The Open University of Sri Lanka
B.Sc./B.Ed. Degree Programme
Final Examination – 2010/2011
Applied Mathematics – Level 04
APU2141 –Regression Analysis I

Date: 19.11.2012   Time: 9.30am – 11.30am

Answer FOUR questions only.

Non programmable calculators are permitted. Statistical Tables are provided.

Where appropriate you may use $V\left(\hat{\beta_1}\right)=\dfrac{\sigma^2}{\sum\left(x_i-\bar{x}\right)^2}$ and Mean squared error $=$

$\dfrac{(n-1)}{(n-2)}\left(S_y^2-\hat{\beta_1}^2 S_x^2\right)$, where $S_y^2=\dfrac{\sum\left(y_i-\bar{y}\right)^2}{(n-1)}$ and $S_x^2=\dfrac{\sum\left(x_i-\bar{x}\right)^2}{(n-1)}$.

1. A researcher interested in studying the effect of temperatures of a certain medium on the growth of cells in a tissue culture, controlled the temperature of 32 samples at desired temperatures and measured the number of cells after 4 weeks in each culture. The following results were obtained by fitting a simple linear regression model with temperature (in $^{\circ}C$), x, as the predictor variable and the number of cells, y, as the response variable.

Parameter	Estimate	Standard error
Intercept	1.13	1.85
Slope	2.85	0.24

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

   ii) Clearly explain what is meant by the value 0.24 in the above output.

   iii) Construct 90% simultaneous confidence intervals for the intercept and the slope parameters.

   iv) Using a 10% significance level, determine whether $1.9+3x$ can be accepted as a possible mean response function or not. Clearly state your findings.

2. The following summary statistics were computed based on the data collected on ? persons in a study to find out whether there is any relationship between the height of the person, y(in inches), and the head circumference, x(in cm).

$$\sum x_i = 1675, \sum x_i^2 = 93765, \sum y_i = 1820, \sum y_i^2 = 111406, \sum x_i y_i = 102037.$$

i) Compute the Pearson correlation coefficient, r.

ii) A student reported that, r will have a higher value if both variables were measured in centimeters. Do you agree with the student? Give reasons for your answer.

iii) Suppose a simple linear regression model is fitted for the data. A student reported that the estimate for the slope parameter obtained from the raw data will change if the heights were converted into centimeters. Do you agree with the student? Give reasons for your answer.

iv) A student reported that the value of the Pearson correlation coefficient, r, alone is not adequate to decide whether there is a relationship between the two variables. Do you agree with the student? Give reasons for your answer.

3. The following summary statistics are computed from data collected on 40 medicinal plants in a study on the relationship between the dried weight of the medicinal plant (in milligrams) and its age (in weeks). The sample included plants of ages in the age range of 4 weeks to 20 weeks. Suppose the researcher wants to fit a simple linear regression model for the data using the method of least squares.

Sample mean of the dried weights of the plants = 6.2 grams

Sample mean of the ages of the plants = 9 weeks.

i) Which variable will you advise the researcher to use as the explanatory variable? Give reasons for your choice.

ii) Clearly explaining the notation you use, write down the model equation.

Suppose the researcher fitted the model using the method of least squares and found that the fitted value for plants of 14 weeks to be 8.9 grams.

iii) A student reported that the given information is not adequate to estimate the parameters indicated in each case below. If you agree with the student, clearly indicate what additional information you need in each case. If you disagree, find the estimates in each case:

a) expected dried weight of a 16 weeks old plant.

b) the age of a plant with a dried weight of 7.4 grams.

2

4. The following information is extracted from the results of a study to find out the relationship between the time taken (in minutes) for a chemical compound to completely dissolve in a solvent and the temperature of the solvent.

Suppose a simple linear regression model is fitted for the data collected on 42 samples at temperatures ranging from $20\,^0C$ to $80\,^0C$ using the method of least squares with y denoting the time (in minutes) and x denoting the temperature (in ^0C).

Based on the fitted model, a 95% confidence interval for the slope parameter is found to have confidence limits -2.1 and -0.9.

i) Give an estimate for the slope parameter.

ii) Obtain an estimate for the error in the estimate computed in Part (i).

iii) Construct a 90% confidence interval for the slope parameter.

iv) Using a 10% significance level, test whether the data provide evidence to support the hypothesis that an increase in temperature by $1\,^0C$ corresponds to a reduction in time to completely dissolve by 3 minutes or not. Clearly state your findings.

v) If the standard deviation of the temperatures used for the experiment is $6\,^0C$, obtain an estimate for the possible variation in the times taken to completely dissolve the component at $12\,^0C$.

5. In a study to find out the effect of a catalyst on the reaction time of a chemical reaction, a researcher fitted the model $y = \beta\,x + \varepsilon$, where y denotes the reaction time (in minutes) and x denotes the amount of catalyst added (in mg). The accompanying figure illustrates a plot of residuals obtained by fitting the above model using ordinary least squares against the predictor variable x.
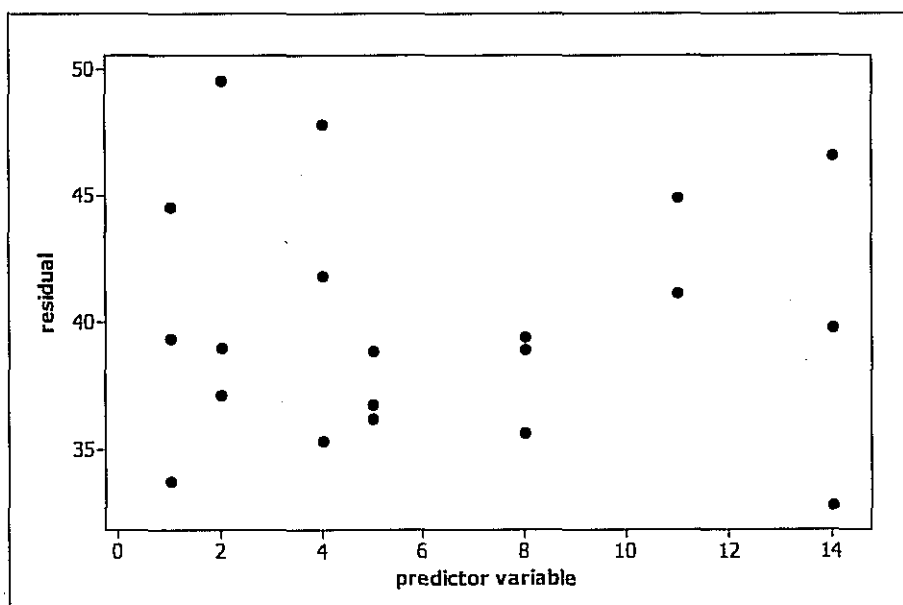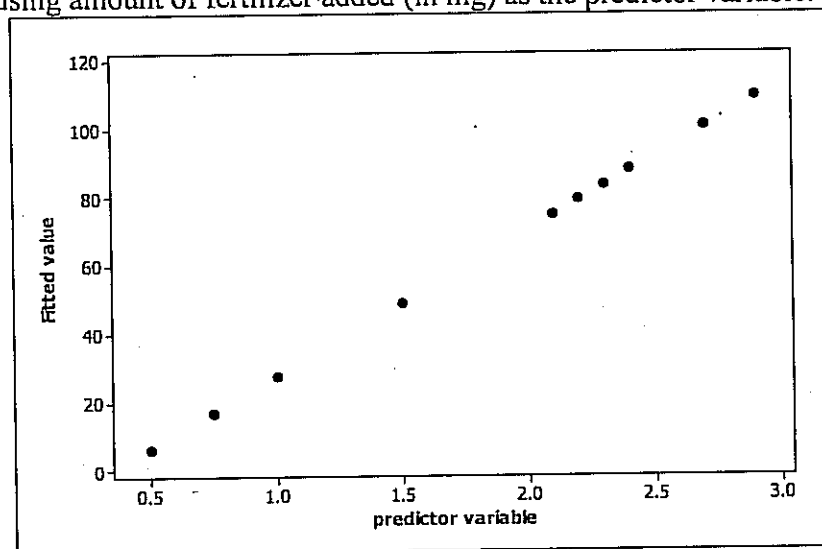
a) State whether each of the following statements is true or false. In each case, give reasons for your answer.
   i) The researcher has fitted a simple linear regression model.
   ii) The plot indicates that the fitted regression function is appropriate.
   iii) The plot can be used to estimate the residual corresponding to a new observation collected at $x = 8\,mg$.

b) Explain the following terms in relation to the above study.
   i) Residual
   ii) Regression function

6. The following is a plot of fitted values against the predictor variable obtained from fitting a simple linear regression model using ordinary least squares to data collected on yield per plot (in grams) using amount of fertilizer added (in mg) as the predictor variable.

i) Find the fitted value corresponding to $x = 1.5\,mg$. In relation to this study, what does it estimate?

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

   a) The plot indicates that the fitted response function is not appropriate.

   b) A plot of residuals against the predictor variable will provide the same information as the above plot.

   c) The plot can be used to estimate the unknown parameters in the regression function.

   d) The plot indicates that the study has no replicates.