CS 130 Homework #4
75 points
Due: January 15, 2008, 3pm

Note: The Lipid Data and Candy Bars files can be found in **CS130 Public**. All analyses must be done in SPSS (except for question #4). Import the Lipid Data and Candy Bars files into SPSS if you have not already done so. You **do not** need to submit these SPSS files. You must submit three (3) Word documents and one (1) Excel file to the **CS 130 Drop** folder! Make sure your name is on the top of all three Word files. You must print out the Word documents and turn them in by the due date.

Question 1
Consider the Lipid Data file we have been using in class. The column labeled "Coffee intake (cups/day)" records how many cups of coffee a person drinks each day. The column labeled "Systolic BP" records the systolic blood pressure of a person.

In a Word document called **Hmwk4_1_PUNetID.doc**, answer each of the following questions in order:

You are to find if people that drink 3 or more cups of coffee a day have a significantly different Systolic BP than people that drink 2 or fewer cups of coffee a day.

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

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

c) Paste in the results of the statistical test.

d) State your conclusion.

e) What is the mean value for Systolic BP for people drinking 3 or more cups of coffee? 2 or fewer cups of coffee?

Question 2
Again, consider the Lipid Data file we have been using in class. The column labeled "HDL" records the level of HDL cholesterol a person has in their blood; this measurement was taken very recently. The column labeled "HDL3yrs" records the level of HDL cholesterol a person had in their blood three years ago. For the last two and a half years, these people have been on a special diet to try to lower their HDL cholesterol.

In a Word document called **Hmwk4_2_PUNetID.doc**, answer each of the following questions in order:

You are to find if the HDL levels after the special diet are statistically significantly different than the HDL levels prior to starting the diet.

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

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

c) Paste in the results of the statistical test.

d) What is the mean value for HDL and HDL-3yrs?

e) State your conclusion.

f) 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 <u>chance occurrences</u>. Someone without a statistics background should be able to understand your explanation.

Question 3

Consider the Candy Bars file we have been using in class. The column labeled "Saturated fat g" records how many grams of saturated fat each candy bar contains.

In a Word document called **Hmwk4_3_PUNetID.doc**, answer each of the following questions in order:

You are to determine if the mean of the saturated fat contained in all the candy bars differs significantly from 20 g (20 g is the recommended daily allowance for saturated fat).

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

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

c) Paste in the results of the statistical test.

d) What is the mean saturated fat for the candy bars?

e) State your conclusion.

Question 4

Open up Lipid Data.xls (from **CS130 Public**) in Excel. Save this file as **Hmwk4_4_PUNetID.xls**. You must submit this file.

Insert a new column after the column labeled Cholesterol. Title this column "Cholesterol vs. Average". In this column, display either "ABOVE AVERAGE", "BELOW AVERAGE", or "AVERAGE" depending on how this particular subject's cholesterol level relates to the *average cholesterol value for all the subjects*.

Insert a new column after the column created above. Title this column "Cholesterol vs. Female Average". In this column, display either "ABOVE AVERAGE", "BELOW AVERAGE", or "AVERAGE" depending on how this particular subject's cholesterol level relates to the average cholesterol value for the subset of subjects that are female. (Hint: look at the **daverage**() function. You may need read the help or search on the web to find a good example of this in action).

Build an XY Scatter Plot to show Cholesterol vs. LDL. Use LDL as your independent variable. Add a Linear trend line and show the equation and R^2 value on the chart. What Cholesterol level would you predict for someone with an LDL level of 160? Write the answer to this question using a comment near the chart. Be sure the comment is visible when you submit your file. Use the **forcast**() function to write a formula to check your answer. Place this formula in a clearly labeled cell near the chart. Use the help in Excel and the web to determine how to use this function.

Submitting your work:

To submit your work copy all three (3) Word files and one (1) Excel file into the **CS130 Drop** folder on Turing. You must submit this before 3pm on January 15th for your assignment to be considered on time. . You must print out the Word documents and turn them in by the due date. The late policy is as follows:

Assignments can be turned in up to 24 hours late with a penalty of 10% of the assigned points. Anything later will NOT be accepted.

Grading:

Grading will be based on:
- Correctness of your results
- Professionalism
- Completeness of your results

Good Luck!