

CS130 Assignment #3

Date assigned: Tuesday, March 1, 2011

Date due: Tuesday, March 8, 2011 4:30 pm

Points: 50

Goals: Excel Regression, PASW data definition, type, measure, descriptive statistics, charts

Create a Word document **WordAnswers3PUNetID.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 PASW 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 taken 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.

- 1) 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.
- 2) 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.
- 3) Build a Bar Chart to show how many earthquakes occurred in each source network (src). Copy the Bar Chart into your Word document.
- 4) Build a Pie Chart to show how many earthquakes occurred in each source network (src). Copy the Bar Chart into your Word document.
- 5) 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 #2 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 Nov/Dec 2010. Save this file as **CEI_PUNet.xlsx**. Answer the following questions.

► 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 - BONUS Use Excel to solve this problem.

We want to know which month has historically had the highest number of Unemployment Insurance Claims. Build an appropriate chart (or charts) to answer this question. You may need to write some formulas in the Excel spreadsheet to calculate values to chart (hint: xxxIF()). Copy this chart into the Word document under the heading **BONUS**. Answer the following questions under the header **BONUS-Question #**

- 1) What statistic did you graph on the Y-Axis and why?
- 2) What is the independent variable? Why? Give a detailed explanation for full credit.
- 3) Which month has historically had the highest number of Unemployment Insurance Claims?

Submitting your work:

To submit your work, copy your single folder **PUNetIDSolution3** (for me that's will4614Solution3) containing (**WordAnswers3PUNuetID.docx**, **EarthQuake7dayM25_PUNetID.sav**, **CEI_PUNet.xlsx**) 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 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.