THE OPEN UNIVERSITY OF SRI LANKA
B.Sc. /B.Ed. Degree Programme, Continuing Education Programme
APPLIED MATHEMATICS-LEVEL 04
APU2141/APE4141- REGRESSION ANALYSIS I
FINAL EXAMINATION 2015/2016
Duration: Two Hours.

Date: 08.01.2017	**Time: 09.30 a.m. - 11.30 a.m.**

Answer FOUR questions only.

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, consider 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. A researcher wants to fit the mean response function $E(y) = \beta_0 + \beta_1 x$ for the dried weight (mg) of a medicinal plant with age (months) as the predictor variable x, assuming additive errors.

   Summary statistics computed from 26 observations are given below.

   $$\sum x = 106.0, \sum y = 64.6, \sum x^2 = 476.0, \sum y^2 = 224.56, \sum xy = 314.30.$$

   i) State any additional assumptions that are needed to obtain the least squares estimates for the regression parameters.

   ii) Calculate the least squares estimate for the slope parameter.

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

   iv) Using part (iii) or otherwise, test the hypothesis that an increase in the age by 2 months is associated with an increase in the dried weight by 2 milligrams and clearly state the findings.

2. The following information is extracted from the output obtained by fitting a linear regression model for the reaction time (minutes), y, of a chemical experiment using temperature as a predictor variable.

Source of variation	Sum of squares	Degrees of freedom
Regression	63075.666	2
Residual		
Total	64221.872	42

i) What is the sample size used for this study?

ii) State whether the model fitted is a simple linear regression model or a multiple linear regression model. Give reasons for your answer.

iii) Construct the analysis of variance table and test whether the variables in the fitted model significantly help to predict the response variable. Clearly state the findings.

iv) Estimate the random variation in the reaction time based on the fitted model.

v) Estimate the random variation in the reaction time, ignoring the information on the temperature.

3. A researcher fitted a simple linear regression model for the amount of dissolved oxygen (milligrams per litre) using temperature of the solution (^0C) as the predictor variable, x. The temperatures used for the study had varied from $0\,^0C$ to $30\,^0C$.
Summary statistics computed from the data are given below.

Variable	Sample mean	Sample standard deviation
Dissolved oxygen (y)	11.081	1.879
Temperature (x)	12.12	9.292

Estimate for the slope parameter from fitting the regression function $E(y) = \beta_0 + \beta_1 x$ assuming additive errors is -0.196.

i) Find the Pearson correlation coefficient, r.

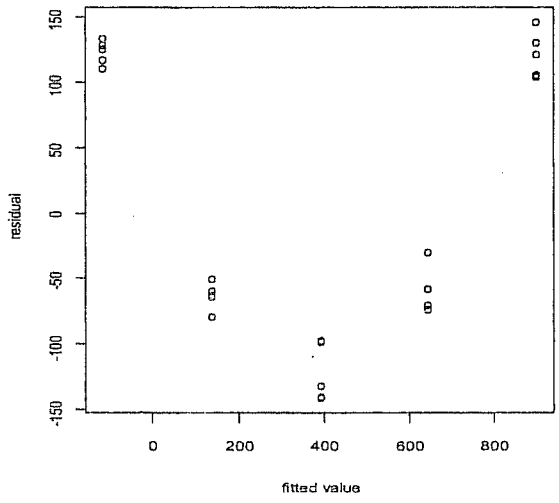
ii) Find the equation of the fitted regression line.

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

iv) Based on the fitted line, estimate the amount of dissolved oxygen in a one-litre sample at temperature $12\,^0C$.

v) In a sample, amount of dissolved oxygen per litre was found to be 9.50mg. Estimate the temperature of the sample.

4. The accompanying figure illustrates a plot of residuals against the fitted values from fitting the regression model $y = \beta_0 + \beta_1 x + \varepsilon$.



State whether each of the following statements is true or false based on the above plot. In each case, give reasons for your answer.

a) The plot indicates the presence of extreme observations.

b) The plot provides evidence against the assumption that the random errors are uncorrelated.

c) The plot indicates inappropriateness of the fitted response function.

d) The plot indicates that there are replicates in the data set.

e) The plot cannot be used to examine the normality assumption.

f) The plot of residuals against the predictor variable will have the same pattern.

5. The increase in heights (inches) of 50 plants, y, were measured two months after adding known amounts of a fertilizer (mg), x to each plant. The amounts of fertilizer added to the plants were 0, 2, 4, 8 and 10 mg. At each level of the predictor, the researcher had collected ten replicates.

The following output was obtained by fitting a simple linear regression model with amount of fertilizer as the predictor variable including an intercept term.

Description	Estimate	Standard error
Intercept	2.536	0.110
Slope	0.426	0.018

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

ii) Write down the equation of the fitted line.

iii) A student refitted the regression model after converting the data as described in each case below. State whether the fitted regression line remain unchanged or not in each case. If the fitted line changes, give the equation of the new fitted line.

a) Convert the amount of fertilizer to grams by dividing by 1000.

b) Convert the amount of fertilizer to deviations by subtracting the sample mean of x values from each value of x.

c) Convert the heights to centimeters by multiplying by 2.5.

6. In a study to estimate the relationship between the reaction time of a chemical reaction, y and the temperature of the sample, x, two researchers independently collected samples each of size 40. Both researchers used temperatures 5, 10, 15, 20 and 25 ^0C with eight replicates at each level. The output from fitting a simple linear regression model to each of the data set is given below.

Description	Researcher1	Researcher2
Fitted line	$y = 2.13 + 1.87x$	$y = 1.32 + 2.41x$
Residual sum of squares	3201.52	4189.48

State whether each of the following statements is true or false. In each case, justify your answer.

a) The fitted line with smaller residual sum of squares gives a better estimate for the unknown population regression line.

b) The estimated standard error of the estimator for the slope parameter from the data collected by Researcher1 is smaller than that of Researcher2.

c) The coefficient of determination, R^2, for the model obtained by Researcher1 is bigger than that of Researcher2.

d) Both model fits indicate positive linear association between the reaction time and temperature.

e) The model with large estimate for the slope parameter indicates stronger linear association between the reaction time and temperature.