Course logistics

Jeffrey Leek, Assistant Professor of Biostatistics
Johns Hopkins Bloomberg School of Public Health

Pre-requisites

- There are no formal pre-requisites
- But all data analysis will be performed in the R programming language: http://www.r-project.org/
- You can find some videos on how to install R here: http://bit.ly/UCJI9M.
- Having a very basic knowledge of R will make the class much more accessible
- If you want to assess your knowledge, take this self-graded R pre-quiz: http://www.biostat.jhsph.edu/~rpeng/coursera/selfquiz/ (this quiz does not count toward your final grade for Data Analysis)
  1. Give yourself 1 point for each correct answer.
  2. If it takes you less than 1 hour and you get a score of 10 or higher you should have no trouble with the level of R in the course
  3. If it takes you more than 1 hour or your score is less than 10, you might want to check out course videos for Computing for Data Analysis here: http://bit.ly/UC5UDc

Why R?

- It is free.
- It is the most popular language for data analysis.
- Typing is better than point-and-click
  - Easier to communicate
  - Reproducible
  - Requires more thought
- It has a huge number of useful packages (as you will see)

Course Structure

- My goal is to make all videos 10-15 minutes
- Several topics may be broken down into sub-components
- R code will be included in the slides
- Slides will be availble in pdf and html form.

Grading

- There will be a total of 8 weekly quizzes each worth 10 points.
- There will be two peer-reviewed data analysis reports worth 40 points each.
- There are 160 total points for the course
- To earn the certificate for the course you need to earn 100 points.
- To earn distinction for the course you need to earn 144 points.

Grading

- You may attempt each quiz up to 4 times. Only the last attempt will count.
- The data analysis you submit will be scored by your peers using a defined rubric. Your final score for the data analysis will be the median of the peer review scores.
- You have up to 5 late days during the course of the term, which you may use on the quizzes.
- You may not use late days on the peer-reviewed assignements.
- See the course logistics page for assignment due dates.

Scoring for the data analysis assignments

- You will get one week after the data analysis deadline to complete peer review of the assignment.
- Each data analysis assignment has four parts: the main text, a figure and caption, the references and R code.
- Each part will be scored on multiple criteria.
- When grading your peers you will give 0-5 points for each criteria.
- The final score will be the percentage of available points multiplied by 40.

Data analysis rubric

Main text

- Does the analysis have an introduction, methods, analysis, and conclusions?
- Are figures labeled and referred to by number in the text?
- Is the analysis written in grammatically correct English?
- Are the names of variables reported in plain language, rather than in coded names?
- Does the analysis report the number of samples?
- Does the analysis report any missing data or other unusual features?
- Does the analysis include a discussion of potential confounders?
- Are the statistical models appropriately applied?
- Are estimates reported with appropriate units and measures of uncertainty?
- Are estimators/predictions appropriately interpreted?
- Does the analysis make concrete conclusions?
- Does the analysis specify potential problems with the conclusions?

Data analysis rubric

Figure

- Is the figure caption descriptive enough to stand alone?
- Does the figure focus on a key issue in the processing/modeling of the data?
- Are axes labeled and are the labels large enough to read?

References

- Does the analysis include references for the statistical methods used?

R script

- Can the analysis be reproduced with the code provided?

Typos/errors/differences of opinion

- I'm prone to a typo or two
- This is my first time giving video lectures
- I'm happy to get feedback in the "Feedback" forum
- I'll try to address as many of the issues as I can
- Keep in mind that currently data analysis is as much art as it is science.

Getting the slides

- Slides for this course were created with Slidify: http://ramnathv.github.com/slidify/.
- They are available from https://github.com/jtleek/dataanalysis.
- To re-compile the slides:
  1. Download the directory containing the lecture from Github
  2. Set the working directory to the lecture directory
  3. Install Slidify
  4. Run the following commands:

```
library(slidify)
slidify("index.Rmd")
```