

EXAM #1

SSC 442

Professor Ben Bushong

Spring 2020

Exam Date: Take Home Exam, February 20

Due Date: Tuesday, February 25 at 12:39 p.m.

You should take this exam as if you had 120 minutes to complete it. There are three primary questions (equally weighted). You absolutely cannot work longer than 150 minutes on this exam; however, you should not feel obligated to use that much time by any means. On all questions, you must write a clear paragraph or two, recognizing that multiple, unclear, or vague answers will be graded with maximal skepticism. You may not work with any other people, but you may consult any internet resources you like. (Again, internet resources  $\neq$  human beings. Disallowed collaboration is explicitly prohibited.).

You must email your answers to [bbushong@msu.edu](mailto:bbushong@msu.edu) with the subject line **[SSC442: Exam 1]**. Failure to include this subject line will result in a 20% reduction in score.

Be mellow. Don't freak out. You are an intelligent and beautiful person. (Obviously the other people in the class are getting the same message. But in his or her case, I am just saying it to build confidence and security as he or she starts the exam. In your case, I really mean it.)

---

PLEASE DO NOT READ PAST HERE UNTIL YOU ARE READY TO BEGIN.

---

---

## START OF EXAM

---

You are in the final round of interviews at Amazon (the large company that sells stuff. If you've never heard of them... uh, Google them?). They are unsure which business group you will work with, and so they are assessing your ability to apply economic and statistical insights to a wide variety of problems. Three groups will ask you questions; your final score will consist of the equal-weighted sum of your answers.

**Workforce Analytics (Human Resources):** We have a large ( $n \approx 250000$ ) dataset of job applications and whether people were hired. If they were eventually hired, we also track their success in the company and whether they were ultimately fired.

1. What might be a good application of statistical learning in this environment? What is the object we are predicting, and what are the inputs?
2. We need to be able to justify our decision-making. Describe how would we could interpret the output of your answer to #1 when speaking with senior management. If the methodology you chose in #1 above is not easily interpreted, explain why.

**Echo Team (Engineering):** We have nearly infinite data on voice commands. People speak to their Amazon device ("Alexa, play me the Beatles") and then we have the data on outputs. If a user has an error, they can report it, though people seldom do.

1. We're looking to bring new engineers in, but we want to make sure they can do some simple programming. Can you think of an example of a simple visualization that they might do using our data? How would this deliver business value to the unit? **Your answer must include reference to the relevant R package for the visualization.**

**Twitch Data Science (Analytics):** We have a small dataset on users who give bits to streamers. Streamers are people who play video games online and stream their play via the Twitch.tv platform. Users can watch them for free—monetized by adrolls—or can pay for premium, ad-free content. Bits are a small monetary gift that encourages streamers and acts as a pay-for-content system. We are looking to expand the bits program and possibly change the price.

1. What is a simple linear model we could run to look at which characteristics of streamers encourage donations? What are some inputs  $X$  and what might our target variable  $Y$  be?
2. We are thinking we might conduct a small-scale experiment in which we change the price on donations. How large of a sample (in terms of the number of users) do you think we need? How does this depend on your answer to #1?

---

**END OF EXAM.**

---