Using Active Learning to Teach Data Analysis to Undergraduate Students

Therri Usher

Thursday, August 13, 2015

Johns Hopkins University

- First research university in the United States
  - "Dedicated to advancing both students' knowledge and the state of human knowledge through research and scholarship"
- Still remains a leader in medical and public health research

---

1“A Brief History of JHU”. [link]

JHU’s Biostatistics Curriculum

Graduate
- 4 terms (half-semesters) of biostatistics
- 4th term course includes a data analysis project

Undergraduate
- 2 semesters of biostatistics
- Optional R programming lab
- Recent introduction of data analysis seminar

JHU’s Biostatistics Curriculum

- Walk into a lecture auditorium
- Listen to a 75-80 minute lecture
- Receive homework assignment or quiz
  - Data is provided, cleaned and processed
  - Purpose of analysis is provided
  - Steps of analysis are clearly outlined
- Provide the correct answer to questions
- Take a multiple-choice final exam
JHU’s Biostatistics Curriculum

Students are expected to become proficient in biostatistics AND are expected to perform data analysis in this learning environment.

Reason For A New Course

The current structure of teaching biostatistics and data analysis does not align with the tenets of data analysis. The course and class structure of this new data analysis course is designed to align with the process, themes, and applied, collaborative nature of data analysis.

Course Structure

- Review of R
- Reproducible Research
- Forming a Research Question
- Data Cleaning and Processing
- Exploratory Data Analysis
- Statistical Analysis
- Model Checking and Evaluation
- Reporting and Defending Results
- Communicating Results / Creating Data Products
Course Themes

1. Data analysis versus statistical analysis
2. Ambiguity of data analysis
3. Data science as an evolving field

Classroom

Class Structure

Announcements, administrative tasks, review: 15-20 minutes
Instruction: 35-45 minutes
  ▶ Interactive lectures
  ▶ Discussions
  ▶ Questions
Active learning exercise: 15-20 minutes

Active Learning Exercises

Purpose:

1. Students can actively apply knowledge that they have gained from the instruction.
2. Students gain practice in skills needed for their data analysis project.
3. ALE’s allow for formative assessment throughout the course.
Active Learning Exercises

Types:

- Programming exercises
- Conceptual exercises
- Debate
- Group teaching

Discussions and questions: Group
Active learning exercises: Individual

Examples of Active Learning Exercises

Use the mpg data set in the ggplot2 package. Model the city miles per gallon and determine if the transmission is associated with it.

- Decide the proper GLM
- Describe why it is appropriate, addressing the assumptions
- Create the linear predictor, including potential confounders
- Fit the model in R
- Interpret the results

Examples of Active Learning Exercises

Create a framework for your final project, i.e. a framework for the relationship between your variable of interest and your outcome.

Draw it out, similar to the diagram presented during the lecture. Be sure to indicate what is known in the scientific literature and what you hypothesize to be true. Discuss ideas with your classmates.

Examples of Active Learning Exercises

Each group will be assigned one of the following topics about data analysis:

- Research questions
- Data cleaning and exploratory data analysis
- Statistical modeling and model fit assessment
- Reproducibility and reporting results

Your group will have to “teach” its assigned topic to the rest of the class and myself. You will have 30 minutes in class to work together and create slide(s) for teaching your topic. Feel free to include personal beliefs and suggestions regarding your topic, but acknowledge that they are your own opinions.

Once the 30 minutes is up, each group will present their topic. Someone from each group must email me their group’s slides after the presentations.
Examples of Active Learning Exercises

Data Analysis Project

Each student was required to complete a data analysis project. They had to:

- Form a research question
- Find a data source
- Clean and process the data
- Perform exploratory data analysis
- Fit an appropriate statistical model
- Evaluate and compare models
- Present their findings in the form of a report and presentation

Impact on Students

1. Increased knowledge of R programming
2. Experience with the entire process of data analysis
3. Formation of their own beliefs and opinions regarding data analysis
4. Learned to think like a data analyst
For More Information

Go to https://github.com/therri-usher/IntroDataAnalysisCourse for course slides, sample code, syllabus, and more!

Thanks

- Gordis Teaching Fellowship
- Public Health Studies at Johns Hopkins University
- Scott Zeger
- Charles Rohde

Contact Info:
Therri Usher
Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
tusher1@jhu.edu