CS130 Assignment #3

Date assigned: Thursday, April 26, 2012
Date due: Tuesday, May 8, 2012
Points: 75

Goals: Excel Regression, SPSS data definition, type, measure, descriptive statistics, charts, hypothesis testing

Create a Word document **WordAnswers3PUNuetID.docx** using your PUNetID that will contain answers to each of the following questions in order. The Word document and all other files created are to be placed in a folder called **PUNetIDSolution3** once again using your PUNetID.

Answer each question in your Word document under the header **Problem# - Question #.**

Problem #1 Use Excel to solve this problem

Many government agencies post raw data online in accessible (Excel, text, csv, 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 Mar/Apr 2012. Save this file as **CEI _PUNet.xlsx.** Answer the following questions with data from the worksheet labeled **CEI3 Leading Indicators.**

►Use correct spelling, punctuation, grammar, and capitalization in your answers.

1)   Run a linear, quadratic (x^2), and exponential regression to determine how well Unemployment Insurance Claims predicts Overtime Hours. Copy the resulting chart, containing labels, a title, trendlines, equations, and R^2 values, into the Word document.

2)   What is Independent and Dependent variable in this case? Why? Give a detailed explanation for full credit.

3)   Which regression gave the best model for the data? What is that equation's R^2 value?

4)   What number of Overtime Hours is predicted by each of the above models (regressions) when Unemployment Insurance Claims is 92,000?

5)   What number of Unemployment Insurance Claims is predicted by the linear regression model when Overtime Hours is 5?

6)   What number of Unemployment Insurance Claims is predicted by the exponential regression model when Overtime Hours is 6?

7)   How well does Unemployment Insurance Claims predict Overtime Hours? Backup your claim with data.

Problem #2 Use SPSS to solve this problem
Copy the file EarthQuake7dayM25.sav from CS130 Public to your Desktop. Rename this file **EarthQuake7dayM25_PUNetID.sav**. Each variable's type has been set to string and each variable's measure has been set to nominal.   The date file was take from this website: **http://earthquake.usgs.gov/earthquakes/catalogs/index.php#csv**   A small paragraph (CSV Files) on that website explains what each variable means. This file contains data on earthquakes that were at least magnitude 2.5 over a seven day period.

8)   Each variable's type has been set to string and each variable's measure has been set to nominal**. Set each variable's type and measure correctly**. List each variable's type and measure in the Word document**.**

9)   Find the mean, sum, standard deviation, minimum, maximum, variance, and range for the magnitude and the depth. Copy the Descriptive Statistics box into your Word document**.**

10)   Build a Bar Chart to show how many earthquakes occurred in each source network (src). Copy the Bar Chart into your Word document.

11)   Build a Pie Chart to show how many earthquakes occurred in each source network (src). Copy the Bar Chart into your Word document.

12)   Build a Scatter Chart to compare Depth and Median Magnitude. Copy the Scatter Chart into your Word document. Be sure to choose an appropriate independent variable.

Problem #3 Use SPSS to solve this problem.
Import (any way you choose) the following data into SPSS. Name the file Steps_PUNetID.sav.

ID	Gender	Height (meters)	Average Steps per Day	Average Calories Burned Per Day
1	f	1.61	8789	2930
2	f	1.44	5072	1652
3	f	1.75	4297	1360
4	f	1.62	7570	2431
5	f	1.59	8997	2980
6	f	1.57	5780	1881
7	f	1.71	6475	2205
8	f	1.56	6646	2194
9	f	1.64	5437	1822
10	f	1.81	5458	1810
11	f	1.59	4630	1568
12	m	2.03	9375	3054
13	m	1.68	4802	1657
14	m	1.91	5974	1914
15	m	1.88	9479	3161
16	m	1.65	9576	3100
17	m	1.96	6654	2139
18	m	2.00	4791	1676
19	m	1.98	7159	2425
20	m	1.92	9812	3192

►Use correct spelling, punctuation, grammar, and capitalization in your answers. Label and format all charts professionally. In the Word document described above, answer each of the following questions in order under the heading specified. That is, list the heading in bold and then give your answer.

►You are to find if the Average Calories Burned Per Day by women is significantly different than the Average Calories Burned Per Day by men.

13)   What hypothesis test will you need to perform to find this result? Fully explain why you selected the test you did.

14)   State the NULL Hypothesis for the statistical test you selected.

15)   Paste in the results of the statistical test.

16)   State your conclusion.

17)   Explain exactly what the Sig. (2-tailed) value means in this case. I don't want just if the value is less than 0.05 reject the null hypothesis or greater than 0.05 accept the null hypothesis, but in terms of the problem explain the meaning of the Sig. (2-tailed) value including the concept of chance occurrences. Someone without a statistics background must be able to understand your explanation.

► Does the mean height of all the subjects differ from 1.77 meters in a statistically significant way?

18)   What hypothesis test will you need to perform to find this result? Fully explain why you selected the test you did.

19)   State the NULL Hypothesis for the statistical test you selected.

20)   Paste in the results of the statistical test.

21)   State your conclusion.

22)   How well does Average Steps per Day predict Average Calories Burned per Day? Back up your answer with data.

Submitting your work:

To submit your work, copy your <u>single folder</u> **PUNetIDSolution3** (for me that's will4614Solution3) containing (**WordAnswers3PUNuetID.docx**, **EarthQuake7dayM25_PUNetID.sav, CEI_PUNet.xlsx,** and **Steps_PUNetID.sav**) into the **CS130 Drop** folder on Turing.

Grading:

1) Correctness of your results
2) Completeness of your results
3) Professional look and correctness of your

Be sure to come see me early with any questions!

This is NOT a group project.