Department of Mathematics and Statistics
Semester 152

STAT510 First Exam Sunday April 3, 2016

Name: ______________________________ ID #: ____________________

Question	Marks	Marks Obtained
1	6	
2	8	
3	2	
4	4	
5	14	
6	6	
7	4	
8	6	
Total	50	

1. In simple linear regression, how many levels of the predictor variables should be investigated, and why?
   a. If the purpose is to estimate the slope.

   b. If the purpose is to estimate the mean response at X_h?

   c. If the purpose is to predict a new observation at X_h?

2. Consider the simple linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$ where β_0 is known.
   a. Find the least squares estimator of β_1.

   b. What is the variance of this estimator?

   c. Compare this variance to the variance of the estimator in the case where both slope and intercept are unknown.

   d. Write down a $100(1-\alpha)\%$ confidence interval for β_1.

3. Show that the residuals from a linear regression model can be written as $\boldsymbol{e} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{\epsilon}$.

4. Consider the model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i;\ i = 1,2,\cdots,n$

   If you wish the test the hypothesis : $H_0: \beta_1 = \beta_3, \beta_2 = 0$.

   a. Write down the hypothesis in matrix notation.

   b. What is the test statistic for this hypothesis in the general linear test approach?

5. An engineer wants to relate gasoline mileage y to engine displacement X_1 and number of carburetor barrels X_2. He fits two models and gets the following results

Model I

$$\hat{y} = 33.7 - 0.47\, X_1$$

Incomplete ANOVA Table

Source of Variation	SS	df
Regression	955.34	
Error	282.2	
Total		31

A 95% confidence interval for the mean gasoline mileage when $X_1 = 275\ in^3$ is (19.5,21.7)

A 95% prediction interval for the gasoline mileage when $X_1 = 275\ in^3$ is (14.3,27.0)

Model II

$$\hat{y} = 32.9 - 0.053\, X_1 + 0.959 X_2$$

Incomplete ANOVA Table

Source of Variation	SS	df
Regression	972.9	
Error		
Total		

Coefficient	Test Statistic
β_1	-8.66
β_2	1.43

A 95% confidence interval for the mean gasoline mileage when $X_1 = 275\ in^3$ and $X_2 = 2$ is (18.8,21.5)

A 95% prediction interval for the gasoline mileage when $X_1 = 275\ in^3$ and $X_2 = 2$ is (13.8,26.4)

a. Complete the ANOVA Tables.
b. Compute R^2 and R^2_{Adj} for Models I and II, and comment on the results.

c. Construct a 95% confidence interval for β_1 in model II

d. What do the confidence and prediction intervals say about the predictor variable number of carburetor barrels?

e. Find the extra sum of squares $SSR(X_2|X_1)$.

f. Test the effect of the number of carburetor barrels.

6. For each of the following regression models, indicate whether it is a general linear regression model. If it is not, state whether it can be expressed in the linear model form by a suitable transformation.
   a. $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 ln X_{i2} + \beta_3 X_{i1}^2 + \varepsilon_i$

   b. $Y_i = \varepsilon_i \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}^2)$

   c. $Y_i = ln(\beta_1 X_{i1}) + \beta_2 X_{i2} + \varepsilon_i$

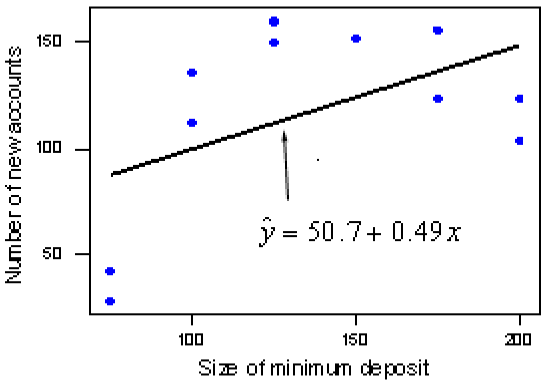
7. For a multiple regression model with four X variables,
   a. what is the relevant extra sum of squares for testing
      i. H_0: $\beta_4 = 0$.

      ii. H_0: $\beta_1 = \beta_3 = 0$.

   b. Show that

$$SSR(X_1, X_2, X_3, X_4) = SSR(X_1) + SSR(X_2, X_3|X_1) + SSR(X_4|X_1, X_2, X_3)$$

8. Consider the following scatter plot, and the associated ANOVA table.



ANOVA Table

Source of Variation	SS	DF
Regression	5141	
Error	14742	
Lack of Fit		
Pure Error	1148	
Total		

Complete the ANOVA Table and test for lack of fit at the 5% significance level (write down the null and alternative hypotheses, the test statistic, decision and conclusion).