Writing a Statistical Report for STAT 313

Introduction

- Give a brief background of the research problem and how the data were collected.

- Clearly outline the question(s) of interest you will address with your statistical analysis. The more specific you define the question of interest here, the easier the rest of the analysis is.
    - Your question should start with something like "What is the relationship…" or "Is there evidence..." and should be as specific as possible.
    - Your *Summary of Statistical Findings* should directly answer the question(s) you pose here!

- *Example for a multiple linear regression: What is the relationship between a credit card balance and income? Does this relationship differ based on credit rating (poor, fair, good, great)?*

Statistical Procedures Used

- This section should lay out the steps, decisions, and logic leading to the final model used to answer the question of interest. Think about telling a story about the question of interest, starting with a description of the variables and ending with a description of the statistical model(s) you will use.

- Describe the response and explanatory variables.

- Use plots to visualize the **raw** data -- describe what you see.
    - Include all plots inline, not in an Appendix!

- Describe the appropriate statistical methods used to answer the research question(s) (e.g. simple linear regression, multiple linear regression, two-sample t-test, etc.).
    - Be specific about which methods are answering which questions of interest (if there are multiple questions).

Summary of Statistical Findings

- In this section you will write up your findings for each research question of interest.

- For questions where you **are not** conducting a statistical test (e.g. hypothesis test), you can summarize your conclusions in a few sentences:
    - Arrive on a conclusion based on the relationship(s) seen in the visualizations **and** the size of the estimated coefficients.

    - *Example for a simple linear regression: Based on the strong positive relationship seen in the scatterplot and the large estimated coefficient of 3.28, we believe there is a substantial relationship between credit card balance and income. For every $10,000 increase in income, we expect the mean credit card balance to increase by approximately $32,800, which is quite substantial.*

- For questions where you are conducting a statistical test, you can usually summarize your conclusions in a few sentences:
    - Arrive on a conclusion based on the strength of evidence provided by the p-value.

- **Do not** use a threshold cutoff!
- **Do not** use the word "significant"!

- Write a conclusion communicating the results of the statistical test to the interested party. Make certain that you include the value of the test statistic **and** the distribution it follows under the null hypothesis.

- *Example for a simple linear regression: There is strong evidence that there is a linear relationship between credit card debt and income (p-value = 0.0263, t-stat = 2.39 on 1623 df). Furthermore, there is strong evidence that this relationship differs based on credit rating (p-value = 0.027, F-stat = 3.067 on an F(3, 1616) distribution).*

- Notice how similar the wording is to the question of interest!

- Make sure you address these questions for **every** statistical method used in the analysis.
  - If you are doing both a simple linear regression and a multiple linear regression, you need to discuss the findings of **both** models!
  - If you are running a multiple comparison test after a one-way ANOVA, make sure to explain why you are doing the follow-up test and what information it provides!

Scope of Inference

- Write a brief statement on the scope of inference for your analysis. Specifically, answer these two questions and comment on their implications:
  - Were the units randomly selected from some larger population? (i.e. What, if any, larger population can you infer the results to?)
  - Was the explanatory variable(s) randomly assigned to groups? (i.e. Are cause-and-effect statements justified?)

- *Example: The 1624 individuals were not randomly selected from any specific population, so we cannot infer our findings to anyone outside of these individuals. Additionally, credit score and income were not randomly assigned, so we cannot conclude that either variable causes a change in the credit card balance. Instead, we can infer that there is an association between credit card balance, credit score, and income, for the individuals included in these data.*

- Make sure you write the Scope of Inference specific to the language of the problem (not just generic statements).
  - For example, **do not** say, "There was no randomization so cause and effect statements cannot be made."

- Make sure you say what was (or was not) randomly assigned to what.
- Make sure you say what was (or was not) randomly selected from what.
- Try to keep your statement to 3 sentences!

Other

- Style:
  - Report must be written in complete sentences in paragraph format.
  - Use the above four headings in your report so that it is clearly organized.

- Length:
  - The body of the report should be no more than 2 pages!

- Appendix:
  - You should include all of the R code used in the R Markdown document associated with your report.
  - I should be able to download your report and click "knit" and get the **same** report you submitted.

- Part of the grade for each project is based on writing (organization, grammar, spelling, appropriate use of statistical terms, etc.)