Intro to NRES 710

Welcome!

Last compiled: 2024-08-19

Introductions

Introductions among faculty (Brian Folt) and students

Class structure

We will discuss the course organization and expectations. Please see course syllabus and schedule documents, provided separately.

Office hours: no office hours, best thing to do is send me an email and find a time to meet; I am happy to meet with you; my office is in KRC 100, enter from the north side of the building.

Prerequisites: Students are expected to have a basic understanding of standard statistical concepts and methods, obtained through other coursework. They should be familiar with what a p-value is, an ANOVA, a t-test, these sorts of basic things. But, we are going to cover all of these topics in class – with a practical emphasis.

Attendance: This is a graduate student class and attendance is not mandatory; you are all adults and make your own decisions. Life happens & you will be busy with field work, other obligations, whatever. It's okay.

Schedule: Tuesdays we will have a lecture, Thursdays we will have a lecture and a mini-lab. Lab assignments will be often due before the Thursday lab. Sometimes these will be readings that we will discuss on Thursdays, other times they will be homework assignments.

Exercises: These will be due periodically throughout the semester. I encourage you to work on these with your peers! The best way to learn something, is if you can communicate it to someone else. However, the goal of the class is to learn how to analyze data. So I encourage you not to lean too heavily on your peers and instead make sure that you are learning the material yourselves.

In this class, we will be learning how to analyze data in R. Why in R? Because it's free, open-source statistical and it is accessible to anyone with a computer and internet connection.

Course website! A central location for class materials.

No text books. Free resources from the internet. See the 'Links' page on the course website.

General overview

NRES 710 is designed to be an introductory, graduate-level statistics class. "Introductory" because you are not assumed to have much in the way of prior statistical knowledge (although an undergrad course under your belt will be helpful) or prior experience in R programming (although some prior exposure will make the learning curve less steep). "Graduate-level" because once we get rolling, it'll be full steam ahead! However, another dimension of a graduate-level course is an emphasis on collaborative knowledge development through group discussion and collective learning, and NRES 710 will provide opportunities for this.

This course provides an introduction to both statistics and computer programming. With a solid grounding in basic rules of probability and statistics, and armed with some basic computer programming abilities (and some human creativity and ingenuity), you can go extremely far in making sense of data and communicating that understanding with others!!

What is science?

Science is the search for *truth* through the accumulation of *facts*. It is a method of answering questions.

- **Truth** is the way things really are, what's really happening, what we want to know about how the world works.
  - The problem is that we can never really know truth. We can seek to understand it and try to know it, but we can never know truth.
- **Sampling error** – there are always errors in measurements.
- How tall is Dr. Brian Folt? If everyone in the class tried to measure me to the mm, we would get many different measurements.
- We can decrease error by standardizing measurers and measurement methods.
- **Process error** – things that contribute to variation in whatever is being measured. This is harder to understand.
- **Q:** How much taller are UNR Students than UNR Faculty? Students are maybe a little taller on average. . . ? But there is overlap. What causes this overlap?
- Other things contribute to the outcome on an individual by individual basis What things might cause height to vary among students and faculty? Age diet, genetics, behavior, etc.
- 'Noise' in the system can be due to process error or random variation.
- Why does process error make it difficult for us to understand truth?
  - How would we test for height differences between Students and Faculty?
  - Sample!!
  - If our sample is representative, then we might be able to approximate 'truth'.
  - But, if our sample may not be representative, due to random chance and process error

Q: How could we ensure TRUTH?

If we sample the entire population, we can approximate truth! But it's often impractical to get the whole population.

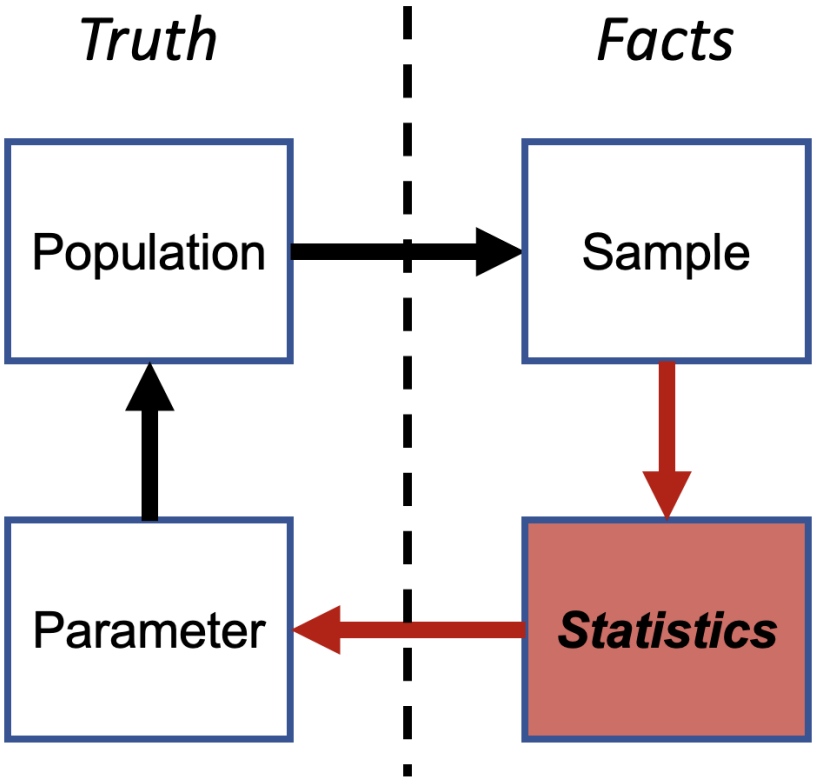
Facts are estimates of truth about our sample

- Measurements. Facts might be accurate, inaccurate; they might be reliable and repeatable, or they might not be reliable and repeatable.
- We go out and we collect data, and these data are facts.

How does 'statistics' come into this? **Statistics is the method by which we use facts (data) to estimate truth.**

- We collect data (facts) and use statistics to calculate an average (a statistic), which is an estimate of truth, based on facts.
- You can't do science without statistics.

We can use statistics as a process to make inference about a property of a *population* (a parameter) from a representative *sample*. In the figure below, the vertical line separates the stuff we can observe and measure directly (*facts*; right side) and the stuff we can't observe but want to make inference about (*truth*; left side).



Do you see the 'p-value' anywhere in here?

- P-values are not statistics. They are not estimates of truth.
- Most statistic classes emphasize P-values way too much, in my opinion. That's why we are going to read papers about on this topic before class on Thursday.

In we are going to learn how to use statistics to estimate truth. We will learn about what a p-value is and how it is used, but try to de-emphasize p-values and place more emphasis on measure effects and uncertainty around those effects. I also mentioned that truth is impossible to know, so how can we know if we are doing a good job of estimating it? We will SIMULATE DATA to know truth, so that we can then fit statistical analyses and ensure that we are correctly estimating truth.

The Statistical Program R

R is an open-source project, and new packages are being added all the time.

R is incredibly powerful and feature rich. You are NOT expected to memorize syntax right away, but rather just know that the answer is always a few clicks away! 'help' files in R are extremely useful, as is information about how to use R that is found in books or on the internet.

'Base R' is the default software built into R that does not including loading any additional packages. Here is a 'base R' cheat sheet; this is a great reference for most of the basic tasks you will need to perform in R.

Learn to use R scripts, and save your scripts frequently! This is the primary record of what you've done and allows you and others to reproduce your workflows.

If you have a problem, Google it! Someone has likely had the same problem as you in the past and asked for solutions online. Or, ask ChatGPT for help on how to solve your coding problem in R. When either

Googling or prompting ChatGPT for help, be mindful that potential solutions presented to you may not be ideal for your problem.

Install R and RStudio

CRAN website for downloading R
RStudio main site

Make sure you have the most recent versions!

Make a new RStudio "project" for this class

R projects are an **extremely useful** feature of RStudio. Just store all the code and data for this course in your project folder and it will make your life much, much easier!

"StatsChats"

Faculty from various departments – Paul Hurtado (math/stats dept), Ken Nussear (geography), Perry Williams (NRES), Kevin Shoemaker (NRES) and others – have hosted informal sessions for grad students in EECB, NRES, Geography etc (and faculty) to discuss data analysis questions. This can be a good opportunity to ask questions about statistics, find out more about what types of data and questions your peers are working with, and help others (thereby reinforcing concepts for yourself!). Keep an eye on that website linked above and if the group identifies a regular meeting time, I will relay that information to you all in class.

Feedback

This is my first time teaching this class. It was originally developed by Dr. Kevin Shoemaker, but I am making adjustments to cater it to my understanding of statistics and the best ways that I can teach you all about it.

Given this is my first time teaching this specific material in this format, I welcome frequent and honest feedback on course topics to understand whether course topics are appropriate and/or useful. So, any feedback is welcomed – thank you!

–go to next lecture–