**Data Analysis And Regression**

**Assignment-1** | **Total points: 10 pts for DSC 323; 15 pts for DSC 423**

Note:

* All assignments should be submitted in a **single MS WORD format**, no PDFs or any other file types will be accepted. If you submit any other file type, it will not be graded.
* No extensions will be given unless for a documented reason specified in the syllabus, no late assignments past the due date even a couple of minutes late will be accepted as you have an extra day (7-days) to submit your assignments.
* Submitting work that is not yours is grounds for an automatic ‘F’ for the entire course – this includes taking content and ideas from others or consulting others to complete your deliverables other than your instructor.
* SAS software and virtual server stalls, gets slow and crashes; so start early and keep multiple backups in multiple places/mediums. Late submission or inability to do the assignment due to server and/or software issues will not be accepted. Any issues relating with SAS, contact IS using the phone number provided in the syllabus, I won’t be able to help you with DePaul software related issues.

**PROBLEM 1 [10 pts] – To be answered by everyone**

The file election.txt attached to this assignment provides data acquired from census records selected counties in the U.S. who voted in the elections. The data show

County – Name of the county

PctVoted – Percentage of people voted

MedianAge – Median age of the voters in that county

MeanIncome – Mean income in U.S. Dollars in that county

PctUnemployment – Percentage of people unemployed in that county

PopulationDensity – Population density (Population divided by square miles) in that county

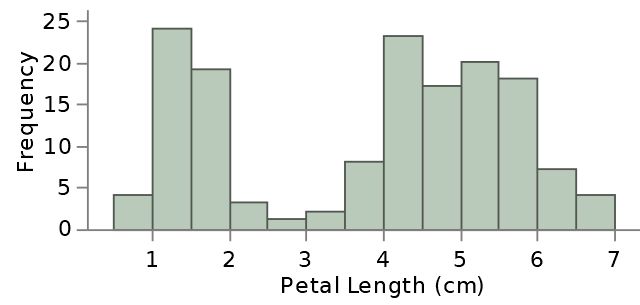
Gender – Dominant gender of the people voted in that county

***Use SAS to compute the analysis below. All the functions are in either the code for the Lab Session-1 we did in class (see code that was posted on D2L). This is the first assignment, and for many of you it may be the first time you use SAS outside of the first lab session. So if you run into an error, post a message on the discussion board or contact me. Make sure to include your code in the message.***

In this exercise you are asked to get the data into a SAS dataset and perform basic exploratory analysis of the data to analyze the characteristics of people voted.

1. Open the dataset and examine the data. Answer the following:
   1. How many Observations are there?
   2. How many fields are there?
   3. Which fields are numerical?
   4. Which fields are text?
2. Write the SAS code to create the SAS dataset using either IMPORT or INFILE statement. If you are using INFILE statement, pay attention to the text fields while writing your code.
3. Run a PROC PRINT to print your dataset in SAS. Do a print screen, copy and paste the first 5 observations of the output.
4. What is the 5-point summary numbers for percentage of people voted, median age and percent unemployment? The 5-point summary numbers are min, max, median or 50% percentile, Q1 and Q3. Include the output. Discuss your findings based on the values you see.
5. Create a histogram to analyze the percent of people voted. Include the histogram output. Using the histogram and the 5-point summary from the previous question, analyze the histogram. Discuss your findings using the 5-point summary, normality (normal or skewed), modality, outliers, etc.?
6. Create a boxplot to analyze percentage of people voted by gender. Include the output. What can you say about the gender and voting patterns? Discuss your findings using the boxplot.
7. What is the gender breakdown in this dataset? (Hint: use PROC FREQ). Include the output. Which is the predominant gender in this dataset? Explain how you came to this conclusion.
8. Copy and paste your FULL SAS code into the word document along with your answers.

**PROBLEM 2 [5 pts] – For Graduate Students ONLY**



Based on the above image answer the following:

1. What type of modality does the image show?
2. What is the minimum petal length?
3. What is the maximum petal length?
4. What size of petal length is higher compared to all other sizes?
5. What is the range of the petal length?