RAJARATA UNIVERSITY OF SRI LANKA
FACULTY OF APPLIED SCIENCES

B.Sc. (General) Degree in Applied Sciences
Third Year – Semester I Examination – Oct / Nov 2015

MAT 3203 – Regression Analysis

Time: *Two hours*

Index No:

Instructions to candidates:

- **This examination paper contains two parts.**
- **Answer all questions in Part *A* and Part *B*.**
- **Part A – Circle the letter of the correct answer.**
- **Part B - Answer all three (03) questions on the answer script provided.**
- **Calculators and statistical tables are permitted to use.**

PART- A

01. The model developed from sample data that has the form of is known as,
a. Regression equation
b. Correlation equation
c. Estimated regression equation
d. Regression model

02. Regression model in which more than one independent variable is used to predict the dependent variable is called,
a. Simple linear regression model
b. Multiple regression model
c. Multivariate regression model
d. None of these alternatives is correct.

03. In regression analysis, the unbiased estimate of the variance is,
a. Coefficient of correlation
b. Coefficient of determination
c. Mean square error
d. Slope of the regression equation

04. In regression analysis, if the independent variable is measured in pounds, the dependent variable,
a. Must also be in pounds
b. Must be in some unit of weight
c. Cannot be in pounds
d. Can be any units

05. In regression analysis, which of the following is not a required assumption about the error term?
a. The expected value of the error term is one.
b. The variance of the error term is the same for all values of X.
c. The values of the error term are independent.
d. The error term is normally distributed.

06. Larger values of imply that the observations are more closely located beside the,
a. Average value of the independent variables
b. Average value of the dependent variable
c. Least squares line
d. Origin

P.T.0.

07. In a regression and correlation analysis if $r^2 = 1$ then,
   a. SSE = SST
   b. SSE = 1
   c. SSR = SSE
   d. SSR = SST

08. A regression analysis between sales (in Rs.1000) and price (in Rupees) resulted in the following equation

$$\hat{y} = 50000 - 8x$$

The above equation implies that an,
   a. Increase of Rs.1 in price is associated with a decrease of Rs.8 in sales
   b. Increase of Rs.8 in price is associated with an increase of Rs.8,000 in sales
   c. Increase of Rs.1 in price is associated with a decrease of Rs.42,000 in sales
   d. Increase of Rs.1 in price is associated with a decrease of Rs.8000 in sales

09. Regression analysis was applied between sales (in RS.1000) and advertising (in Rs.100) and the following regression function was obtained.

$$\hat{y} = 500 + 4x$$

Based on the above estimated regression line if advertising is Rs.10,000 , then the point estimate for sales (in Rupees) is,
   a. Rs.900
   b. Rs.900,000
   c. Rs.40,500
   d. Rs.505,000

10. In a multiple regression model, the error term ε is assumed to be a random variable with a mean of,
   a. Zero
   b. -1
   c. 1
   d. Any value

Below you are given a partial analysis output based on a sample of 25 observations. Use this results to answer questions *11* & *12*.

	Coefficient	Standard Error
Constant	145.321	48.682
X_1	25.625	9.150
X_2	-5.720	3.575
X_3	0.823	0.183

11. The estimated regression equation is,
   a. $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$
   b. $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$
   c. $\hat{Y} = 145.321 + 25.625\, X_1 - 5.720\, X_2 + 0.823\, X_3$
   d. $\hat{Y} = 48.682 + 9.15\, X_1 + 3.575\, X_2 + 1.183\, X_3$

12. The interpretation of the coefficient on x_1 is that,
   a. A one unit change in X_1 will lead to a 25.625 unit change in Y
   b. A one unit change in X_1 will lead to a 25.625 unit increase in Y when all other variables are held constant
   c. A one unit change in X_1 will lead to a 25.625 unit increase in X_2 when all other variables are held constant
   d. It is impossible to interpret the coefficient.

13. The least squares regression line is,
   a. The line which makes the sample correlation as close to +1 or –1 as possible.
   b. The line which best splits the data in half, with half of the data points lying above the regression line and half of the data points lying below the regression line.
   c. The line which minimizes the sum of the squared residuals.
   d. The line which minimizes the number of points that do not pass through the line.

14. The residual is the difference between the observed value of the independent variable and the predicted value of the independent variable.
    a. Above statement is true
    b. Above statement is false

15. In the context of a simple linear regression, the point $(\bar{x}, \bar{y})$ ..... .
    a. will always be one of the points in the data set.
    b. will always fall on the fitted line.
    c. is not informative.

PART - B

Answer all three (03) questions on the answer script provided

01. a. State the four assumptions that are used in simple linear regression.

b. For the model $Y_i = \beta X_i + \varepsilon_i$ for $i=1,2,3,\ldots,n$

i. Find the least square estimation $\hat{\beta}$ of β

ii. Show that $\hat{\beta}$ is an unbiased estimator for β.

iii. Find $var(\hat{\beta})$.

c. Suppose the following summary statistics are given for a dependent variable Y and an independent variable X.

$n = 10, S_{yy} = 3, S_{xx} = 54, S_{xy} = 12, \bar{X} = 6, \bar{Y} = 2$

i. Find b_0 and b_1 and write down the regression line of Y on X.

ii. Check the hypothesis stated below ,

$H_0: \beta_1 = 0 \; vs \; H_1: \beta_1 \neq 0$

d. Is it appropriate to fit a simple linear regression model that predicts survival (live or die) of a mouse using the quantitative variable of drug dosage (in milligrams) as a predictor variable? Briefly justify your answer.

02. Following data represents the mid semester examination marks and end semester examination marks for OR class.

Mid	75	60	70	71	80	69	73	66	79	81
End	80	82	73	77	85	77	79	68	84	89

a. State the independent variable and dependent variable for the above data set

b. Use the following Excel output to fit a regression curve for the model of the form, $y = \beta_0 + \beta_1 x + \varepsilon$ (Should mention the real name of y and x either Mid-marks and End-marks.)

Regression Statistics	
Multiple R	0.616344
R Square	0.379879
Adjusted R Square	0.302364

	Coefficients	*Standard Error*	*t Stat*
Intercept	38.41473	18.58377	2.067111
X Variable 1	0.566095	0.255717	2.213756

Cont...

c. Furthermore the lecturer added the tutorial marks to the above mark sheet as follows

Tutorial	55	50	53	60	59	51	60	63	53	54

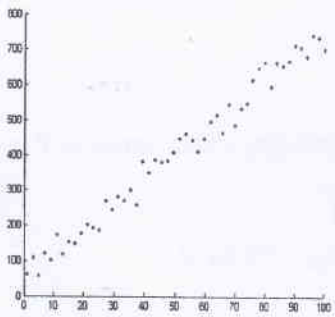
Express the new model as the matrix form as follows, $\underline{Y} = \beta\underline{X} + \underline{\varepsilon}$.
Hence note the model specification matrix for the model.

d. Estimate the above (c) model using the additional information given below.

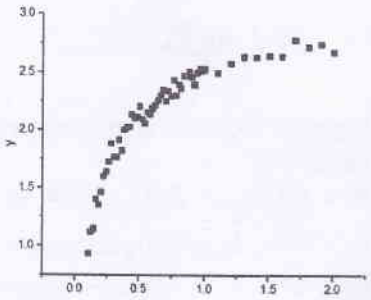
$$(X^TX) = \begin{pmatrix} 10 & 724 & 557 \\ 724 & 52814 & 40354 \\ 557 & 40354 & 31205 \end{pmatrix} \text{ and } (X^TY) = \begin{pmatrix} 797 \\ 57710 \\ 44129 \end{pmatrix}$$

e. Compare the fit of models (b) and (d) and state your comments.

03. a. Consider the scatter plots given below,



I



II

Which of the above would be intrinsically linear model? Justify your answer.

b. Assume that you have given a data set with 10,000 data points. When you are asked to fit a simple linear regression model for this data discuss the whole process that you should follow.

c. Consider the multiple linear regression output given below,

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	82624266	16524853	18.79356	9.16E-06
Residual	14	12309961	879282.9		
Total	19	94934227			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-1350.67	1326.782	-1.01801	0.325946	-4196.34	1494.996
X1	105.1368	37.21172	2.825368	0.013489	25.32554	184.9481
X2	-905.579	688.1833	-1.3159	0.209349	-2381.59	570.4283
X3	4.038254	33.28221	0.121334	0.905151	-67.3451	75.42157
X4	732.1831	257.4505	2.843976	0.013003	180.0062	1284.36
X5	23.08303	10.08736	2.288312	0.038187	1.447773	44.71829

i. Write down the general regression model which is adequate to the initial status of the above output. Hence write down the estimated regression model.

ii. Using the available information, analyse the validity of the model and make your recommendations.