Teaching Pedagogy for an Intro Stats Class
Project Case Study #1

Statistics 469: Analysis of Correlated Data

Problem Background

As you are no doubt aware from personal experience, effective college instruction should include student learning activities that lead to content mastery. In order to promote higher learning, class activities that are not effective in helping students learn should be discarded (or altered) and more time placed on those activities most helpful in promoting desired learning outcomes.

In order to evaluate the effectiveness of learning activities, a statistics department (I wonder which one) gathered data from past semesters on their introductory statistics class. The dataset, *ClassAssessment.txt*, consists of the performance of the students in the most recent 5 academic years. Because of the different demographics of students taking the class during the summer, the summer semesters were not included in the assessment. Specifically, the dataset consists of the following variables:

Variable Name	Description
Semester	The semester when the course was given (note there were multiple sections per semester)
NStudents	The number of students who completed the course
Exam1	The average score (in percent) on Exam 1
Exam2	The average score (in percent) on Exam 2
Exam3	The average score (in percent) on Exam 3
HW	The average score (in percent) on the homework
Quiz	The average score (in percent) on class quizzes
Final	The average score (in percent) on the final exam

Primary Research Questions

The goal of gathering the data was to assess if the semester learning activities (e.g. homework, quizzes and exams) was associated with successful learning of the course material (as measured by the final exam score). Specifically, the department researchers want to answer the following questions:

1. What activities, if any, are associated with improved learning? Are there activities that are not associated with improved learning?
2. Of those activities that are associated with improved learning, which have the strongest effect on learning? How large are these effects?
3. How well do the class activities explain student learning?
4. Historically, were there any semesters that had either better or worse student learning than average?

Research Paper Outline

Your assignment for this exam is to write a report to the department on your analysis of *ClassAssessment.txt* and your subsequent answers to their questions. For your report, use the outline below. Your report should

be technical (including equations as necessary) and grammatically correct. Your report should be divided into the sections outlined below with no additional sectioning (meaning no Section 1(b) for example). Each section should include rhetoric on each point mentioned in the outline.

0. Section 0: Executive Summary
   - In 1 paragraph, summarize the main results of your analysis so that readers can understand your conclusions without getting into the details of the statistical modeling.
1. Section 1: Introduction and Problem Background (12pts)
   - In your own words, describe the background of the problem and the goals of the study. Give a short description of the data along with a few graphics to summarize the data.
   - Discuss any potential issues associated with the data that you should make sure to account for in your analysis.
   - For any issues identified above, discuss the consequences of ignoring these issues in an analysis along with any benefits gained by accounting for these issues.
   - Provide a 1 paragraph "birds eye description" of the method you will use to analyze your data (mathematical details will be given later). If appropriate, describe why your chosen method accounts for the challenges you mentioned in 1 (b) above.
2. Section 2: Statistical Model (10pts)
   - In mathematical detail, write out the statistical model that you will use to achieve the goals of the analysis. Be sure to define any parameters and notation that you use in the context of the problem. Provide intuition on what these parameters mean.
   - Explain (you will justify these in the next section) any assumptions that your model uses.
3. Section 3: Model Validation (10pts)
   - Justify any assumptions you made in your model using graphics, summary statistics or hypothesis tests.
   - Describe how well your model fits the data. Justify your answer using appropriate summary statistics or graphics.
   - If appropriate for the problem, describe how well your model does at predicting. Describe how you know and support your conclusions with numerical summaries or figures.
4. Section 4: Analysis Results (13pts)
   - Using results from your model, do you think there are any class activities associated with improved learning? If so, which ones?
   - Of those activities you identified as having a positive association on learning, how much of an impact does each activity have?
   - How well do the class activities currently explain learning?
   - Did you identify any semesters that were better or worse in terms of student learning? If so, which and by how much?
5. Section 5: Conclusions (2pts)
   - Briefly summarize the main findings of your analysis in 1 paragraph and without using statistical jargon.
   - Identify 1-2 "next steps'' that the department should consider that would help them create a better learning environment.
6. The report is well written, formatted appropriately, grammatically correct and proof-read (3pts)

Project Rules

The following rules apply to this exam:

1. The exam is open everything including ONE partner (if you choose to do this alone then it is closed to everyone else). This means that you are free to use old homework, class notes, old code, Google, etc. but cannot talk anyone other than your partner about the exam until after it is over.
2. You can ask the TAs or myself questions about the exam but we will likely refuse to answer.
3. The exception to rule #2 is in regards to debugging code.
4. Please attach your code to the end of the exam as an appendix.