47

RAJARATA UNIVERSITY OF SRI LANKA
FACULTY OF APPLIED SCIENCES

B.Sc. (General) Degree
Third Year - Semester I Examination – March/April 2014

MAT 3203 – Regression Analysis

Answer all questions in Part A and B **Time: 2 $^1/_2$ hours**

Statistical tables and calculators are allowed.
Consider an α level of 0.05 for all statistical tests.

PART - A

Given below is the output from a linear regression analysis, which was carried out to examine the impact of annual stock rate on stock prices. Encircle the most appropriate answers.

Results Output				
Independent Variable	Regression Coefficient	Standard Error	t-value $H_0 : \beta_i=0$	Probability Level
Intercept	-7.964633	3.11101359	-2.560	0.0166
Gradient	12.548580	1.27081204	9.874	0.0001

01. What statistical inference could you make about the impact of the annual stock rate on stock prices
    a. Since 11.309 > table value, reject the null hypothesis.
    b. Since 12.546 > table value, reject the null hypothesis.
    c. Since 9.874 < table value, reject the null hypothesis.
    d. Since 9.874 > table value, reject the null hypothesis.
    e. Since 12.546 < table value, reject the null hypothesis.

02. Which one of the following assumptions is incorrect
    a. The stock price is normally distributed for any dividend rate.
    b. The stock price has the same variability for any dividend rate.
    c. The stock price for any dividend rate is a linear function of dividend rate.
    d. The difference between the stock price and the expected stock price given the dividend rate is independent form company to company.

03. If the value of R-square (the coefficient of determination) is 0.7895 :
    a. 78.95% of the sample stock prices (around the mean stock price) can be attributed to a linear relationship with the dividend rate in the population.
    b. The mean stock price will be estimated to increase $97.50 for each point increase in the rate.
    c. The mean stock price will be increase $78.95 for each point increase in the rate.
    d. The stock price will increase $78.95 for each point increase in the rate.
    e. 78.95% of the sample variability in stock price (around the mean stock price) can be attributed to a linear relationship with the dividend rate.

48

PART- B

01. (i) Using scatter plot diagrams and correlation coefficients, illustrate the three types of relationships that could exist between two quantitative variables. Setup hypotheses to test the significance of the correlation and state the rejection criteria for null hypothesis.

[15]

(ii). One hundred data values relevant to particular predictor and response variable are available. Discuss the procedure you would follow to fit and validate a simple linear regression model to the given data.
(Note : Necessary to indicate that the portion of data you use in each stage)

[15]

(iii). In the simple linear regression model $Y = \beta_0 + \beta_1 X + U, E(U) \neq 0$. Letting $\alpha_0 = E(U)$, show that the model can always be rewritten with the same slope, but with a new intercept and error, where the error has a zero expected value.

[20]

(iv). An auto part is manufactured once a month in lots that varies in size as demand fluctuates. The data below represents observations on lot size (y), and number of man-hours of labor (x) for 10 recent production runs. You are required to fit the simple regression model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ and discuss the adequacy of the model. In vector form the data are:

$$Y = \begin{pmatrix} 73 \\ 50 \\ 128 \\ 170 \\ 87 \\ 108 \\ 135 \\ 69 \\ 148 \\ 132 \end{pmatrix} \qquad X = \begin{pmatrix} 1 & 30 \\ 1 & 20 \\ 1 & 60 \\ 1 & 80 \\ 1 & 40 \\ 1 & 50 \\ 1 & 60 \\ 1 & 30 \\ 1 & 70 \\ 1 & 60 \end{pmatrix}$$

It turns out that,

$$(X^l X)^{-1} = \begin{pmatrix} 0.83529412 & -0.01470588 \\ -0.01470588 & 0.00029412 \end{pmatrix}$$

$$(X^l Y) = \begin{pmatrix} 1100 \\ 61800 \end{pmatrix}$$

$$Y^l Y = 134660$$

[50]

02. The following table contains ACT and GPA (grade point average) scores for 8 college students. Grade point average is based on a four-point scale and has been rounded to the nearest decimal point.

Student	GPA	ACT
1	2.8	2.1
2	3.4	24
3	3.0	26
4	3.5	27
5	3.6	27
6	3.0	25
7	2.7	25
8	3.7	30

49

(i). Estimate the relationship between GPA and ACT using ordinary least squares regression; that is, obtain the intercept and slope in the equation

$$G\hat{P}A = \hat{\beta}_0 + \hat{\beta}_1 ACT$$

[40]

(ii). Compute the fitted values and the residuals for each observation, and verify that the residuals (approximately) sum to zero.

[30]

(iii). What is the predicted value of GPA when ACT = 20?

[10]

(iv). How much of the variation in GPA for the 8 students is explained by ACT. Explain your answer.

[20]

03. (i). Let X denote the number of lines of executable SAS code and Y denote the execution time in seconds. You may use the following:

$n = 10$ $\quad \sum_{i=1}^{10} x_i = 16.75$ $\quad \sum_{i=1}^{10} y_i = 170$ $\quad \sum_{i=1}^{10} {x_i}^2 = 170$

$\sum_{i=1}^{10} {y_i}^2 = 2898$ $\quad \sum_{i=1}^{10} x_i y_i = 285.625$

(a). Fit the regression model $Y = \beta_0 + \beta_1 X + \varepsilon$

[10]

(b). Write down the *ANOVA* table.

[15]

(c). Perform a significance test for regression coefficient (β_1) in the model

[30]

(ii). Consider the multiple linear regression model given below

$$Y = \beta + \gamma X_1 + \eta X_2 + \varepsilon$$

(a). Obtained the least square equations for the model

[15]

(b). Using standard notations write down the general equation for the multiple linear regression model, with *n* observations and *p* regressor variables in the model. Using this model show that $E\left(\underline{\hat{\beta}}\right) = \underline{\beta}$ and $var\left(\underline{\hat{\beta}}\right) = \sigma^2 (X^I X)^{-1}$.

[25]

50

04. (i). Define the term *"intrinsically linear regression model"*

Using appropriate transformations write down the fallowing models as linear.

a. $y = \beta_0 + \beta_1 \left(\frac{1}{x}\right) + \varepsilon$

b. $y = \beta_0 x^{\beta_1} + \varepsilon$

c. $y = \beta_0 e^{\beta_1 x} + \varepsilon$

[40]

(ii) 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

(a). Complete the following ANOVA table:

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

[10]

(b). Write down the hypothesis that you would test using the above *ANOVA* table and test this hypothesis.

[20]

(c). If $\hat{\beta}_3 = 0.004087$ and the standard error of $\hat{\beta}_3$ is 0.0039, test whether the product term should be included in the model.

[20]

(d). Find the coefficient of determination and interpret the result.

[10]

****END****

t Distribution: Critical Values of *t*

51

Degrees of freedom		*Significance level*					
	Two-tailed test:	10%	5%	2%	1%	0.2%	0.1%
	One-tailed test:	5%	2.5%	1%	0.5%	0.1%	0.05%
1		6.314	12.706	31.821	63.657	318.309	636.619
2		2.920	4.303	6.965	9.925	22.327	31.599
3		2.353	3.182	4.541	5.841	10.215	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
6		1.943	2.447	3.143	3.707	5.208	5.959
7		1.894	2.365	2.998	3.499	4.785	5.408
8		1.860	2.306	2.896	3.355	4.501	5.041
9		1.833	2.262	2.821	3.250	4.297	4.781
10		1.812	2.228	2.764	3.169	4.144	4.587
11		1.796	2.201	2.718	3.106	4.025	4.437
12		1.782	2.179	2.681	3.055	3.930	4.318
13		1.771	2.160	2.650	3.012	3.852	4.221
14		1.761	2.145	2.624	2.977	3.787	4.140
15		1.753	2.131	2.602	2.947	3.733	4.073
16		1.746	2.120	2.583	2.921	3.686	4.015
17		1.740	2.110	2.567	2.898	3.646	3.965
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
21		1.721	2.080	2.518	2.831	3.527	3.819

F Distribution: Critical Values of *F* (5% significance level)

v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.36	246.46	247.32	248.01
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.42	19.43	19.44	19.45
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.71	8.69	8.67	8.66
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.87	5.84	5.82	5.80
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.64	4.60	4.58	4.56
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.96	3.92	3.90	3.87
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.53	3.49	3.47	3.44
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.24	3.20	3.17	3.15
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.03	2.99	2.96	2.94
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.86	2.83	2.80	2.77
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.74	2.70	2.67	2.65
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.64	2.60	2.57	2.54
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.55	2.51	2.48	2.46
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.48	2.44	2.41	2.39
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.42	2.38	2.35	2.33
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.37	2.33	2.30	2.28
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.33	2.29	2.26	2.23
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.29	2.25	2.22	2.19
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.26	2.21	2.18	2.16
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.22	2.18	2.15	2.12