Weeks 10 and 11 - Answering your own questions

Nicholas Kortessis

2025-03-26

Let's do some science and collect some data

This week's lab is going to be different. Up to now, we've been using data sets to help illustrate estimation and hypothesis testing. That's all good and well, but there is a lot to be learned by doing things yourself.

So here is your task in its most general form:

1. Ask two questions.
2. Determine whether your questions are about estimation or about testing hypotheses.
3. Identify statistical individuals and the characteristics to measure that are relevant to your questions.
4. Design a sampling scheme. Do your best to get a random sample.
5. Identify the kind of statistical analysis you will do. Will you be constructing confidence intervals? Will you be running a hypothesis test?
6. Go out and collect the data.
7. Input the data into a spreadsheet and save it as a .csv file. Statistical individuals go in as rows. Individual characteristics go in as columns.
8. Load the data into R and do your analysis.
9. Make figures to help graphically illustrate your results.
10. Make a conclusion using the data and analysis about the answer to your question.

At each step, check in with me, Dr. Kortessis, to get feedback on your decisions. I can help narrow or broaden questions as needed. I can help determine whether you are doing estimation or hypothesis testing and help you strategize about sampling designs and decisions about analysis.

DO NOT limit yourself by focusing on questions that can be answered using analyses we have discussed in class. I can help you decide on the right kind of analysis. Let the question be primary and the analysis be secondary.

This is a group exercise

I want you to work in groups of 3 or 4. Here are your groups (randomly selected of course!). Let me know if people are missing.

```
students <- c('Dom', 'Olivia', 'Reesa', 'Lily', 'Janani', 'Chris', 'Ray', 'Luke',
              'Hope', 'Mishel', 'Katie', 'Nate', 'Alton')

set.seed(1)
groups <- matrix(
  c(sample(students, length(students), replace = F),
    rep(NA, 16-length(students))),
                nrow = 4, ncol = 4, byrow = F)
row.names(groups) <- c('Group 1', 'Group 2', 'Group 3', 'Group 4')
colnames(groups) <- c('Student 1', 'Student 2', 'Student 3', 'Student 4')

print(groups)
```

```
##         Student 1 Student 2 Student 3 Student 4
## Group 1 "Hope"    "Olivia"  "Alton"   "Luke"
## Group 2 "Lily"    "Janani"  "Chris"   NA
## Group 3 "Ray"     "Katie"   "Nate"    NA
## Group 4 "Dom"     "Reesa"   "Mishel"  NA
```

What you need to submit

1. A document detailing your question, approach, analysis, and findings. It must include the following:

a. The names of all your group members.

b. Your group's question.

c. The statistical individuals and their characteristics.

d. A sampling design. You need a description of how you selected individuals. Saying you sampled randomly is insufficient. You need to describe exactly what you sampling protocol was such that I could replicate it if I wanted.

e. A description of the analysis you will do and why it is warranted for the question and data you have.

f. Results of the analysis, including at least one figure.

g. A paragraph that answers the question you posed. The answer should be supported by test results and figures that clearly show the answer.

h. Finally, you need some discussion of how your conclusions might be influenced by your sampling design. For example, you couldn't randomly sample, and so your data are biased in some way. But you know that it is biased, so you can say something about that bias' effect on the statistical test.

2. The data you collected in .csv file form.
3. The code used to run the analysis.

Timeline

We will do this for the next two weeks. That means you have plenty of time to get this done during class time.

For the first week, I want you to prioritize getting as much **quality** data as possible. There are multiple members of each group. I expect you to collect quite a bit of data given the fact that there are multiple members per group and you have multiple hours to do the task.

For the second week, I want you to prioritize data analysis and inference (i.e., learning as much from the data you have as possible).

Due for lab 10 (due date: Wednesday, April 2nd)

1. The document including only parts a-d detailed above
2. The data saved as a .csv file

Due for lab 11 (due date: Wednesday, April 9th)

1. An updated document from lab 10 that now includes parts e-h detailed above.
2. The code used to do the analysis as an .R script.

How to get started

Wake Forest has a lot of features that are amenable to asking questions. Here are just some places to begin thinking about asking questions.

The wooded area to the west of Winston Hall.

- There are many flowers currently in bloom. You could questions about the determinants of flowering, their spatial distribution, or even how many are flowering.
- There are lots of trees as well. You might have questions about tree diversity, tree size, tree spatial distribution, or whether they have currently leafed out after being dormant during the winter.
- There are many birds in and around Winston Hall and the surrounding woods. You could ask about the use of different places by birds or about bird diversity.
- Pollinators are starting to emerge this spring with all the flowering plants. You might follow them around to ask about pollinator behavior or you might sit at some flowers and measure characteristics about how frequently pollinators show up and what they do.
- All these are pretty easy to do. If you want help identifying species of animals or plants, you can do two things: 1) Take pictures and ask me. 2) Use apps like "Merlin ID" (for birds) and "Seek" for other plants and animals.

The buildings on campus

- You might ask questions about the how common particular features are of buildings on campus. How many departments does each building house? How many windows does the average building have? What is the average age of the buildings (they have dates of construction on the sides of the building)? Do some building have more elevators than others?

The cars around campus

- Do students drive nicer cars than faculty? Do freshmen drive nicer cars than seniors? What type of car is driven by students and faculty at WFU? What is the frequency of electric cars versus gas cars on campus? How many people at WFU take motorcycles to campus, rather than cars? How long is it between arrivals of the shuttles? How long does the average students wait for a shuttle to go downtown?
- These questions and many others can be answered pretty easily and the measurements about individuals are pretty straightforward to do. Many of the features of cars can be measured just by looking, or at least by searching for information on the model of the vehicle.

The people on campus

- There are people all over campus and they have characteristics that are also pretty easily measurable.
- You can ask about how many people come in and out buildings every day. You could ask whether business school students dress differently than students in the sciences, the arts, and the humanities. What is the (perceived) gender ratio of
- If you are feeling particularly outgoing today, you could ask students questions about current topics by running a survey. You could ask how many classes they have each day fo the week to estimate what the average student's course schedule looks like. You could ask about current political issues. You could ask about current issues in popular culture. It's currently March Madness, and so you might ask how much everyone loathes Duke basketball on a scale of 1-10. You could also ask students their major to find out whether these answers differ by students in different majors.

How your work will be evaluated

Each lab will be graded separately (data collection and data analysis). Here is the points breakdown.

- 50% will be for submitting all parts
- 50% will be for evaluation of the suitability of the approach and analysis to answering the question.

Grades will be identical for all members of the group. Talk now about splitting up effort and how to work collaboratively.