CS130 Assignment #3
Date assigned: Monday, March 29, 2010
Date due: Monday, April 5, 2010
Points: 50

Create a Word document **WordAnswers3PUNuetID.docx** using your PUNetID that will contain answers to each of the following questions in order when applicable. The Word document and any other files created are to be placed in a folder called **PUNetID_Solution3** once again using your PUNetID.

Problem #1

The website **82games.com** provides a wealth of data and analysis for NBA games. The page **http://www.82games.com/0910/0910POR2.HTM** lists data on the 2009-2010 Portland Trail Blazers that describes how each 5 player group performs. The columns are explained at the bottom of the page.

Import the two tables on the above page into Excel. Use the **Data | From Web** interface to do this. Once the data is imported move the columns **eFG** to **T/O** from the bottom table to the right side of the top table, to make one large table that has columns (A-P) labeled **#** on the left and **T/O** on the right. Save this spreadsheet as the file **PUNetID_82Games_PT**. Import this spreadsheet into SPSS and use SPSS to perform the required analysis to to answer the following questions. Answer the following questions in the Word document listed above.

► Note that SPSS cannot handle values of the form: 37%. In the Excel spreadsheet, before you import the data into SPSS, you need to change the formatting of the cells that contain a % to **Number with 2 places after the decimal point**. Your Excel spreadsheet should then display 0.37 rather than 37%.

Clean up the data in SPSS by changing the type and measure as needed and renaming variables as needed. Save this file as **PUNetID_82Games_PT.sav.**

1) List each variable in the dataset. Further, list the <u>type</u> and <u>measure</u> that each variable should be and briefly explain why. Answer this question under a heading of **Problem #1 - Question 1**.

2) What is the mean, median, mode, and standard deviation for each of the variables: (a) **+/-** and (b) **eFG**. Paste the table(s) from SPSS hat answer this question under a heading of **Problem #1 - Question 2**.

3) Construct a bar chart that shows the Win % of each Unit. Paste in the answer to this question under a heading of **Problem #1 - Question 3a**.

4) Using SPSS, perform the correct linear regression on eFG and Win%. Make sure you properly identify the Dependent and Independent variable. Paste in <u>only</u> table(s) containing the R^2 value and the equation coefficients. table. Answer this question under a heading of **Problem #1 - Question 4**.

5) In question 4, which variable is independent? Why? Give a detailed explanation for full credit. Answer this question under a heading of **Problem #1 - Question 5**.

6) In question 4, which variable is dependent? Why? Give a detailed explanation for full credit. Answer this question under a heading of **Problem #1 - Question 6**.

Note1: Make sure all of your graphs are properly and accurately labeled.

Note2: The Word document answers are to be in the exact order as specified above.

Problem #2

Many government agencies post raw data online in accessible (Excel, text, or XML) formats. The federal government maintains the website **data.gov** which contains federal data and links to web pages hosting data collected by state governments. The state of California provides a good deal of data including "California Economic Indicators ". You can download this data from the following website:
http://www.dof.ca.gov/HTML/FS_DATA/indicatr/ei_home.htm

Download **Excel Table 4** of the California Economic Indicator data for **Sep/Oct 2009.** Save this file as **CEI _PUNet.xlsx.** Import the table in worksheet "CEI3 Leading Indicators" into SPSS. You will need to determine which cells in the worksheet to import into SPSS. To more easily import the data into SPSS you might need to alter the spreadsheet.

Once the data is imported, check the type and measure. Make any changes necessary. Save this SPSS file as **CEI_PUNetID.sav.**   Be sure to add **values** correctly for the months.

Run a linear, quadratic, and cubic regression to determine how closely **Overtime Hours** predicts **Unemployment Insurance Claims**.

Copy the resulting scatter plot to the Word document under the heading **Problem#2-Question1**.
Answer the following questions under the heading **Problem#2-Question2:**
>   What is the dependent variable? Why? Give a detailed explanation for full credit.
>   What is the independent variable? Why? Give a detailed explanation for full credit.

Answer the following questions under the heading **Problem#2-Question3:** Clearly label each R^2 value and equation. Do not paste in the output tables from SPSS.
>   What is the R^2 value for each regression?
>   What is the equation of each of the the regression lines?

Answer the following questions under the heading **Problem#2-Question4:**
>   Which regression best represents the data?
>   How well does **Overtime Hours** predicts **Unemployment Insurance Claims?** Backup your claim with
data.

We want to know which month has historically had the highest number of **Unemployment Insurance Claims**. Build an appropriate chart to answer this question. Copy this chart into the Word document under the heading **Problem#2-Question5.**
>   What is the dependent variable? Why? Give a detailed explanation for full credit.
>   What statistic did you graph on the Y-Axis and why?
>   What is the independent variable? Why? Give a detailed explanation for full credit.

Open the **CEI_PUNetID** Excel file again and run the same regressions as you did in SPSS. Copy the scatter plot to the Word document under the heading **Problem#2-Question6**. Be sure the scatter plot contains the R^2 values and the equations.

Submitting your work:

To submit your work, copy your <u>single folder</u> **PUNetIDSolution3** (for me that's will4614Solution3) containing (**WordAnswers3PUNuetID.docx**, **PUNetID_82Games_PT.xlsx, PUNetID_82Games_PU.sav**, **CEI_PUNetID.xlsx,** and **CEI_PUNetID.sav**) into the **CS130 Drop** folder on Turing. You must submit your work by the time specified above for your assignment to be considered on time. Reread the syllabus for the late policy. The official time can be found http://time.gov/timezone.cgi?Pacific/d/-8/java.

Grading:

Grading will be based on:
1) Correctness of your results
2) Completeness of your results
3) Professional look and correctness of your Word document answers.

Be sure to come see me early with any questions! Also, make sure you reread the Academic Dishonesty policy from the course syllabus.

This is NOT a group project.