Week #9 Assignment

Poisson Regression

Due: Sunday, 11:59 PM

- **Data Descriptions**

The daily total of bike counts was conducted monthly on the Brooklyn Bridge, Manhattan Bridge, Williamsburg Bridge, and Queensboro Bridge. To keep count of cyclists entering and leaving Queens, Manhattan, and Brooklyn via the East River Bridges. The Traffic Information Management System (TIMS) collects the count data. Each record represents the total number of cyclists per 24 hours at Brooklyn Bridge, Manhattan Bridge, Williamsburg Bridge, and Queensboro Bridge.

- **Data Formats and Loading**

To save you time in finding a data set for Poisson regression, I created several subsets that contain the relevant information you need for this week's assignment. The data was saved in Excel format with multiple tabs. Please find the tab with your last name and then copy-and-paste your data file to a new Excel sheet and save it as a CSV format file or simply copy-and-paste to Notepad to create a TXT format file so you can read the file to R. You can also read the Excel file directly to R using appropriate R functions in relevant R libraries.

Here is the link to the Excel data set:NYC cyclist data set (AssignDataSet.xlsx).

- **Assignment Instructions**

Your analysis and report should be similar to section 3 of my class note. PLEASE TELL STORIES BEHIND ALL R OUTPUTS (tables and figures) YOU GENERATED IN THE ANALYSIS. You can earn half of the credit if you only generate relevant outputs correctly but with no good storytelling. *The model diagnostic is not required in this assignment but will be required in the next assignment.*

The following components must be included in your analysis report.

- Your description of the data and the variables
  - data collection
  - variable names and definitions. Keep in mind that the variable **Date** is the observation ID.
- What is the research/practical question?
  - Explicit definition of the questions
  - what is the response variable that captures the information to address the research/practical questions
- What statistical model will be used in the analysis
  - Assumptions and conditions - give a brief description of the models
  - Build a Poison regression model on the counts only.
    * Use the p-values to perform variable selection - keep only significant variables in the final model

* Interpret the regression coefficients of the Poison regression model
* Explain/describe the steps of your analysis, motivation, and findings.
- Build a Poison regression on the proportions (rates) of cyclists entering and leaving the bridge in your data.
  * Use the p-values to select the significant variables
  * Interpret the regression coefficients of the Poison rate model
  * Explain/describe the steps of your analysis, motivation, and findings.
- Summarize the findings of the above two models.