The Open University of Sri Lanka
B.Sc.Degree/B.Ed. Degree/Continuing Education Programme
Final Examination – 2013/2014
Applied Mathematics – Level 04
APU2141/APE4141 –Regression Analysis I

Date: 18.11.2014 Time: 9.30am – 11.30am

Instructions:

- **The duration of this examination is two hours.**
- **There are six questions of which answer four questions only.**
- **Non programmable calculators are permitted.**
- **Statistical Tables are provided.**
- **You may use** $V(\hat{\beta}) = \dfrac{\sigma^2}{\sum (x_i - \bar{x})^2}$ **and Mean squared error =** $\dfrac{(n-1)}{(n-2)}\left(S_y^2 - \hat{\beta}_1^2 S_x^2\right)$, **where** $S_y^2 = \dfrac{\sum (y_i - \bar{y})^2}{(n-1)}$ **and** $S_x^2 = \dfrac{\sum (x_i - \bar{x})^2}{(n-1)}$.

01. A researcher is interested in studying whether the age at which a fertilizer is initially applied has a significant effect on the dried weight of a medicinal plant at four months. Twenty plants in the age range of 2 weeks to 10 weeks were used for the study. The summary statistics computed from the dried weight, y (in milligrams) and the age (in weeks) are given below.

$$\sum x_i = 111, \sum y_i = 68.4, \sum x_i^2 = 747, \sum y_i^2 = 285.1, \sum x_i y_i = 302.4$$

Suppose the simple linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$ is appropriate and the model is fitted using the method of least squares.

i) Clearly state the hypothesis you would test to meet the objectives of the researcher.

ii) Construct an analysis of variance (ANOVA) table that can be used to test the hypothesis stated in part (i).

iii) Using the ANOVA table, test the hypothesis using a 5% significance level and clearly state the findings.

02. In a study to estimate the effect of nitrogen on the growth of bean plants, a researcher added known amounts of some nitrogen rich dried leaves to each of 26 tomato plants and recorded the increase in height (in cm) after four weeks. The dried weights of the nitrogen leaves used for the study had varied from 0mg to 14mg. The model $y = \beta_0 + \beta_1 x + \varepsilon$, given in the usual notation, was fitted to the data using the method of least squares. Consider the following information obtained from the data.

Variable	Sample	
	Mean x	Standard deviation
Response	3.858	0.894
Predictor	7.462	4.785

A 95% confidence interval for the slope parameter is (0.172, 0.196).

i) Estimate the parameter β_1 and describe what it measures in relation to this study.

ii) Estimate the parameter β_0 and describe what it measures in relation to this study.

iii) Briefly explain what is measured by the value 0.894 in the given output.

iv) Construct a 90% confidence interval for β_1.

v) Using a 5% significance level, determine the validity of the claim that an increase of 1mg of the dried nitrogen rich leaves is associated with a 0.1cm increase in the expected height of tomato plants.

03. The following summary statistics were computed from the data collected on the length of direct sunlight exposure, x (hours) and the stem length (cm) of 42 plants measured after five days from germination of the seeds.

$$\sum x_i = 124.6, \sum y_i = 174.47, \sum x_i^2 = 468.04, \sum y_i^2 = 833.352, \sum x_i y_i = 620.36$$

i) Find the Pearson correlation coefficient between the response and the predictor variable and clearly explain what it measures in relation to this study.

ii) Find the coefficient of determination and explain what it measures in relation to this study.

iii) Estimate the random variation in the stem length without using the information on the sunlight exposure.

iv) Estimate the random variation in the yield relying on the model $y = \beta_0 + \beta_1 x + \varepsilon$.

v) Out of the two measures computed in parts (iii) and (iv), which one do you recommend in this study? Give reasons for your answer.

04. In a study to assess the effect of a catalyst on the reaction rate of a chemical reaction, a researcher recorded reaction times, y (seconds) of 13 samples after adding known amounts of the catalyst, x (mg). The researcher fitted two models to the data collected, using the method of least squares.

Model 1: $y = \beta_0 + \beta_1 x + \varepsilon$.

Model 2: $y = \gamma_0 + \gamma_1 x^2 + \varepsilon$.

The observed y and the fitted y from the two models are given below.

Observed Y	1.7	1.1	1.5	3.7	1.1	2.5	3.2	3.2	2.8	2	3.8	4.4	2.7
Fitted Model 1	1.65	1.65	1.65	2.23	2.23	2.23	2.82	2.82	2.82	3.4	3.4	3.4	3.4
Fitted Model 2	1.96	1.96	1.96	2.12	2.12	2.12	2.6	2.6	2.6	3.4	3.4	3.4	3.4

i) State whether each of the following statements is true or false. In each case, give reasons for your answer.
   a) Both models fitted by the researcher are simple linear regression models.
   b) Given information is not adequate to decide whether the study has replicates.
   c) According to the criteria used in the method of least squares, Model 2 fits better than Model 1.

ii) Estimate the random variation in the observed response using Model 1.

iii) In relation to this study, clearly explain what is measured by your estimate computed in part (ii).

05. In a study to assess the effectiveness of a training program on the production output, the length of training, x, (in days) of 17 workers and the number of units of the product assembled in a day (y) were recorded. It was decided to fit the regression model $y = \beta_0 + \beta_1 x + \varepsilon$, using the method of least squares. The summary statistics computed from the data are given below.

$$\sum x_i = 46, \sum y_i = 68, \sum x_i^2 = 188, \sum y_i^2 = 352, \sum x_i y_i = 249$$

The management claims that the training program significantly improves the production output.

i) Estimate the parameter β_1.

ii) Estimate the standard error of the estimator used in part (i).

iii) Write down the hypothesis you would test to assess the validity of the said claim.

iv) Using a 5% significance level, test the validity of the claim and clearly state your findings.

06. In a study to estimate the effect of a chemical compound on the melting time, y (minutes), of ice, a researcher added known amounts of the compound to 25 samples each of 1 liter, prior to freezing. The researcher had used 5 replicates at each amount of the compound used, which are, 0, 2, 5, 8 and 10mg. The following information was obtained by fitting the model $y = \beta_0 + \beta_1 x^{1/2} + \varepsilon$ using the method of least squares.

parameter	estimate	standard error
intercept	1.05	0.162
slope	0.554	0.026

i) Clearly explain what the ε stands for in the model equation.

ii) Write down the equation of the fitted model.

iii) Use the fitted model to estimate the melting time of a sample of 1 liter with 4 mg of the compound added.

iv) The observed melting time of the sample described in part (iii) was 10 minutes. Will a plot of ordinary residuals indicate this as an extreme observation? Give reasons for your answer.

v) Is the observation described in part (iii) influential for estimating the slope parameter in the model? Give reasons for your answer.