RAJARATA UNIVERSITY OF SRI LANKA
FACULTY OF APPLIED SCIENCES

B.Sc. (General) Degree in Applied Sciences
Third Year - Semester I Examination – June/July 2018

MAT 3203 – REGRESSION ANALYSIS

Time: Two(02) hours

Answer All Questions.
Statistical tables and calculators will be provided.
Unless otherwise specified, symbols have their usual meaning.

1. a) Suppose $Y_1, Y_2, \ldots, Y_n$ are a set of uncorrelated random variables with common variance σ^2 and $E(Y_i) = \widehat{\beta_0} + \widehat{\beta_1} x_i$ for I = 1, 2, .., n where $\widehat{\beta_0}$ and $\widehat{\beta_1}$ are unknown parameters and the x_i's are known constants.

   Using the least square method, obtain the equations for $\widehat{\beta_0}$ and $\widehat{\beta_1}$.

   (35 marks)

   b) Consider the summary quantities obtained from a regression model derived for a chemical process, where the yield of the process (Y) is assumed to be related to the reaction temperature (X). Assume that $y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ regression model is appropriate, where $\epsilon_i \sim N(0, \sigma^2)$.

   In the usual notation, n = 20, $\sum Y_i = 50$, $\sum X_i = 100$, $\sum Y_i^2 = 134.84$, $\sum X_i^2 = 509.12$, and $\sum X_i Y_i = 257.66$.

   i. Obtain values for $\widehat{\beta_0}$ and $\widehat{\beta_1}$, and state the estimated regression function.

ii. Show that Sum of Squares of Errors (SSE) can be written as $(S_{YY} - \beta_1 S_{XY})$. Hence, find the coefficient of determination and interpret your answer.

iii. Obtain 95% confidence interval for $\widehat{\beta_1}$.

iv. Can the hypothesis H_0: $\widehat{\beta_1} = 0.6$ be accepted?

(65 marks)

2. a) Briefly explain *multiple linear regression*. **(10 marks)**

b) Consider the model $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$, for the following data set:

Y	x_1	x_2
293	1.6	851
230	15.5	816
172	22	1058
91	43	1201
113	33	1357
125	40	1115

Complete the following ANOVA table:

Source	DF	SS	MS	F
Regression	.....	29951.4	.....	.....
Residual	.....	.....	.....	
Total	.....	30245.3		

Using above details, construct the F test to check linear association between X and Y with $\alpha = 0.05$. **(50 marks)**

c) Considering the following simple linear regression output which was obtained from the **MINITAB** software, answer the following questions.

i. Write down the regression equation and interpret the coefficients.

ii. Briefly interpret the **MINITAB** output.

(40 marks)

Predictor	Coef	SE Coef	T	P
Constant	389.19	23.81	16.34	0.000
Lat	-5.9776	0.5984	-9.99	0.000

S = 19.12 R-Sq = 68.0% R-Sq(adj) = 67.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	36464	36464	99.80	0.000
Residual Error	47	17173	365		
Total	48	53637			

3. A study was conducted on the effect of temperature on the yield of a chemical process. The following data (in coded form) were collected:

X	-5	-4	-3	-2	-1	0	1	2	3	4	5
Y	1	5	4	7	10	8	9	13	14	13	18

i. Find the summary statistics $S_{YY}, S_{XX}, S_{XY}, \bar{X}$ and $\bar{Y}$.

ii. In the usual notation, use $\sum Y_i^2$, $\sum X_i^2$ and $\sum XY$ to find the matrix $\boldsymbol{X'X}$ and vector $\boldsymbol{X'Y}$. Write the normal equations in matrix form.

iii. Estimate the parameters, and find the dispersion matrix of the parameter vector.

iv. Find the Sum of Squares of Error (SSE), and Sum of Squares of Regression (SSR), and hence, find the coefficient of determination.

v. Prepare the ANOVA table and interpret the result. **(100 marks)**

4. a) i. Write down the three procedures of model selection. **(15 marks)**

ii. Briefly explain one procedure mentioned in part (a) with steps. **(15 marks)**

iii. Explain the uses of R^2, R^2_{adj}, MSE and C_p statistics in Model selection procedure.

(20marks)

b) For a multiple regression model with 4 independent variables, the following statistics for variable selection were obtained:

Model	Variable	R^2	R^2_{adj}	C_p	$\sqrt{MSE}$	1	2	3	4
A	1	67.5	64.5	138.7	8.9639				X
B	1	68.6	63.6	142.5	9.0771		X		
C	1	53.4	49.2	202.5	10.727	X			
D	1	28.6	22.1	315.2	13.278			X	
E	2	97.9	97.4	2.7	5.4063	X	X		
F	2	98.2	96.7	5.5	2.7343	X			X
G	2	93.5	92.2	22.4	4.1921			X	X
H	2	84.7	81.6	62.4	6.4455		X	X	
I	2	68.0	61.1	138.2	9.3214		X		X
J	3	98.2	97.6	3.0	2.3087	X	X		X
K	3	98.2	97.6	3.0	2.3121		X	X	X
L	3	98.1	97.5	3.5	2.3766	X		X	X
M	3	97.3	96.4	7.3	2.8638	X	X	X	
N	4	98.2	97.4	5.0	2.4460	X	X	X	X

Select the best one variable, two variable and three variable models and justify your answer. Also select the best among these models.

(50 marks)

END