**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: 15.08.2011 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**

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

01. The following summary statistics were computed based on 40 uncorrelated observations collected in a study on how a particular fertilizer affects the yield of beans. Let $y$ denotes the yield per plot (in kg) and $x$ denotes the amount of fertilizer applied to the plot (in grams).

$$\sum x_i=617.11, \sum y_i=281.45, \sum x_i^2=150.70, \sum y_i^2=2088.88.$$

Pearson correlation coefficient = 0.874.

i) Suppose the researcher wants to fit a simple linear regression model. Are the given information adequate to fit the model using the method of least squares? If not clearly state what additional information you need. If yes, obtain the equation of the fitted regression line.

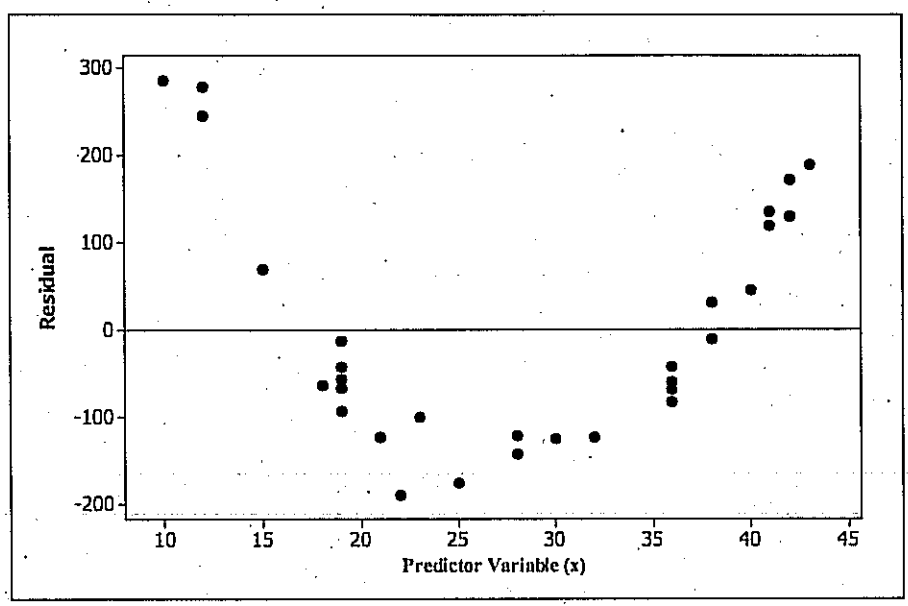
ii) State whether each of the following statements is true or false. In each case, give reasons for your answer.

a) Given information is not adequate to decide whether a simple linear regression model is appropriate or not to describe the relationship between the response and the predictor variable.

b) Based on the value of the Pearson correlation coefficient, we can conclude that the application of the said fertilizer cause an improvement in the yield if applied in amounts used for the study.

c) Change in the estimated yield per plot associated with an increase in the amount of fertilizer applied by 1 gram will be 0.874 kilograms.

02. A simple linear regression model was fitted for the data collected on the amount of quarterly sales, $y$ and the amount spent on advertising for the period (in thousand rupees), $x$ by a company. The accompanying figure illustrates a plot of residuals 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 does not indicate the presence of extreme observations in the data.

b) We can conclude that the responses collected on different experimental units are correlated.

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

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

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

a) random error

b) residual

03. In a study conducted to examine the effectiveness of a training program on the performance of carrying out a specific task, the time taken to successfully complete the task (in hours), $y$, and the length of training (in days), $x$, were recorded. The data collected on a sample of 16 participants are given below.

| $x$ | 8 | 8 | 8 | 6 | 6 | 6 | 6 | 5 |
|---|---|---|---|---|---|---|---|---|
| $y$ | 4.5 | 4.5 | 5.0 | 5.0 | 5.5 | 5.5 | 6.0 | 6.0 |

| $x$ | 5 | 5 | 5 | 4 | 4 | 3 | 3 | 3 |
|---|---|---|---|---|---|---|---|---|
| $y$ | 6.5 | 6.5 | 7.0 | 7.0 | 7.0 | 8.5 | 8.5 | 8.5 |

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

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

ii) Clearly describe all the parameters involved, in relation to this study.

iii) Fit the model using the method of least squares and write down the equation of the fitted model.

iv) Estimate the expected time taken to successfully complete the task by a randomly chosen person, if a training of five and half days is given.

04. In a study on how protein affects the weight gain of an ornamental fish species, a researcher recorded the amount of protein given (in milligrams) and the weight gain (in grams) after one month of giving the protein diet. The following summary statistics, given in the usual notation, were computed based on the data collected on 30 fish.

$$\sum x_i = 19.9, \sum y_i = 42.7, \sum x_i^2 = 15.83, \sum y_i^2 = 64.35, \sum x_i y_i = 31.02$$

Suppose a simple linear regression model is appropriate to describe the relationship between the weight gain in one month and the amount of protein given.

i) Out of the two variables given, which one will you choose as the explanatory variable? Give reasons for your answer.

ii) Give an estimate for the random variation in the response assuming that no information is available on the explanatory variable.

iii) Give an estimate for the random variation in the response, assuming that the information on the explanatory variable was available.

iv) In relation to this study, briefly explain the following.

a) Population regression line

b) Fitted regression line

05. The following summary statistics were computed based on the data collected from 32 samples by a researcher interested in finding out whether the temperature of the sample $(^0C)$ has an effect on the amount of precipitate formed ( in $mg$ ).

$\sum x_i = 982.0$ $\quad \sum y_i = 1018.5$ $\quad \sum x_i y_i = 32282.9$

$\sum x_i^2 = 31412.0$ $\quad \sum y_i^2 = 33280.7$

Suppose a simple linear regression model of the form $y = \beta_0 + \beta_1 x + \in$ is appropriate to model the relationship between $x$ and $y$.

The researcher claims that an increase of temperature by $5^0C$ is associated with an increase of the expected amount of precipitate by $4\,mg$.

i) Write down the null and the alternative hypothesis you would test to examine the validity of the researchers claim.

ii) Estimate for the slope parameter.

iii) Estimate for the error in the estimate computed in Part (ii).

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

v) Using Part (iv) or otherwise test the validity of the researcher's claim.. Clearly state the findings.

06. In a study on how the dried weight of a medicinal plant changes with age, a researcher collected data on 40 plants in the age range of 4 weeks to 10 weeks.
The researcher fitted a simple linear regression model using the method of least squares with dried weight $y$ (in grams) as the response and the age (in weeks) as the explanatory variable and obtained the following estimates.

| Parameter | estimate | standard error |
|---|---|---|
| intercept | 0.81 | 0.53 |
| slope | 0.14 | 0.07 |

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 0.07 in the above output.

iii) Estimate the expected dried weight of a randomly chosen plant of 6 weeks old.

iv) Estimate the difference in the expected dried weights of two plants that are 6 weeks and 8 weeks old.

v) Estimate the error in the estimate computed in part (iii).