AMS 7L **LAB #8** Spring, 2009

Linear Regression

Name: _____

Lab Section: _____

Instructions:

- The TAs/lab assistants are available to help you if you have any questions about this lab exercise. If you have any questions please raise your hand and they will get to you as quickly as possible.
- At the end of class, you will need to turn in this cover sheet to your lab instructor. *If you do not turn it in, you will not get credit for this lab.* Be sure to write your name and section above.
- The following symbol ↬ at the beginning of a question means that after you answer that question you should raise your hand and have a TA or lab assistant review your answers up to that point. Once they have reviewed your work they will initial in the appropriate space in the table below. The purpose of this check is to be sure you have answered the questions correctly.

Check-Problem ↬	Lab Instructor's Initials
5	
13	
19	
26	

Linear Regression

Objectives:

1. To practice using JMP to perform linear regression
2. To practice interpreting JMP output and drawing conclusions

Getting Started: There are two files to download today. Log onto your machine and download the `anscombe.txt` and `mercury.txt` datafiles from the class webpage:
`http://www.soe.ucsc.edu/classes/ams007/Winter08/`
After you have downloaded these files, start JMP.

Part I. Simple Linear Regression

Open the file `anscombe.txt` from `File` "Open Data Table", with the `Text Import Files` option. You should get four columns of "X" variables and four columns of "Y". These are data constructed by F.J. Anscombe to illustrate some ideas of regression ("Graphs in Statistical Analysis," *The American Statistician*, 27: 17-21(1973). Fit the regression line for predicting `y1` from `x1`. To do this, from the JMP Starter window, go to `Basic` "Bivariate" and put `x1` as "X, Regressor" and `y1` as "Y, Response". JMP should bring up a scatterplot of `x1` and `y1`.

Question #1 What is your guess for the correlation between `x1` and `y1`?

Question #2 Find the correlation using JMP by going back to the JMP Starter window and going to `Multivariate` "Multivariate" and putting both `x1` and `y1` in "Y, Columns" and hitting "OK". What does JMP compute the correlation to be?

Question #3 Go back to the "Bivariate" output window. Get the fitted regression line information by going to the Hot Spot by `Bivariate Fit of y1 By x1` and choosing `Fit Line`. What is the equation for the fitted line (round to a reasonable number of decimal places)?

Question #4 What is the value of the coefficient of determination, R^2?

↬ **Question #5** What is the p-value for testing the significance of the slope, i.e., whether there is a linear relationship between `x1` and `y1`? Is it significant?

Question #6 Make a plot of the residuals by going to the Hot Spot by `Linear Fit` and choosing `Plot Residuals`. How do the residuals look?

Question #7 Keeping that analysis window open, now do a regression of `y2` on `x2` (i.e., predicting `y2` from `x2`). What is the fitted regression line?

Question #8 What are the R^2 and p-value? How do these compare to the results for `y1` and `x1`?

Question #9 How do the residuals look? Why?

Question #10 What are the regression line, R^2, and p-value for predicting `y3` from `x3`?

Question #11 What are the regression line, R^2, and p-value for predicting `y4` from `x4`?

Question #12 Should linear regression be used in either the third or fourth cases? Why or why not?

↬ **Question #13** Why is it important to look at the scatterplot and residual plots? Why can't we just look at the regression equation, the R^2 and the p-value?

After you've been checked off, you can close all of your windows except the JMP Starter Window.

Part II. More on Regression and Residuals

Now open the `mercury.txt` file, which you hopefully will remember from the previous lab. If JMP puts everything into a single column, go to `File` "Preferences", `Text Data Files` and check `Space` for `End of Field` and try again. This datafile contains information on mercury concentrations in river fish caught at 16 stations on two rivers. For each fish caught, the length (in centimeters), weight (in grams), and mercury concentration (in parts per million) were measured.

Question #14 Consider fitting a regression of mercury on weight (i.e., predicting mercury from weight). State the null and alternative hypotheses for testing the significance of the regression, and be sure to define your population parameter.

Question #15 What is the regression equation?

Question #16 What is the predicted mercury concentration in a fish that weighs 1800 grams?

Question #17 Interpret the meaning of the R^2 value in the context of this problem.

Question #18 What is the p-value for your hypotheses?

↬ **Question #19** What do you conclude from the hypothesis test? (State your conclusions in the context of the problem, i.e., in terms of mercury and fish weight.)

Question #20 How does the plot of residuals look?

Question #21 Now fit the regression of mercury on length. What is the fitted regression equation?

Question #22 Interpret the meaning of the slope in the context of this problem.

Question #23 Interpret the meaning of the intercept in the context of this problem. Does this make physical sense?

Question #24 What is the predicted mercury concentration for a fish that is 45 centimeters long?

Question #25 What are the R^2 and p-value?

↬ **Question #26** How does the plot of residuals look?

Quit JMP and please remember to **Log Off**.