Stat 601 Questionnaire

Name: James Young

1. Where do you come from?

The upper peninsula of Michigan, originally, then North Dakota and now South Dakota.

2. Of the statistical topics you have studied, which do you like the most?

I enjoyed predictive analytics and transferring some of those skills for competitions in Kaggle and the Dream Challenges.

2a- Why?

It’s exciting, like a video game, trying to beat the score of a previous model while being careful not to overfit.

3. Of the courses you have taken, which statistics course did you like the least?

It’s hard to pick one that I haven’t enjoyed, and I’ve learned useful techniques in all of them, but my least favorite was Stat 600 (I still really enjoyed it though).

3a- Why?

It required SAS, which I was not familiar with and didn’t see a huge incentive to learn.

4. What is your past experience with analysis of data?

I first gained interest in analyzing data in undergrad when we looked at opposing football team’s tendencies conditional on factors such as down, distance, previous play, etc. Currently, in grad school I have taken more formal routes at pursuing this interest through stat courses based on making explanatory and predictive models. I also am starting to deal with some larger datasets from my biological “omics” experiments.

5. What do you expect of this course?

To increase my formally accepted background and ability in inferential and predictive thinking and to improve my tool kit for both areas.

6. What do you expect of the instructor?

Consistency and helpfulness in addition to delivering useful content.

7. Why do you think statistics is important (for you)? If you like statistics, why?

I think gaining further statistics experience can improve my thought process for approaching problems, both in my research and day to day life. I enjoy statistics because of the quantitative trait it can bring to otherwise “fluffy” problems. How the data is gathered and interpreted is another story, though. I especially like the competitive and creative side of using statistical learning in online competitions, as mentioned earlier.

8. What is your experience with programming and statistical programming?

I started using R when I started grad school in 2017. It was an on/off affair at first, but I’ve been more consistent this past year and have also picked up some Python skills. I’ve used R formally in STAT 541 and STAT 600 as well as in some of my research.

9. What is your preferred programming language for data analysis?

I am currently more comfortable with R but am increasingly using Python. It seems like there are more helpful online forms for targeted questions with Python. I think I will probably transfer to using Python more in the future.

9a- Why?

Python and R are close in ease of use, but it seems industry jobs that I look at ask for Python more. Both can make reproducible analysis and interactive web apps (Shiny vs. Flask or Dash) but Python has more resources that I value such as free GPU us on Google Colab.

10. Do you prefer statistical software to writing your own programming script for data analysis?

I guess it depends what is meant by statistical software. Is R or Python a statistical software or a platform for statistical programming? Maybe it’s both depending on how much you depend on previously written packages for your analysis. I feel more satisfied programming a task (if I’m able to) rather than pointing and clicking.

10a- Why

I think I have a better understanding of the analysis if I code more of it versus pointing and clicking. I think there is also more flexibility in the way you can interrogate a problem if you can program sufficiently.