Stat 615/415 (Regression) Homework #4

Analysis of Variance (sec. 2.6-2.9).

1. (**2.17**) An analyst fitted normal error regression model and conducted an F test of $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$. The P-value of the test was 0.033, and the analyst concluded that $\beta_1 \neq 0$. Was the α level used by the analyst greater than or smaller than 0.033? If the α level had been 0.01, what would have been the appropriate conclusion?

2. (**2.18**) For conducting statistical tests concerning the parameter β_1, why is the t-test more versatile than the F-test?

3. (**2.19**) When testing $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$, why is the F-test a one-sided test even though H_1 includes both cases $\beta_1 < 0$ and $\beta_1 > 0$?

4. (Continued from HW-2,3) The time it takes to transmit a file always depends on the file size. Suppose you transmitted 30 files, with the average size of 126 Kbytes and the standard deviation of 35 Kbytes. The average transmittance time was 0.04 seconds with the standard deviation of 0.01 seconds. The correlation coefficient between the time and the size was 0.86.

   In the previous homework, we fit a regression model that predicted the time it will take to transmit a 400 Kbyte file.

   (a) Compute the total, regression, and error sums of squares.

   (b) Complete the ANOVA table. Include degrees of freedom, mean squares, and the F-statistic.

   (c) Use this F-statistic to test significance of our regression model that relates transmission time to the size of the file. State H_0 and H_1, calculate the p-value, and make a conclusion. How does your test statistic and the p-value relate to your results in problem #3c of Homework-3?

   (d) Calculate R^2 and explain what it means.

5. (Continued from HW-2,3) At a gas station, 180 drivers were asked to record the mileage of their cars and the number of miles per gallon. The results are summarized in the table.

	Sample mean	Standard deviation
Mileage	24,598	14,634
Miles per gallon	23.8	3.4

   The sample correlation coefficient is $r = -0.17$. In the previous homework, we fit a regression model that described how the number of miles per gallon depends on the mileage.

   (a) Complete the ANOVA table. Include sums of squares, degrees of freedom, mean squares, and the F-statistic.

   (b) What statement can be tested using this F-statistic? Calculate the p-value and state a conclusion for this ANOVA F-test.

   (c) From your ANOVA table, estimate the standard deviation of responses, $\sigma = \text{Std}(Y_i)$.

   (d) Calculate R^2 and comment on the goodness of fit in this regression problem.

6. Computer project (**2.23, 2.67**).
Grade point average (this data set was already used in Homework-2,3).

   (a) Set up the ANOVA table. Use it to answer questions (b-e).

   (b) (Stat-615 only) What is estimated by MSR in your ANOVA table? by MSE? Under what conditions do MSR and MSE estimate the same quantity?

   (c) Conduct an F-test of whether or not $\beta_1 = 0$. Control the α level at 0.01. State the alternative and your conclusion.

   (d) How much does the variation of Y reduce when X is introduced into the regression model? What is the relative reduction?

   (e) Obtain the sample correlation coefficient and attach the appropriate sign to it, positive or negative.

   (f) (leftover from the last homework) On the same graph, plot

   - the data
   - the least squares regression line for ACT scores
   - the 95 percent confidence band for the true regression line for ACT scores between 20 and 30.