The Open University of Sri Lanka
B.Sc./B.Ed. Degree Programme
Final Examination - 2012/2013
Applied Mathematics – Level 04
APU2141/APE4141 – Regression Analysis I

Date: 02.12.2013 Time: 9.30am – 11.30am

Answer FOUR questions only.

Non programmable calculators are permitted. Statistical Tables are provided.

Where appropriate you may use $V\left(\hat{\beta}_1\right)=\frac{\sigma^2}{\sum(x_i-\bar{x})^2}$ **and**

residual sum of squares $=\left(S_y^2-\hat{\beta}_1^2 S_x^2\right)$, **where** $S_y^2=\sum(y_i-\bar{y})^2$ **and** $S_x^2=\sum(x_i-\bar{x})^2$.

1. A researcher is interested in finding out the effect of a certain chemical compound on the diesel consumption of motor cars. After adding different amounts of the compound to diesel (in mg), varying from 2 milligrams per litre to 20 milligrams per litre, the researcher recorded the distances traveled per litre (km) by 26 cars. The summary statistics computed from the data are given below.

$$\sum x_i=223.0, \sum y_i=187.4, \sum x_i^2=2799, \sum y_i^2=1786.04, \sum x_i y_i=2207.8$$

The researcher claims that addition of 2 milligrams of the compound is associated with an increase of the expected traveling distance by 1.5 kilometers.

Suppose a simple linear regression model is appropriate to describe the relationship between the two variables.

i) Compute the Pearson correlation coefficient between the variables involved in this study. Clearly describe what it measures.

ii) Which variable would you choose as the response variable? Give reasons for your choice.

iii) In the usual notation, write down the model equation.

iv) Clearly describe all the parameters in the regression function, in relation to this study.

v) Clearly state the assumptions you would make to fit the model stated in part (iii) using the method of least squares.

2. The expected yield from a chemical process is suspected to be associated with the reaction temperature. A researcher measured the yields (in milligrams), y, from 30 samples maintained at different temperatures (in ^0C), x, varying from -10^0C to 20^0C.

   A preliminary analysis indicated that all the assumptions of a usual simple linear regression model were reasonable. The following output was obtained by fitting the simple linear regression model using the method of least squares with temperature as the predictor variable and the yield as the response variable.

Parameter	Estimate	Standard error
Intercept	1.90	0.034
Slope	0.19	0.006

   i) In relation to this study, clearly explain what is estimated by the value 1.90 in the above output.

   ii) In relation to this study, clearly explain what is estimated by the value 0.006 in the above output.

   iii) Construct a 95% confidence interval for the slope parameter.

   iv) Using part (iii) or otherwise, at 5% significance level, determine whether the data support the claim that a change of temperature by 2^0C is associated with an increase in the yield by 0.4 milligrams. Clearly state your findings.

3. In a study to understand the effect of a fertilizer on the dried weight of a medicinal plant, a researcher recorded the dried weights (in milligrams) of 20 plants four weeks after applying the fertilizer. The data collected are summarized in the accompanying table.

x (mg)	0	0	0	0	1	1	1	1	2	2
y (mg)	1.5	1.5	1.5	1.5	2.0	2.1	2.1	2.2	2.5	2.5

x (mg)	2	2	3	3	3	3	4	4	4	4
y (mg)	2.6	2.6	3.0	3.0	3.1	3.1	4.0	4.0	4.3	4.

   Consider fitting a simple linear regression model to the mean dried weight with the amount of fertilizer as the predictor variable.

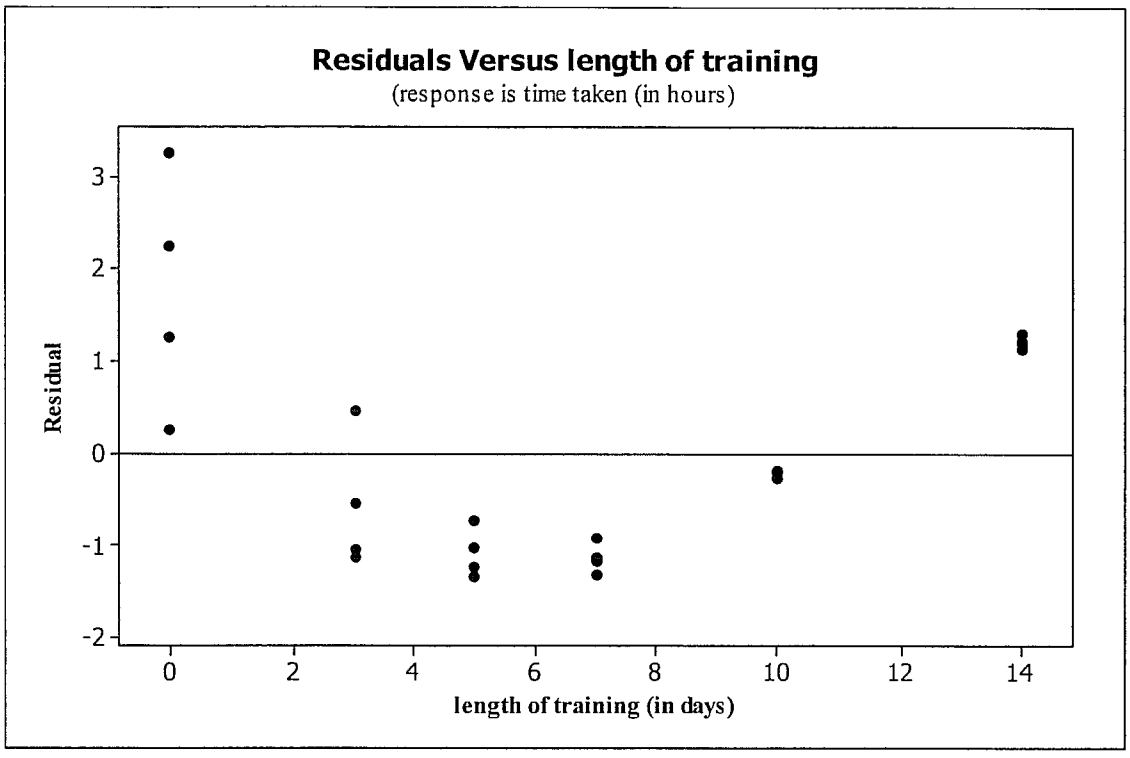
   i) Obtain the least squares estimate for the slope parameter.

   ii) Obtain the least squares estimate for the intercept parameter.

   iii) Estimate the mean squared error.

   iv) Find the coefficient of determination, R^2,, and explain what it measures in relation to this study.

4. A study was conducted to find out how a special training programme affects the time to accomplish a specific task. The researcher had used 24 persons for the study. Four replicates are used at each of the six lengths of training (days). A simple linear regression model was fitted to the time taken to accomplish the task (in minutes) using the length of training (in days) as the predictor variable. A plot of resulting residuals against the predictor variable is given below.



i) State whether each of the following statements is true or false according to the above plot. In each case, give reasons for your answer.

   a) The persons with no training have taken more time to accomplish the task than what is predicted from the fitted model.

   b) The plot indicates that the training has an effect on the time to accomplish the task.

   c) The plot cannot be used to find out whether the association between the two variables is positive or negative. .

ii) Explain the following terms in relation to this study.

   a) random error

   b) residual

5. A researcher is interested in studying the effect of a catalyst on the melting temperature of a chemical compound. The following information were obtained by fitting a simple linear regression model to the melting temperature ((in ^0C), measured on 35 samples with the amount of catalyst (in mg), x, as the predictor variable.

Parameter	Estimate	Standard error
Intercept	3.47	0.58
Slope	2.18	0.11

i) Find the equation of the fitted regression line.

ii) Estimate the expected melting temperature of samples that had not received the catalyst.

iii) Find an estimate for the standard error of the estimate computed in part (ii).

iv) Estimate the expected change in the melting temperature when the amount of catalyst is increased by 3 milligrams.

v) Find an estimate for the standard error of the estimate computed in part (iv).

6. In a study on how the dried weight of a medicinal plant varies with age, a researcher recorded the age (in weeks) and the dried weight (in mg) of 20 plants. The summary statistics computed from the data are given below.

$$\sum x_i = 120.0, \sum y_i = 374.4, \sum x_i^2 = 880.0, \sum y_i^2 = 7999.8, \sum x_i y_i = 2617.0$$

Consider the simple linear regression model for the mean dried weight using age as a predictor variable.

i) Construct an analysis of variance table for the regression analysis.

ii) Give an estimate for the random variation in the dried weight.

iii) Using a 5% significance level, test whether the age significantly help to predict the dried weight. Clearly state your findings. Assume that all the usual model assumptions are satisfied.