

Homework 0

Noah Simon

Jan 2, 2017

Please turn this in Thursday, 1/5/17 by 5:00 pm.

Background Questions

Please answer these questions honestly and concisely! (the background questions shouldn't take more than 15 minutes max) Your answers will be very helpful in guiding the course.

- 1) Why are you taking this course / What are you aiming to get out of this course? (eg. "I have some data to wrangle/analyze for my thesis, and was interested in learning how to do it." or "I have a CSE background, but am not interested in the usual programmer route, and have heard that applying programming skills to 'data' is new, exciting, and lucrative!" or "I am interested in epidemiological problems, and know that the ability to deal with data is important. I was hoping to build those skills")
- 2) Are there particular topics you would like to see? (eg. databases, analysis of text or image data, high throughput 'omics data, machine learning tools for prediction)
- 3) What is your programming background like? (eg. "I have basically never programmed, but saw a little bit of R in biostat 511". or "I took a semester of programming in Java" or "I'm basically a computer. 01101011 01101001 01101100 01101100 00100000 01100001 01101100 01101100 00100000 01101000 01110101 01101101 01100001 01101110 01110011")
- 4) What is your familiarity with biomedical data; and what type of data are you most interested in looking at (eg. 'omics data, electronic health record data, health survey data, imaging data, etc...)
- 5) Do you have a question/data in mind for the project? If so, what is it? Also, would you be more interested in an individual or group project?

Thought question

You do not need to write anything for this, but come to class prepared to discuss.

In class, for the nslc-data, we talked about evaluating if various values of π_T (the response probability for a patient on treatment) were consistent with the our data (the proportion of patients on the treatment arm who responded). In practice we are generally more interested in checking if the statement

The response probability on treatment (π_T) is no higher than the response probability on control (π_C)

is consistent with the data. How might we evaluate this statement, using the programmatic ideas we have been considering?