Homework 3 Due Date:

You may either type up your results or perform you analysis by hand. Where ever you use R, include the code and output with the problem. Do so in an organized fashion. Make sure you complete all parts of all five problems.

1. Use the data for ch0502.txt to answer the questions below

   (a) Using R, create the linear regression for SBP(Y) regressed on QUET(X). Write the equation of the estimated regression line.

   (b) Test the hypothesis that the slope parameter is 0 vs not 0. Write out the hypothesis. Use your R output to find the test statistic and p-value. State your conclusion.

   (c) Using R, make a scatter plot of the data with the regression line included on the plot. Create the 95 % prediction interval and include this on the plot.

   (d) Predict the SBP of a person with a QUET = 3.4. Calculate a the 95% prediction interval around this predicted value and interpret.

   (e) In R create a residual plot and a histogram of the residuals. Do any of the assumptions of linearity, constant variance or normal distribution of residuals appear to be violated? Explain.

2. Use the data for problem 7, ch0507.txt

   (a) Using R create the scatter plot of DIST regressed on MPH.

   (b) Determine the estimated linear regression line for DIST regressed on MPH.

   (c) Create the residual plot for the regression above and assess the validity of this model.

3. Use the data for problem 7, ch0507.txt

   (a) Using R create the scatter plot of SQRTDIST regressed on MPH.

   (b) Determine the estimated linear regression line for SQRTDIST regressed on MPH.

   (c) Create the residual plot for the regression above and assess the validity of this model.

   (d) Find a 98% confidence interval β_1 and interpret.

   (e) Perform a hypothesis test to determine if the true slope is equal to 1 at a significance level of $\alpha = 0.01$. State the null and alternative hypothesis. Calculate the test statistic and state your conclusion.

   (f) Using R, create the scatter plot of the data with the regression line and the 95% confidence bands on included on the plot.

   (g) Find the predicted value for SQRTDIST for X= 45. Using the formula for the 95% confidence bands, estimate the population mean for SQRTDIST at X = 45 and interpret the meaning of this confidence interval.

4. Use the data for problem 8 ch0508.txt to answer the questions below

   (a) Using R, find the equation of the estimated regression line for SAL(Y) regressed on CGPA(X).

   (b) Create a scatter plot of the data in R and overlay the estimated line.

   (c) In R create a residual plot and a histogram of the residuals. Do any of the assumptions of linearity, constant variance or normal distribution of residuals appear to be violated? Explain.

5. Revisit problem 12 from the textbook, the problem about the child's vocabulary. The data is from a single child. What assumption of a linear regression is violated?

6. The problem uses the file lift.txt. The file has three variables INDIVIDUAL, ARMSTRENGTH, and DYNAMICLIFT.

   (a) Using R, make a scatter plot of DYNAMICLIFT vs ARMSTRENGTH. Do you expect a low or high value for the correlation coefficient?

   (b) Using R, calculate the correlation coefficient between ARMSTRENGTH and DYNAMICLIFT. Describe the strength and direction of the data.

   (c) Perform a hypothesis test to determine if $\rho \neq 0$.

   (d) Calculate R^2 by squaring the correlation coefficient. Re-calculate R^2 using the formula (SSY-SSE)/SSY. (You can find SSY and SSE from the anova(fit) command.) Compare the results.

   (e) Interpret the R^2 you calculated by (SSY- SSE)/SSY.

   (f) If you were going to use this model, how would you use it? (Prediction, demonstrating a trend, other).