Midterm Exam - Stat 313

The Week of February 8, 2021

These are the exam questions given to students for their oral midterm exams. Students answered one randomly selected question from each bank (Linear Models & Additional Topics), and were informed of their assigned questions when they entered the "exam room."

The numbered questions are what were given to students the Friday before exam week. The *follow-up questions* are a sample of the questions I asked students following their explanation. These questions are not exhaustive, but provide an outline of the types of questions I asked students.

Linear Models - Question Bank

You will be asked one question from this bank of questions during your oral exam.

1. A linear regression is one type of model in the broad class of "linear models." What are the characteristics of these linear models? What are the requirements for using these types of models?

Possible follow-up questions:

- How is the relationship between the variables described?
- What are the conditions of a linear model?
- How are a difference in means and a linear regression both linear models?

2. Describe to someone who has never taken a Statistics course the difference between simple linear regression and multiple linear regression.

Possible follow-up questions:

- Are there differences in the conditions for these two models?
- How does the interpretation of the slope coefficient change in a multiple linear regression?
- What are the benefits of using a multiple linear regression instead of a simple linear regression?

3. What are the benefits and weaknesses of linear regression?

Possible follow-up questions:

- What are the conditions of linear regression?
- Are the conditions for regression difficult to be "met"?

5. Some textbooks state that regression should not be used if there are "outliers" in the data. Why is this an important condition to consider? What are the implications of removing every "outlier" from a dataset?

Possible follow-up questions:

- What constitutes an "outlier"?
- How would an outlier impact a linear regression?
- How would you determine if a point was an "outlier"?
- How does removing every "outlier" point from a dataset change what can be inferred from the regression?

6. Most people are familiar with the saying that "correlation does not imply causation." However, when some researchers find a strong correlation between their explanatory and response variables ($R > 0.8$), they often interpret the relationship as "larger values in x *lead to* larger values in y." What is an issue with this interpretation? In this scenario, what could researchers conclude?

Possible follow-up questions:

- How do you interpret the phrase "lead to"?
- What type of relationship is implied by this phrasing?
- What is required to make causal statements?
- What does random assignment look like in the context of a linear regression?

Additional Topics - Question Bank

You will be asked one question from this bank of questions during your oral exam.

Statistical Models

1. Describe to someone who has never taken a Statistics course what the principle of parsimony is.

Possible follow-up questions:

- What does it mean for a model to be "simpler" than another model?
- How does one choose the "simplest" model?
- How could you determine if it's worth it to add another explanatory variable to the model?

Reproducibility

2. Much of what we have done thus far in `R` could have been done in Excel. What are the benefits of programming in `R`? What are the costs?

Possible follow-up questions:

- Do the costs of learning to code in `R` outweigh the benefits?
- What have you learned to do in `R` that you hadn't done in Excel?
- If you were to use Excel to analyze your data, how would you get that analysis into a report?
- Are there any issues with copying-and-pasting graphs or analyses from Excel into Word?

Data Visualization

3. We have spent a great deal of time producing and critiquing data visualizations, and have yet to discuss statistical testing (e.g. hypothesis tests). What information can a data visualization provide you that a data summary or a statistical test cannot?

What would you say to someone that said someone could manipulate a visualization to say whatever they want, but a mean doesn't lie?

Possible follow-up questions:

- What type of information does a statistical test (like a hypothesis test) tell you?
- What type of information does a data visualization tell you?
- If you weren't sure about the results of a statistical test, how would you determine if the results were sound?

Study Design

4. Random samples and experiments are often thought of as the "gold standard" of study design. Why are these aspects of a study so highly valued? What difficulties may researchers encounter when attempting to utilize these designs in their study?

Possible follow-up questions:

- What does random sampling allow for researchers to say?
- What does random assignment allow for researchers to say?
- How would it be difficult for a researcher to use random sampling?
- How would it be difficult for a researcher to use random assignment?
- Can you always use random assignment? Why or why not?

5. How is statistics used in the generation of scientific knowledge? How is dichotomous thinking used in statistics? How is objectivity related to each of these areas?

Possible follow-up questions:

- How is statistics used in your discipline?
- What is the connection between dichotomous thinking and the use of statistics?
- Where is dichotomous thinking used in statistics?
- What is the relationship between dichotomous thinking and "objectivity"?