Statistical Methods for Discrete Response, Time Series, and Panel Data (W271): Lab 2

W271 Instructional Team

June 22, 2017

Instructions:

- **Due Date: 7/2/2017 (11:59 p.m. PST, Sunday)**

- 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:

    * SectionNumber_hw01_FirstNameLastNameFirstInitial.fileExtension
    * For example, if you are in Section 1 and have two students named John Smith and Jane Doe, you should name your file the following
      · Section1_hw01_JohnS_JaneD.Rmd
      · Section1_hw01_JohnS_JaneD.pdf

  - Although it sounds obvious, please write the name of each members of your group on page 1 of your report.

  - 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, we encourage you to work in a group.

  - When working in a group, do not use the "division-of-labor" approach to complete the lab. That is, do not divide the lab by having Student 1 completed questions 1 - 3, Student 2 completed questions 4 - 6, etc. Asking your teammate to do the questions for you takes away your own opportunity to learn.

- Other general guidelines:

  - Please read the instructions carefully.

  - Please read the questions carefully.

  - 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 function you use is appropriate for the hypothesis you are asked to test

  - 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.

– 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.

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

The Lab

The purpose of this lab is to practice using Poission Regression models, conduct residual diagnostics, and examine goodness of fit.

- **Do question 23, 24, 25 in C Chapter 5 (page 351 and 352) of Bilder and Loughin's *Analysis of Categorical Data with R*.**

- Even though it is not included in these questions, conduct a comprehensive Exploratory Data Analysis (EDA) analysis, which includes both graphical and tabular analysis, as taught in this course.

- In your report, type the questions (before you answer each of the questions) in your report.