THE OPEN UNIVERSITY OF SRI LANKA

B.Sc. /B.Ed. Degree Programme, Continuing Education Programme

APPLIED MATHEMATICS-LEVEL 05

ADU5301/APU2141- REGRESSION ANALYSIS 1

FINAL EXAMINATION 2017/2018

Duration: Two Hours.

Date: 31.03.2019	**Time: 9.30am-11.30am**

Instructions:

- **This question paper consists of 06 questions. Answer only four questions.**
- **Statistical Tables are provided. When reading values, you may use the closest degrees of freedom given in the table.**
- **Where appropriate, assume that the regression models are fitted using the method of least squares.**
- **In all tests, use the significance level as 0.05.**
- **Non-programmable calculators are permitted.**

1. The following results were obtained from fitting the model $y = \beta_0 + \beta_1 x + \varepsilon$, to the weights (in kg), y of 25 children, where x denotes the age of the child measured to the nearest year. The ages of children in the sample were 8, 10, 12, 14 and 16 years and there were five replicates at each level of age.

   Estimate for the slope parameter = 1.84

   Total sum of squares = 685.84

   i) Calculate $\sum (x_i - \bar{x})^2$.

   ii) Calculate the regression sum of squares and explain what it measures in relation to this study.

   iii) Calculate the residual sum of squares and the mean squared error.

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

   v) Using the results obtained in part (iv) or otherwise, test the validity of the hypothesis that a one year change in age is associated with a $2kg$ change in the weight.

2. The wastage of raw material (in grams per unit of product), y was measured in 40 mixtures used in a production with known amounts of a chemical added to the mixtures (in grams). The amounts of chemical added to the mixtures was in the range of 0 to 20 grams per unit of product.

   A simple linear regression model was fitted for y using x as the predictor variable. The least squares estimates for the intercept and the slope parameters were found to be 3.85 and 0.95 respectively. Descriptive statistics calculated from the data are as follows:

Variable	Descriptive Statistic	
	Mean	Standard deviation
Wastage (y)	11.94	5.85
Amount of chemical added (x)	8.5	5.98

   a) i) Write down the equation of the fitted line.

   ii) Using the fitted line, estimate the expected wastage per unit of product in a mixture that had received 15 grams of the chemical added.

   b) State whether each of the following statements is true or false in relation to this study. **In each case, give reasons for your answer.**

   i) Since estimated slope of the regression line is less than 1, the chemical added is associated with a reduction in the wastage of raw material.

   ii) Given information is not adequate to calculate the Pearson correlation coefficient between the amount of chemical added and the wastage of raw material. If the given information is not adequate, state the additional information that is needed. If not, calculate the Pearson correlation coefficient.

3. i) State whether each of the following statements is true or false. **In each case, give reasons for your answer.**

   a) If the Pearson correlation coefficient calculated from the data is close to 0, a simple linear regression model cannot provide a good fit to the data.

   b) The Pearson correlation coefficient between two variables y and x was found to be 0.68. If 0.04 is added to all the x values and Pearson correlation coefficient is recalculated, the new value will be 0.72.

   c) If the data are tightly scattered about a line parallel to the X-axis, Pearson correlation coefficient will be close to 1.

ii) The following summary statistics were computed from the amounts of a catalyst added (x) and the reaction rates measured (y) on 25 chemical samples.

$$\sum x_i = 121.0,\ \sum y_i = 115.5,\ \sum x_i^2 = 805.0,\ \sum y_i^2 = 594.57,\ \sum x_i y_i = 671.0$$

a) Compute the Pearson correlation coefficient between the reaction rate and the amount of catalyst added.

b) Clearly state all that can be concluded from the value of the Pearson correlation coefficient calculated in part (a). Assume that there were no extreme observations in the data.

4. In a study on how the dried weight of a medicinal plant changes with age, a researcher collected data on 30 plants in the age range of 4 weeks to 15 weeks.

The researcher fitted the model $y = \beta_0 + \beta_1 x + \epsilon$ using the method of least squares with dried weight y (in milligrams) as the response and the age (in weeks) as the explanatory variable and obtained the following results.

Parameter	Estimate	Standard error
β_0	5.44	0.68
β_1	1.04	0.08

Mean squared error = 1.94

i) In relation to this study, clearly state the assumptions that the researcher has to make to obtain the given information.

ii) In relation to this study, clearly explain what is measured by the mean squared error.

iii) Estimate the expected dried weight of a randomly chosen plant of 10 weeks old.

iv) Estimate the difference in the expected dried weights of two plants that are 12 weeks and 10 weeks old.

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

5. A researcher recorded the preservation times (in days) of a particular yogurt kept at different temperatures (^0C). The temperatures used for the study (in ^0C) were 3,4,5,6 and 8. The researcher had used 5 replicates at each temperature level.

Assume that in the range of temperatures used for the study, a simple linear regression model for the preservation time with temperature as the predictor variable provides a good fit. The following summary statistics were computed from the data.

$\sum x = 130, \sum y = 726.0, \sum x^2 = 750, \sum y^2 = 21362.0, \sum xy = 3638.0.$

The researcher wants to test whether the temperature has a significant effect on the preservation time or not.

i) Write down the null and the alternative hypotheses you would test to address the researcher's objective. Clearly describe the notation you use.

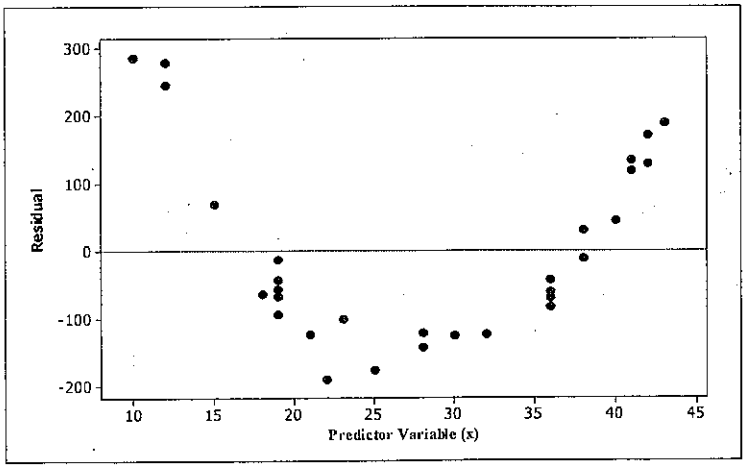
ii) What is the sample size used in this study?

iii) Compute the least squares estimate for the slope parameter.

iv) Construct an analysis of variance (ANOVA) table that can be used to address the researcher's objective.

v) Using the ANOVA table constructed in part (iii), test the hypothesis using a 0.05 significance level, and clearly state the findings.

6. A simple linear regression model was fitted for data collected on the amount of quarterly sales, y and the amount spent on advertising (in Rs) for the period, x by a company. A simple linear regression model was fitted to the data. The accompanying figure illustrates a plot of residuals from this fit against the predictor variable x.



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 plot indicates the presence of extreme observations in the data.

b) The plot indicates that the responses are correlated.

c) A plot of residuals against the fitted values will have similar pattern.

d) The plot indicates that the random errors do not have constant variance.

e) The plot indicates inadequacy of the regression function used for the model.

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

a) random error

b) residual