RAJARATA UNIVERSITY OF SRI LANKA
FACULTY OF APPLIED SCIENCES

B.SC. (GENERAL) DEGREE
THIRD YEAR - SEMESTER I EXAMINATION – FEBRUARY/ MARCH 2013

MAT 3203 – REGRESSION ANALYSIS

Answer all questions **Time: 2 hours**

Statistical tables and calculators are permitted.

01. (i). The table below gives the total grain production and the production of cereals in lakhs of tonnes (rounded figures) for nine consecutive years.

Year	1	2	3	4	5	6	7	8	9
Total grain production	400	440	480	550	620	650	660	740	760
Cereal production	50	60	70	85	95	100	105	115	120

(a). Using a suitable graphical method comment on the relationship between the total grain production and the cereal production.

(b). Statistically asses the comment you have made in (a).
Hint : Pearson correlation is 0.996.

(ii). ***X*** denotes the number of lines of executable SAS code and ***Y*** denote the execution time in seconds.
You may use the followings:

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

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

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

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

(c). Test the hypothesis $\beta_0 = 0$ at $\alpha = 0.05$.

(d). How much variation is explained by the fitted regression?

02. Consider the regression model given below:

$Y = \alpha(x - \bar{x}) + \varepsilon$, *where* $\varepsilon \sim N(0, \sigma^2)$, in the usual notations

Suppose (x_i, Y_i), $i = 1,2,3,\ldots,n$ is a data set from this model.

(i). Determine the least squares estimator $\hat{\alpha}$ of α.

(ii). What is the distribution of $\hat{\alpha}$?

(iii). Derive a test for $H_0 : \alpha = 1$ *vs* $H_1 : \alpha \neq 1$

(iv). Using the data given below and taking the value $\alpha = 0.05$, test the hypothesis stated in part (iii).

x	0.5	1.5	3.2	4.2	5.1	6.5
Y	1.3	3.4	6.7	8.0	10.0	13.2

03. You are given the following data :

X	1	2	3	4	5	6	7
Y	8	17	29	34	46	42	52

(i). Use the following *XL* output to fit a regression curve for the model of the form, $Y = \beta_0 + \beta_1 x + \varepsilon$.

SUMMARY OUTPUT

Regression Statistics	
R Square	0.938057
Standard Error	4.321871
Observations	7

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	4.142857	3.652648	1.134207	0.30814
X Variable 1	7.107143	0.816757	8.701662	0.000332

(ii). Find the model specification matrix for a model of the form $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$.

(iii). Estimate the model (ii) using the additional information given below.

$$(X^I X)^{-1} = \begin{bmatrix} 2.428571 & -1.28571 & 0.1428571 \\ -1.28571 & 0.797619 & -0.0952381 \\ 0.1428571 & -0.0952381 & 0.01190476 \end{bmatrix}, \quad X^I Y = \begin{bmatrix} 228 \\ 1111 \\ 6091 \end{bmatrix}$$

$s^2 = 12.27380952$

(iv). Analyze the significance of the model (ii).

(v). Compare the fit of model (i) and (ii) to the given data.

04 (i). (a). Using standard notations write down the general equation of the multiple linear regression model, with n observations and p regressor variables in the model.

(b). Express the above in the form of a matrix notation and using the usual assumptions, show that the $E\left(\underline{\hat{\beta}}\right) = \underline{\beta}$ and $var\left(\underline{\hat{\beta}}\right) = \sigma^2(X'X)^{-1}$.

(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		

(b). Write down the hypothesis that you would test using the above *ANOVA* table and test this hypothesis at $\alpha = 0.05$.

(c). If $\hat{\beta}_3 = 0.004087$ and the standard error of $\hat{\beta}_3$ is 0.0039, using an $\alpha = 0.05$ test whether the product term should be included in the model.

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

t Distribution: Critical Values of *t*

		Significance level					
Degrees of freedom	*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)

ν_1 / ν_2	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