Midterm Exam - Stat 313

October 25 & 26, 2021

Linear Models - Question Bank

You will be asked one randomly selected 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?

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

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

4. 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?

5. 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?

Additional Topics - Question Bank

You will be asked one randomly selected 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 (or Occam's razor) is.

Reproducibility

2. Much of what we have done thus far using `R` and R Markdown documents could have been done using Excel and Word. What are the benefits of using `R` and R Markdown documents? What are the costs?

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 and p-values). What information can a data visualization provide you that a data summary or a statistical test cannot?

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?

Philosophy

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?