Department of Mathematics and Statistics
Semester 142

STAT310 First Major Exam Wednesday March 18, 2014

Name: _____ ID #: _____

1) A researcher is investigating measures of the physical fitness of persons entering a 10-kilomerter race by measuring maximum oxygen uptake. He wanted to determine if maximum oxygen uptake can be obtained from a prediction equation using easily measured explanatory variables from the runners. In a preliminary study, the researcher randomly selects 54 males and records data for the variables:
    - Oxygen intake (liters/minute)
    - Weight (kilograms)
    - Age (years)
    - Time necessary to walk 1 kilometer (minutes)
    - Heart rate (beats/minute)

a) Consider the following output

Predictor	Coefficient	Standard Error
Constant	5.588	1.030
Weight	0.013	0.002
Age	-0.083	0.034
Time	-0.158	0.026
Pulse	-0.009	0.002

i) Identify the dependent and the independent variables.

ii) Write down the regression equation.

iii) Explain the coefficients of weight and age.

iv) Find a 95% confidence interval for the coefficient of age.

b) Complete the ANOVA Table

Source	SS	df	MS	F
Regression	6.106			
Residual				
Total	10.500			

i) Estimate the variance of the model.

ii) Compute R^2 and adjusted R^2.

iii) Write the hypothesis of the significance of the regression and test it.

c) Another researcher fit a simple linear regression model using pulse as the explanatory variable, and found the following

Predictor	Coefficient	Standard Error
Constant	2.545	0.494
Pulse	-0.004	0.004

$$\sum y = 46.9, \sum x = 3280.6, \sum xy = 6376$$

i) What is the meaning of the coefficient of pulse in this model?

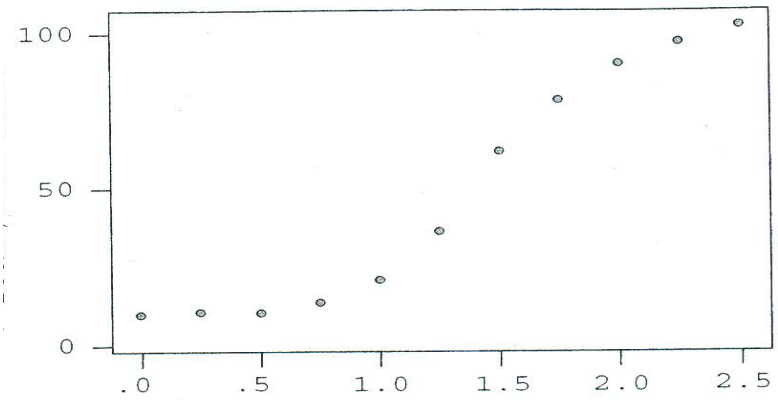
ii) Compare the coefficients of pulse in both models, and explain the difference.

iii) If we wanted to use both models to test a hypothesis about the parameters, what would that hypothesis be?

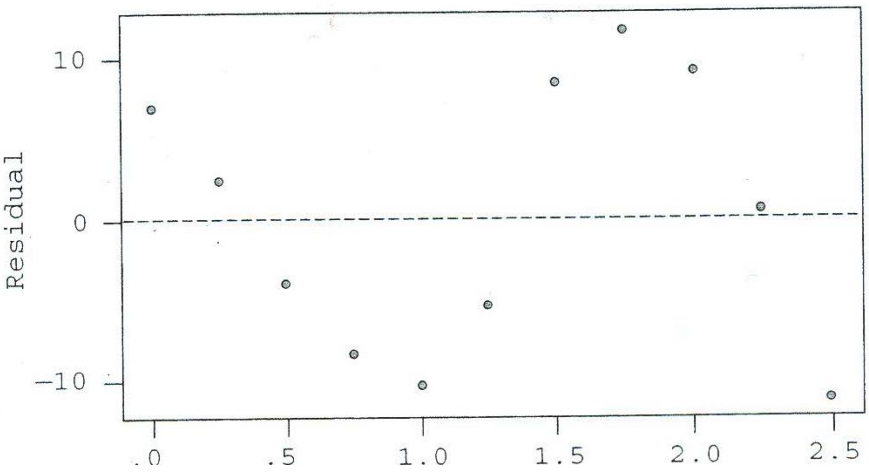
iv) Test that hypothesis.

2) Explain the following 2 plots.

a) Plot of the response vs the independent variable



b) Plot of the residuals vs the independent variable



3) Show that an equivalent way to test the significance of the regression in multiple regression is to use the statistic

$$F = \frac{R^2(n-p)}{(p-1)(1-R^2)}$$

4) In multiple regression, can the sum of squares of regression equal 0? If so, what does it mean?

5) Discuss ways to standardize regression coefficients, and reasons why.

6) Are the magnitudes of standardized regression coefficients a measure of the relative importance of the regressors? Explain.

7) What are the conditions needed to carry out a lack of fit test, and why?

8) Express the vector of residuals **e** in terms of the vector of errors **ε.**