

Scope of Inference

ST551 Lecture 2

Charlotte Wickham

2017-09-22

Announcements

Homework 0 is posted

Due midnight Thursday Sep 28th on canvas.

Part 1, like lecture, find a news article and answer questions about the inferences implied and justified.

Part 2, get up and running with R.

From last time. . .

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.

Headlines from last time

- “59 Percent Of You Will Share This Article Without Even Reading It”
- “A Glass Of Red Wine Is The Equivalent To An Hour At The Gym, Says New Study”
- “Gut Bacteria May Play a Role in Weight Loss”
- “Get up at least once every 30 minutes. Failure to do so may shorten your life, study finds”
- “Bats crash into windows because of a glitch with their ‘sonar’ ”
- “Lightning storms triggered by exhaust from cargo ships”

When are inferences justified?

Two common study designs

1. Random Sampling study
2. Randomized Experiment

Random sampling study

- A population(s) is defined
- Units are **randomly sampled** from the population(s)
- Units are observed

Statistical methods focus on quantifying the uncertainty in values population parameters.

Randomized experiment study

- A group of units is selected
- Units are **randomly assigned** to different levels of a treatment variable
- Units are observed

Statistical methods focus focus on quantifying the uncertainty in the existence and size of treatment effects

What kind of study?

“A Glass Of Red Wine Is The Equivalent To An Hour At The Gym, Says New Study” <http://onlinelibrary.wiley.com/doi/10.1113/jphysiol.2012.230490/full>

Fifty 8-week-old male Wistar rats were obtained from Charles River Laboratories (Sherbrooke, QC, Canada).

Following acclimation to the treadmill, 10-week-old male Wistar rats were randomly divided into four groups, which included sedentary rats or exercise trained [...] rats that received control AIN-93G diet (Control) or the AIN-93G diet supplemented with resveratrol (RESV; 4 g RESV/kg food).

What kind of study?

“59 Percent Of You Will Share This Article Without Even Reading It” <https://hal.inria.fr/hal-01281190/document>:

Using multiple data collection techniques, we are able to jointly study Twitter conversations and clicks for URLs from five reputable news domains during a month of summer 2015.

One may specify to access a 1% sample of tweets. This is the way we discover URLs for this study.

What kind of study? Bats Part I

Is this random sampling?

Field experiment part of “Bats crash into windows because of a glitch with their ‘sonar’ ”

*We performed experiments at three different bat colonies in South-West Hungary in June 2011. The first colony [...] consisted of greater mouse-eared bats (*Myotis myotis*) and Schreiber's bats (*Miniopterus schreibersii*). The second [...] and third colonies [...] consisted of soprano pipistrelles (*Pipistrellus pygmaeus*).*

from Supplementary Materials @ [http:](http://science.sciencemag.org/content/357/6355/1045.full)

[//science.sciencemag.org/content/357/6355/1045.full](http://science.sciencemag.org/content/357/6355/1045.full)

Is this a randomized experiment?

After all individuals had left the colony we placed one or two smooth, black, flexible plastic plates (2 x 1 m or 2 x 2 m when combined) vertically 1-3 m from the colony entrance but never on the actual main flight path. We observed returning bats for four hours with an infrared camera (SONY HDR-CX560V, 60 frame/sec, night vision mode) while presenting the plate uncovered (i.e. smooth) or covered with a rough, ribbed plastic mat or branches (alternated in 15 minute intervals).

Short answer to the big question

Population inferences are statistically justified only if the study is a random sampling study. Inference can be made to the population that was sampled from.

Causal inferences are statistically justified only if the study is a randomized experiment. Effects can be attributed to the treatment that was randomized.

Both? If a random sample from a population is taken, **and then** units are randomly assigned to treatments, both population and causal inferences can be made. The effects can be attributed to the treatment, and generalized to the population.

Justified inferences?

“A Glass Of Red Wine Is The Equivalent To An Hour At The Gym, Says New Study”

Population inferences: not justified

Causal inferences: justified (but limited to actual treatments on studied rats on measured endpoints)

Can you figure out what inferences are statistically justified in the other articles?

Why random sampling?

To make a population inference that data must be **representative** of the population.

A random sample ensures the representative-ness (and gives us a probability model for quantifying our inferences):

Random sampling could include:

- Simple random sampling
- Systematic sampling
- Cluster sampling
- Stratified random sampling . . .

Why a randomized experiment?

A study that isn't a **randomized experiment** is often referred to as an **observational study**.

Causation is hard to assess in an observational study because:

- no way to determine the direction of causation
- both “treatment” and response could be related to some other variable: **a confounder**

Randomization ensures no other variables (whether we know they exist or not) impact which treatment a unit receives.

Confounders

A relationship between a predictor variable X and an outcome of interest Y is affected by a confounding variable Z if:

- Z is related to the values of the predictor variable X , **AND**
- Z affects the values of the outcome variable value Y

E.g. *Commute type and time*: the way you commute is associated with how far you commute, and how far you commute affects your commute time. Commute distance is a confounding variable.

Your Turn:

From last lecture: I survey 50 OSU students with a GPA ≥ 3 and 50 OSU students with GPA < 3 .

Discuss with neighbor

How would the study need to be done to ensure I can make a **population** inference to OSU students?

How would the study need to be done to ensure I can make inference about whether high GPA **causes** fewer friends on Facebook?

How would the study need to be done to ensure I can make inference about whether fewer friends on Facebook **causes** high GPA?

Real life isn't always easy

Random sampling studies and/or randomized experiments aren't always feasible:

- It's unethical to randomly assign people to something that may harm them (i.e. smoking)
- *Global warming* - we can't observe an Earth without increased CO₂ emissions
- *Bat study* - how can you randomly sample from all bats?

Non-statistical justifications

We've focused on inferences that can be justified statistically. That means we can quantify the uncertainty in our inferences.

But there are other ways to make inference:

- **Expert knowledge:** Perhaps echo-location is similar enough across all bats that it's reasonable to infer what is true for the bats in the study is applicable to all bats.
- **Bradford-Hill Criteria:** principles for establishing evidence of a causal relationship often in public health research.
- You don't have a random sample, but you can convince people it is somehow representative enough of a population to be useful (be very wary of any statistical inferences in this setting)

Next time . . .

We'll spend a few weeks talking in depth about inference in the random sampling model.

Coming up next week:

- review describing distributions, the Normal distribution,
- sampling distributions,
- Central Limit Theorem