, shirt size).

- **Summarize**
 - Succinctly describe a dataset (ideally using fewer numbers/words than it would take to describe the entire dataset)
 - Provide plots/graphics to convey information about the dataset
- **Predict** Use current dataset to predict values for new observations
- **Infer** Use dataset to provide probabilistic answer to a question in a broader context

Two common types of statistical inference

A.K.A “Scopes of inference”

Causal inference

The pattern seen in the data was **caused** by an intervention or treatment (or value of another variable).

Population inference

The pattern seen in the data can be inferred to a wider population.

Example

I compare the commute times for students in the class that walked versus those that drove.

Observational unit: student

Variables: commute type, commute time

Observed pattern: Students in this class who walk have a lower average commute time than those who drive.

- **A population inference:** OSU students who walk have a lower average commute time than those who drive.
- **A causal inference:** Walking decreases your commute time, for students in ST551 Fall 2017.
- **Both population and causal inference:** Walking decreases commute time.

Your turn: Fill in the blanks

I survey 50 OSU students with a GPA ≥ 3 and 50 OSU students with GPA < 3 .

Observation: The students surveyed with “high” GPA have, on average, fewer friends on Facebook, than those surveyed with “low” GPA.

Observational unit: _____

Variables: _____

Population Inference: _____

Causal Inference: _____

Inferential language in the media

Get in groups of 3-4. I'll provide you an article.

Discuss then answer:

1. What type of inference is being **implied** by the headline?
2.
 - If population inference, describe the population(s)?
 - If causal inference, what is the treatment/intervention?(It may be both or neither)
3. Rewrite the headline to clarify the inference you think is being implied.
4. If you have time, can you figure out from the body of the article: what the observational units were, what variables were measured, and what pattern was observed in the data?

Next time . . .

When are population and/or causal inferences justified?