

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/Fall107/>

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 y_1 from x_1 . To do this, from the JMP Starter window, go to Basic “Bivariate” and put x_1 as “X, Regressor” and y_1 as “Y, Response”. JMP should bring up a scatterplot of x_1 and y_1 .

Question #1 What is your guess for the correlation between x_1 and y_1 ?

Question #2 Find the correlation using JMP by going back to the JMP Starter window and going to Multivariate “Multivariate” and putting both x_1 and y_1 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 y_1 By x_1 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 x_1 and y_1 ? 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 y_2 on x_2 (i.e., predicting y_2 from x_2). What is the fitted regression line?

Question #8 What are the R^2 and p -value? How do these compare to the results for y_1 and x_1 ?

Question #9 How do the residuals look? Why?

Question #10 What are the regression line, R^2 , and p -value for predicting y_3 from x_3 ?

Question #11 What are the regression line, R^2 , and p -value for predicting y_4 from x_4 ?

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. Regression, Residuals, and Transformations

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?



The residuals show variability that may be increasing as length increases. That would be a violation of an important model assumption, and our regression equation would not be valid. So we'll need to fix this. We'll take logarithms of both length and mercury concentration, and fit the regression on the log-log scale. First, go back to the data table. Create a new column and call it "Log Length". Bring up the **Formula** dialog box for this column, and from the **Functions** (grouped) menu, choose **Transcendental** and **Log** and then click on **Length** and "OK" to put the natural logarithms of the fish lengths into this new column. Next create another new column called "Log Mercury" that has the log of the mercury concentrations. Fit the regression of Log Mercury on Log Length.

Question #27 What is the regression equation for predicting Log Mercury from Log Length?

Question #28 What is the R^2 , and how does it compare to your answer to Question #25?

Question #29 What is the p -value, and how does it compare to your answer to Question #25?

Question #30 Now how does the plot of residuals look?

Question #31 Which regression equation (from Questions #21 or #27) is better to use, and why?

Question #32 What is the predicted mercury concentration for a fish that is 45 centimeters long? (Don't forget to take the log of the length, then exponentiate your result.)

Question #33 Does this match your answer to question #24? Why or why not?

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