MET CS 555 Term Project

20 points

Select a small data set from the available public data sets (you can find a list of public data sets here http://www.teymourian.de/public-data-sets-for-data-analytic-projects/).

Describe a research scenario and specify a research question based on data analytic methods that we learned in class. For example, methods such as *one and two sample mean tests, correlation tests, simple and multiple linear regression, ANOVA and ANCOVA, one and two-sample tests for proportions, and logistic regression* are fair game. Perform your analysis, and then report your results and conclusions.

Clean up your data and randomly sample 1000 observations from it if your data set is large. Sample Projects: https://www.teymourian.de/for-students/ .

1. Describe your research scenario and question(s).

Briefly describe your research scenario. Similar to our class examples, you should first describe the overall scenario and then specify a specific research question (or questions) based on it.

2. Describe the data set.

Briefly describe the data set. Describe each variable of the data set that you plan to use in your analysis. Describe any data cleaning you have performed. If possible, provide a link to the main data set source.

3. Describe the statistical methods you plan to use.

Briefly describe the statistical methods you will be utilizing to investigate your research question(s).

4. Report your results.

Write up the results of your analysis. You should present tables and figures when relevant, and you should have a short write-up describing your results.

5. State your conclusions and discuss any limitations.

State the conclusion so that a none-statistician can understand. Discuss any potential limitations of your analysis. For example, are you suspicious that the assumptions of your test may not hold? Do you feel the analysis may have limitations for any other reasons?

Solution Submission

1. **Upload a write-up document.**
2. **Upload your data set. This is the data set after cleaning (a small CSV file).**
3. **Upload your R script.**
4. **Upload a 2-minute presentation (up-to-12-slides, prepared with PowerPoint or the equivalents, and saved as pdf file)**

Grading will be based on:

1. **Development and description of a research scenario and question(s).**
2. **Data preparation and data cleaning (when relevant).**
3. **Correctness of statistical analysis methods chosen to answer the research question.**
4. **Correctness of R code.**
5. **Quality of figures, tables, and write-up.**
6. **Correct and thoughtful discussion of conclusions and limitations.**
7. **Verbal and written communication and presentation skills of your project. [Extra Credit: 4 pts]**

Submission Schedule

1. **One-page project proposal describing your research scenario and question(s), Due March 23, 2023, 5pm ET, on Blackboard.**
2. **Up-to-8-page write-up document along with your data set, your R script and your 2-minute presentation, Due April 20, 2023, 5pm ET, on Blackboard.**
3. **2-minute in-classroom presentation, Due April 27, 2023, 6pm ET, in the classroom.**