Statistical Methods for Discrete Response, Time Series, and Panel Data (W271): Lab 1

Professor Jeffrey Yau

Instructions:

- **Due Date: 2/4/2018 (Sunday)**

- **Page limit of the pdf report: 15**

- Submission:

    - Submit your own assignment via ISVC
    - Submit 2 files:
        1. A pdf file including the summary, the details of your analysis, and all the R codes used to produce the analysis. Please do not suppress the codes in your pdf file.
        2. R markdown file used to produce the pdf file
    - Each group only needs to submit one set of files
    - Use the following file naming convensation; fail to do so will receive 10% reduction in the grade:
        * FirstNameLastName_LabNumber.fileExtension
        * For example, if you have two students named John Smith and Jane Doe, you should name your file the following
            · JohnSmith_JaneDoe_Lab1.Rmd
            · JohnSmith_JaneDoe_Lab1.pdf
    - Although it sounds obvious, please write the name of each members of your group on page 1 of your pdf and Rmd reports.
    - This lab can be completed in a group of up to 3 people. Each group only needs to make one submission. Although you can work by yourself, I encourage you to work in a group.

- Other general guidelines:

    - Try to use only techniques and R libraries that are covered in this course.

    - If you use R libraries and/or functions to conduct hypothesis tests not covered in this course, you will have to explain why the functions you use are appropriate for the hypothesis you are asked to test. Lacking explanations will result in a score of zero for the corresponding question.

    - Thoroughly analyze the given dataset. Detect any anomalies, including missing values, potential of top and/or bottom code, etc, in each of the variables.

    - Your report needs to include a comprehensive Exploratory Data Analysis (EDA) analysis, which includes both graphical and tabular analysis, as taught in this course. Output-dump (that is, graphs and tables that don't come with explanations) will result in a very low, if not zero, score.

    - Your analysis needs to be accompanied by detailed narrative. Remember, make sure your that when your audience (in this case, the professors and your classmates) can easily understand your your main conclusion and follow your the logic of your analysis. Note that just printing a bunch of graphs and model results, which we call "output dump", will likely receive a very low score.

    - Your rationale of any decisions made in your modeling needs to be explained and supported with empirical evidence. Remember to use the insights generated from your EDA step to guide your modeling step, as we discussed in live sessions.

    - All the steps to arrive at your final model need to be shown and explained very clearly.

- Other requirements:

- Groups will be asked to present their lab in the live session following the submission date.

- Students are expected to act with regards to UC Berkeley Academic Integrity.

Investigation of the 1989 Space Shuttel Challenger Accident

1. Read the Dala et al (1989) paper (attached in this zip file).

2. Conduct a thorough analysis and EDA of the given dataset "challenger.csv", as we did in live session 2 and 3. Pay attention to the instructions given above.

3. Answer question 4 and 5 on Chapter 2 (page 129 and 130) of Bilder and Loughin's *"Analysis of Categorical Data with R"*

4. In addition to the questions in Question 4 and 5, answer the following questions:

   a. Interpret the main result of your final model in terms of both odds and probability of failure

   b. Plot the main effect of your final model with the y-axis being probability of failure and x-axis being *temperature*.