Stat 696 Old DAR Evaluation

Andrew Bates

9/26/2018

Exam 1

Rank: 4

This paper was poorly organized and lacked clarity. With the entire paper consisting of two sections, the analysis and conclusion under the 'Data Preparation' section, and short text which doesn't explain what they did, it's not clear to me what analysis was performed. The claim that a better test score in high school is a "free ticket to any colleges" seems to come out of nowhere. If they were analyzing college acceptance rates this may have some grounding but they state they are analyzing high school test scores.

There are also a number of issues at a more detailed level. The variable names were all abbreviations that one would need a data dictionary to interpret. The tables were unformatted and look like code output. There were also too many digits in the tables to be readily interpretable. The plot is labeled "Avegrade" which is too brief, especially since they did a poor job explaining what 'Avegrade' is. At best, this paper feels like an outline/first draft that was thrown together in a few minutes.

Exam 2

Rank: 1

The first word in paper is a grammatical error. The summary is good in that it provides background, methods, and results. However, they should have expanded a bit more. For example, they say teacher salary and enrollment influence drop-out rate but are these the only variables? Also, I'm not sure how controllable the covariates used are by a school/school board. The mean drop-out rate is 2. Apparently, this means 2%? They should have made this clear instead of leaving it up to the reader to infer.

In the first paragraph of the Statical Analysis section, I like how they justify their use of linear regression via its interpretability. The second paragraph provides an explanation of backwards elimination, which is good, however, I thought they went into too much detail. This paragraph is also a bit jumbled. It should be split into two paragraphs: one explaining the method, and another explaining the results. Actually, this first explanation of the model should be removed because this is not their final model. As readers, we don't really care about all their intermediate models.

Exam 5

Rank: 2

Right off the bat, the title contains a grammatical error which is not a good sign. I like how this paper has an executive summary to provide a short summary of the problem and results. However, they say they are predicting several different responses, but the paper doesn't explain these models beyond R^2. Also, it's not clear to me what the final model was. This is in part because they explain several models, one with a transformation, then state in the conclusion that a transformation was not used.

They say variable selection was used but don't explain what method was used. Also, they list significant variables but list different variables after model selection. Which ones were used? The conclusion is about 1.5 pages while the text for the rest of the paper is about 2 pages. The conclusion is more like a summary of the analysis than a conclusion.

Exam 11

Rank: 3

The executive summary is a nice addition but they don't provide a full summary. Specifically, the last sentence states they analyzed several variables but they never state the results of this analysis. Also, this last sentence should be incorportated into the first, which explains what they did in this report. In the introduction, they state that proportional hypothesis testing" is used for one set of percentages and "hypothesis testing for equal means" for another set. I'm not sure why they used different tests. They also say that logistic regression is used but this conflicts with the analysis section which states the response is a real number.

One part of the analysis I don't understand is why they performed KS tests on all variables. They should have explained the reason for this. In fourth paragraph of the analyis section, they state that one "data point was altered"! I don't agree with this decision at all. On page 5, they tested for a difference in proportion of students who went to 4 year vs. 2 year colleges. They say that this was done using the previous method. But the previous method was a regression and I don't know how this tests for a difference in proportions.