STAT 4/579: STATISTICAL LEARNING Quiz 1 Name: _____

This is Exercise 1.12.2 in the notes. These are things you should have learned in the prerequisite course.

The information below relates y, a second measurement on wood volume, to x_1, a first measurement on wood volume, x_2, the number of trees, x_3, the average age of trees, and x_4, the average volume per tree. Note that $x_4 = x_1/x_2$. Some of the information has not been reported, so that you can figure it out on your own.

Table of Coefficients

Predictor	$\hat{\beta}_k$	SE($\hat{\beta}_k$)	t	P
Constant	23.45	14.90		0.122
x_1	0.93209	0.08602		0.000
x_2		0.4721	1.5554	0.126
x_3	−0.4982	0.1520		0.002
x_4	3.486	2.274		0.132

Analysis of Variance

Source	*df*	SS	MS	F	P
Regression	4	887994			0.000
Error					
Total	54	902773			

Source	*df*	Sequential SS
x_1	1	883880
x_2	1	183
x_3	1	3237
x_4	1	694

(a) How many observations are in the data?

(b) What is R^2 for this model?

(c) What is the mean squared error?

(d) Give a 95% confidence interval for β_2.

(e) Test the null hypothesis $\beta_3 = 0$ with $\alpha = 0.05$.

(f) Test the null hypothesis $\beta_1 = 1$ with $\alpha = 0.05$.

(g) Give the F statistic for testing the null hypothesis $\beta_3 = 0$.

(h) Give $SSR(x_3|x_1, x_2)$ and find $SSR(x_3|x_1, x_2, x_4)$.

(i) Test the model with only variables x_1 and x_2 against the model with all of variables x_1, x_2, x_3, and x_4.

(j) Test the model with only variables x_1 and x_2 against the model with variables x_1, x_2, and x_3.

(k) Should the test in part (g) be the same as the test in part (j)? Why or why not?

(l) For estimating the point on the regression surface at $(x_1, x_2, x_3, x_4) = (100, 25, 50, 4)$, the standard error of the estimate for the point on the surface is 2.62. Give the estimated point on the surface, a 95% confidence interval for the point on the surface, and a 95% prediction interval for a new point with these x values.

(m) Test the null hypothesis $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ with $\alpha = 0.05$.