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

Professor Jeffrey Yau

Instructions:

- **Due Date: 6/3/2018 (Sunday)**
- **Page limit of the pdf report: 15, which does not include the table of content page**
- Do not play around with the margin, linespace, and font size;
- Use the one I specify below:
  - fontsize=11pt
  - margin=1in
  - line_spacing=single
- Submission:
  - Each group only makes to make one submission to ISVC; have one of your team members made the submission
  - 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:
    - FirstNameLastName1_FirstNameLastName2_FirstNameLastName3_LabNumber.fileExtension
    - For example, if you have three students in the group for Lab 1, and their names are Paul Laskowski, Drew Paulin, and Jeffrey Yau, then you should name your file the following
      - PaulLaskowski_DrewPaulin_JeffreyYau_Lab1.Rmd
      - PaulLaskowski_DrewPaulin_JeffreyYau_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 4 people. Each group only needs to make one submission. I strongly encourage students to work in groups for the lab.
- Other general guidelines:
  - For statistical methods that we cover in this course, use only the R libraries and functions that are covered in this course. If you use libraries and functions that we have not covered but are for the statistical methods we learn in this course, you have to provide (1) explanation of why such libraries and functions are used instead and (2) reference to the library documentation. Lacking the explanation and reference to the documentation will result in a score of zero for the corresponding question.

  - Your report needs to include
    * A thorough analysis of the given dataset, which includ examiniation of anomalies, missing values, potential of top and/or bottom code, etc, in each of the variables.
    * An introduction section that summarize the question(s) being asked, the methodology employed (including the final model specification), and a highlight of the results.
    * 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. Since the report has a page-limit, you will have to selectively include the visuals that are most relevant for the analysis and concise explanation of the visuals. Please do not ramble. Please remember that your report will have to "walk me through" your analysis.
    * A modeling section that include a detailed narrative. Make sure that your audience (in this case, the professors and your classmates) can easily follow the logic of your analysis that leads to your final model.
      · The rationale ofdecisions made in your modeling, supported by sufficient empirical evidence. Use the insights generated from your EDA step to guide your modeling step, as we discussed in live sessions.
      · All the steps used to arrive at your final model; these steps must be clearly shown and explained.
    * A conclusion that summarize the final result with respect to the question(s) being asked and key takeaways from the analysis.
- Other requirements:
  - Students are expected to act with regards to UC Berkeley Academic Integrity.

Investigation of the 1989 Space Shuttel Challenger Accident

1. Carefullly read the Dala et al (1989) paper (attached in this zip file).
2. Answer question 4 and 5 on Chapter 2 (page 129 and 130) of Bilder and Loughin's *"Analysis of Categorical Data with R"*
3. 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. With the same set of explanatory variables in your final model, estimate a linear regression model. Explain the model results; conduct model diagnostic; and assess the validity of the model assumptions. Would you use the linear regression model or binary logistic regression in this case. Why? Or, why not?