# CS130 Assignment #3

**Date assigned:** Monday, October 10, 2016
**Date due:** Monday, October 17, 2016 (by 2:15pm ... there is no late grace for this assignment)
**Points:** 50

Create a Word document **PUNuetIDAnswers.doc** using your PUNetID that will contain answers to each of the following questions in order when applicable. Your answer document and any other files created are to be placed in a folder called **PUNetIDSolution3** once again using your PUNetID. The entire folder **PUNetIDSolution3** is to be dropped in the CS130Drop folder on grace.

## Problem #1

Create Worksheet #5 - named **Internet Hosts**

The following table represents the number of worldwide internet hosts in the domain name system (DNS)

| Year | Internet Hosts (in millions) |
|------|------------------------------|
| 1996 | 9.47 |
| 2000 | 72.4 |
| 2004 | 233.1 |
| 2008 | 541.68 |
| 2012 | 908.59 |

1) Enter the data into worksheet #5 of your existing file of worksheets so that x represents the number of years since 1990. Column headings are to be **Years Since 1990** (e.g. enter x = 6 for 1996) and **Internet Hosts (in millions)**. Make your worksheet look professional and display the internet host values to 2 decimal places.

2) Perform a polynomial regression of order 2 on the data. Make sure that the independent variable is on the X-axis and the dependent variable is on the Y-axis.

3) Paste a copy of the Scatterplot with the regression equation and R^2 value in your Word document under a heading **Problem #1 - Question 1**. Your regression equation is to show values to 6 decimal places.

4) Would this correlation be considered a "strong" correlation based on our discussions in class? Why or why not? Answer this question in your Word document under the heading **Problem #1 - Question 2**.

5) What is the predicted number of internet hosts in the year 2016? Answer this question in your Word document to two decimal places under the heading **Problem #1 - Question 3**. Show your work for full credit.

6) What is the actual number of internet hosts in the year 2016? Answer this question in your Word document to two decimal places under the heading **Problem #1 - Question 4.**

6) How many internet hosts can we expect in the year 2020 using the correct regression equation? Answer this question in your Word document under the heading **Problem #1 - Question 5**.

## Problem #2

Using the sample data set entitled "TreeData.xlsx" found in the CS130Public folder and R, answer each of the following questions in the Word document PUNetIDAnswers.doc as specified. When asked for, place a graph into your document with the appropriate explanation.

1) For each variable in TreeData.xlsx, identify a) variable name, b) qualitative or quantitative, c) discrete, continuous, neither d) nominal, ordinal, neither. For b), c), and d) briefly explain your answer for each variable. Answer this question under a heading of **Problem #2 - Question 1**.

2) Write an R script called TreeData.R that does the following:

a) Changes the location of your data set to the desktop of the computer you are on

b) Installs the Readxl package and then activates the package

c) Imports the TreeData.xlsx into a variable called treedata

d) Computes and shows on the screen the mean and standard deviation of the variable trunkgirth and weight.

e) Creates a Scatterplot of trunkgirth and weight. Yes, you will have to look this up. It's very straight forward given what we've learned in class. Make sure your plot is properly titled (i.e. plot title, proper x-axis legend, and proper y-axis legend). Paste in your answer to this question under a heading of **Problem #2 - Question 2**.

Note: I will run your R script from the Desktop. If the script does not work correctly, you will lose significant points.

3) Which variable is independent? Why? Give a detailed explanation for full credit. Answer this question under a heading of **Problem #2 - Question 3**.

## Problem #3

A common disease that affects apple trees is called cedar-apple rust. An obvious symptom is rust-colored spots on the apple leaves. The immediate cause of the fungus is red cedar trees. Removing all red cedar trees within a several mile radius should eliminate the disease. In year 1 of this experiment, eight trees were studied … specifically the number of rusted leaves on each tree were noted. After removing all red cedar trees within a specified distance from the orchard, the same trees were examined the following year in a similar manner. Here are the results:

| Tree | # of rusted leaves (Year 1) | # of rusted leaves (Year 2) |
|------|------|------|
| 001 | 37 | 31 |
| 002 | 9 | 15 |
| 003 | 83 | 56 |
| 004 | 35 | 27 |
| 005 | 49 | 54 |
| 006 | 34 | 11 |
| 007 | 72 | 60 |
| 008 | 47 | 28 |

a) You need to perform a hypothesis test to determine whether removing the cedar trees significantly affected the amount of cedar rust on the apple leaves from year 1 to year 2. What hypothesis test will you use? Explain your answer under the heading **Problem #3 – Question 1**.

Write an R script called CedarRust.R that does the following:

1) Creates a data frame for the three variables and their associated values from the above table. Assign the data frame to a variable called cedarrustdata.

2) Run the test you selected from a) above

b) Paste in your results from R under a heading **Problem #3 – Question 2**.

c) State the Null Hypothesis for the above problem under the heading **Problem #3 – Question 3**.

d) Should we accept the Null Hypothesis? Why or why not? Explain under the heading **Problem #3 – Question 4**.

e) Finally, state your conclusion under the heading **Problem #3 – Question 5**.


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. As an example, do not put Problem #3 – Question 4 before results Problem #3 – Question 2 in the Word document.

## How to Submit and Grading Policies

A copy of your single **folder** (PUNetIDSolution3) with all of the files (properly named) is to be placed in the CS130Drop folder by **2:15pm** on the due date to be considered on time.

Here are the files I'm expecting in the folder PUNetIDSoltuions3

1. PUNetIDAnswers.doc

2. PUNetIDExcel.xlsx

3. TreeData.R

4. CedarRust.R

Further, you are to turn in a printed copy of PUNetIDAnswers.doc. Page 1 of the document is to contain ALL answers to Problem #1. Page 2 is to contain ALL answers to Problem #2 and page 3 is to contain ALL answers to Problem #3. Do not use less than a 10-pt font size. Print this document out (double-sided and stapled in the upper left corner) and place on the instructor's desk no later than 2:15pm on the day the assignment is due.

Grading will be based on:

• 	Correctness of your results

• 	Completeness of your results

• 	Professional look of the worksheets, charts, and document as described above and discussed in class

• 	Ability to perform a what-if analysis in any worksheet by changing any of the user input data with accurate results computed and displayed in the worksheet.