CS130 Assignment #3

Date Assigned: Wednesday, March 8th, 2017
Date Due: Wednesday, March 15th, 2017 2:15 pm
Points: 50

Create a Word document PUNetIDHW3Solutions.doc using your PUNetID that will contain answers to each of the following questions in order when applicable. Your Solutions document and any other files created are to be placed in a folder called PUNetIDHW3 once again using your PUNetID. The entire folder PUNetIDHW3 is to be dropped in the CS130Drop folder on Grace.

Problem #1
Create a Worksheet named Enrollment in Excel Workbook called PUNetIDHW3Excel.xlsx and use this worksheet to answer the following questions in your Word document.

The following table represents the freshmen enrollment from 2008 to 2015 at Pacific University on the Forest Grove campus:

<table>
<thead>
<tr>
<th>Year</th>
<th>FG UG Freshmen Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>345</td>
</tr>
<tr>
<td>2009</td>
<td>333</td>
</tr>
<tr>
<td>2010</td>
<td>475</td>
</tr>
<tr>
<td>2011</td>
<td>374</td>
</tr>
<tr>
<td>2012</td>
<td>460</td>
</tr>
<tr>
<td>2013</td>
<td>421</td>
</tr>
<tr>
<td>2014</td>
<td>459</td>
</tr>
<tr>
<td>2015</td>
<td>479</td>
</tr>
</tbody>
</table>

Source: Office of Institutional Research and Assessment, Pacific University

0) Enter the data into the Enrollment worksheet in your PUNetIDHW3Excel.xlsx workbook so that x represents the number of years since 2007 (so x = 1 corresponds to 2008). Column headings are to be Years Since 2007 and FG UG Freshmen Enrollment. Make your worksheet look professional and display the Total Undergraduate Enrollment value as an integer.

1) Produce a scatterplot and 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. Note: This data does produce an odd looking trendline (real life data is messy). 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.

2) 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.

3) What is the predicted freshmen enrollment at Pacific in the year 2010 using your regression equation? Answer this question in your Word document under the heading Problem #1 - Question 3. Show your work for full credit.

4) What is the actual undergraduate enrollment at Pacific in the year 2010 (see table above)? Why is there a discrepancy? Answer this question in your Word document under the heading Problem #1 - Question 4.

5) How many undergraduates 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 "Basketball.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) Create a directory c:\r-data. Copy Basketball.xlsx into c:\r-data

2) Write an R script called **Basketball.R** that does the following:
   a) Changes the location of your working directory to "c:\r-data"
   b) Installs the Readxl package and then activates the package
   c) Imports the Basketball.xlsx into a variable called `basketballDataFrame`
   d) Create Ordinal Values in `basketballDataFrame$positionOrdinal` from the `basketballDataFrame$position` variable. The levels are F, G, and C and the labels are “Forward”, “Guard” and “Center”.
   e) Computes and shows on the screen the mean and standard deviation of the variables `weight` and `heightInches`.
   f) Creates a Scatterplot of `weight` versus `heightInches` (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 1**.
   g) Create a barplot showing the number of players of each type (“Forward”, “Guard” and “Center”). 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**.

3) For each variable in `basketballDataFrame`, 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 3**.

4) For your Scatterplot in f) above, which variable is independent? Why? Answer this question under a heading of **Problem #2 - Question 4**.

Note: I will run your R script from the Desktop. If the script does not work correctly, you will lose significant points.
Problem #3
We want to determine if the senior basketball players weigh more than the Freshmen. The data you will use is listed below. You must build a data frame in R by hand for this data. Do not try to load Basketball.xlsx and use that data.

<table>
<thead>
<tr>
<th>year</th>
<th>weightPounds</th>
<th>heightInches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fr</td>
<td>220</td>
<td>81</td>
</tr>
<tr>
<td>Sr</td>
<td>210</td>
<td>79</td>
</tr>
<tr>
<td>Fr</td>
<td>185</td>
<td>77</td>
</tr>
<tr>
<td>Fr</td>
<td>200</td>
<td>77</td>
</tr>
<tr>
<td>Fr</td>
<td>185</td>
<td>77</td>
</tr>
<tr>
<td>Sr</td>
<td>180</td>
<td>72</td>
</tr>
<tr>
<td>Fr</td>
<td>135</td>
<td>66</td>
</tr>
<tr>
<td>Sr</td>
<td>150</td>
<td>70</td>
</tr>
<tr>
<td>Sr</td>
<td>200</td>
<td>76</td>
</tr>
</tbody>
</table>

1) You need to perform a hypothesis test to if, on average, the Freshmen weights are different statistically than the Senior weights. What hypothesis test will you use? Explain your answer under the heading Problem #3 – Question 1 in your Word document PUNetIDHW3Solutions.doc

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

2) Creates a data frame for the three variables and their associated values from the above table. Assign the data frame to a variable called basketballWeightDataFrame. Be sure to appropriate data labels for type as in problem #2.
3) Runs the test you selected from 1) above

Using the results of the R script you created above, answer the following questions in your Word document:

4) Paste in your results from R under a heading Problem #3 – Question 2.
5) State the Null Hypothesis for the above problem under the heading Problem #3 – Question 3.
6) Should we accept the Null Hypothesis? Why or why not? Explain under the heading Problem #3 – Question 4.
7) 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 (PUNetIDHW3) 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 PUNetIDHW3:

1. PUNetIDHW3Solutions.doc
2. PUNetIDHW3Excel.xlsx
3. Basketball.R
4. BasketballWeight.R

Further, you are to turn in a **printed copy of PUNetIDHW3Solutions.doc**, and to make things consistently organized, 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 and R output as described above and discussed in class
- (Excel) Ability to perform a what-if analysis by changing any of the user input data with accurate results computed and displayed in the worksheet.