STAT 1361   Name _____
Statistical Learning and Data Science
University of Pittsburgh

Written Report Rubric

1. Introduction _____ / 15

   This should include an overview of the problem of interest and details of the specific dataset as well as a clear description of the problems of interest. Specific things to consider:

   - Who might be interested in the results of your analyses and what kinds of questions might they find particularly relevant?
   - How big is the data and how is it structured? Are there different ways you could think about the structure?
   - Is there any missing data and if so, how much?

2. Methods Overview _____ / 15

   This should contain a summary of the methods explored and the various approaches that were considered. Specific things to consider:

   - How might the size and structure of the data determine the kinds of methods that you will concentrate on?
   - Is the response continuous or categorical? Could it be treated as either?
   - Would different methods be more appropriate for different questions you might want to answer?

3. Method Details _____ / 30

   This is where you should provide details on the methods that you personally were involved in fitting and how those results influenced other models that were attempted. You may include plots here, but they should not overwhelm the document. If you need (or would like) to provide many plots and/or code, these can be put into an appendix. Specific things to consider:

   - Are there any specific challenges with employing these method(s) on the type of data you have?
   - How did you select final models and/or choose tuning parameters?
   - What kinds of plots are most relevant for the method(s) you're employing?
   - How do these methods tie in with the other method(s) your group is using?
   - What variables are routinely determined to be important/unimportant?
   - How did you address missing data issues?
   - Do you trust the inference methods available for the method(s) you are using? Did you try alternative measures based on deleting/permuting variables?

- How much do you trust the results you observe and how well do the method(s)/model(s) seem to perform?

4. Summary of Results _____ / 15

This should provide an overview of all of the results obtained by everyone in your group. Comment on overall trends, contradictions between models, etc. You can include a table here if it helps summarize the findings. Specific things to consider:

- Are the models largely consistent or inconsistent (i.e. are the same variables generally selected as important etc.)?
- Are there particular methods or classes of methods that seem to be doing far better than the others?

5. Conclusions and Takeaway _____ / 25

Based on the results described in the previous section, describe what you feel can safely be concluded. If there are further tests/models that you think would be relevant to pursue given the overall results, you should do so here (again, continuing with details in an appendix if necessary). Specific things to consider:

- Which models do you trust the most and why?
- What might explain the agreement/disagreement between methods/models that you reported in the previous section? Are there things you could do to try and test this?
- Which models would you employ in practice? What if you only cared about inference? What if you only cared about prediction?
- What do you feel can safely be concluded from your analyses? Why? What are some practical takeaways? Are any of these surprising?
- How would you explain the results to someone interested in your findings that doesn't have a strong statistics background?

Comments: