Midterm Exam - Stat 313

October 23, 2020

Linear Models - Question Bank

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

1. Describe to someone who has never taken a Statistics course what a linear model is.
2. Describe to someone who has never taken a Statistics course the difference between a simple linear regression and a multiple linear regression.
3. What are the benefits and weaknesses of linear regression?
4. Many disciplines employ regression, but do not check the model conditions. Why should we care about the model conditions? What are the implications of not checking these conditions?
5. Some textbooks list "lack of influential points" as a model condition for regression. Why is this a potentially important condition to consider? What issues might arise from removing all "influential" observations from a dataset?
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?

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.
2. Describe to someone who has never taken a Statistics course what Simpson's paradox is.
3. How would you determine if your statistical model should be used for predicting what could be expected to happen for new data?

Reproducibility

4. 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?

Data Visualization

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

6. 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?

7. What is the purpose of blocking in a study? If an experiment fails to block on a variable that influences the response, what are the implications for the inference of the study?