UNIVERSITI SAINS MALAYSIA

Second Semester Examination
Academic Session 2007/2008

April 2008

MST 565 – Linear Models
[Model Linear]

Duration : 3 hours
[Masa : 3 jam]

Please check that this examination paper consists of <u>TEN</u> pages of printed material before you begin the examination.

[Sila pastikan bahawa kertas peperiksaan ini mengandungi <u>SEPULUH</u> muka surat yang bercetak sebelum anda memulakan peperiksaan ini.]

<u>Instructions</u>: Answer **<u>all four</u>** [4] questions.

*[**<u>Arahan:</u>** Jawab **<u>semua empat</u>** [4] soalan.]*

1. (a) Let

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_3 = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$$

(i) Show whether $\mathbf{x}_1$ and $\mathbf{x}_2$ are orthogonal?
(ii) Show whether vectors $\mathbf{x}_1, \mathbf{x}_2$ and $\mathbf{x}_3$ are linearly independent?
(iii) Show whether $\{ \mathbf{x}_1, \mathbf{x}_2 , \mathbf{x}_3 \}$ form an orthonormal set?

(b) Let

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

and let $z = \mathbf{y}'\mathbf{y}$. Suppose that when $\frac{\partial z}{\partial \mathbf{y}}$ is evaluated at the point $\mathbf{y}_0$, we obtain

$$\frac{\partial z}{\partial \mathbf{y}} = \begin{bmatrix} 6 \\ -4 \\ 10 \end{bmatrix}.$$

Find $\mathbf{y}_0$.

(c) Let

$$X = \begin{bmatrix} 1 & 1 & 2 \\ 4 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

(i) Determine the rank of this matrix.
(ii) Show that X is a full rank matrix.

(d) Let X be an $n \times k$ matrix of full rank. Show that the $n \times n$ matrix of $H = X(X'X)^{-1}X'$ is an idempotent matrix.

[25 marks]

1. (a) *Andaikan*

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad \textit{and} \quad \mathbf{x}_3 = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$$

(i) *Tunjukkan, adakah* $\mathbf{x}_1$ *dan* $\mathbf{x}_2$ *berortogonal?*

(ii) *Tunjukkan, adakah vector* $\mathbf{x}_1, \mathbf{x}_2$ *dan* $\mathbf{x}_3$ *saling tak bersandar secara linear?*

(iii) *Tunjukkan, adakah {* $\mathbf{x}_1, \mathbf{x}_2$, $\mathbf{x}_3$*} suatu set yang ortonormal?*

(b) *Andaikan*

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}.$$

dan biarkan $z = \mathbf{y}'\mathbf{y}$. *Andaikan apabila* $\frac{\partial z}{\partial \mathbf{y}}$ *ditaksirkan pada titik* $\mathbf{y}_0$, *kita memperolehi*

$$\frac{\partial z}{\partial \mathbf{y}} = \begin{bmatrix} 6 \\ -4 \\ 10 \end{bmatrix}.$$

Tentukan $\mathbf{y}_0$ *tersebut?*

(c) *Andaikan*

$$X = \begin{bmatrix} 1 & 1 & 2 \\ 4 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

(i) *Tentukan pangkat bagi matriks ini.*

(ii) *Tunjukkan bahawa* X *merupakan satu matriks berpangkat penuh.*

(d) *Andaikan* X *merupakan suatu matriks* $n \times k$ *berpangkat penuh. Tunjukkan bahawa matriks* $n \times n$, $H = X(X'X)^{-1}X'$ *merupakan suatu matriks idempotent.*

[25 markah]

2. (a) Let $X' = (X_1, X_2, X_3)$ be a random vector which is distributed as $N_3(\mu, \Sigma)$ where

$$\mu = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 3 \end{pmatrix}$$

(i) Find the marginal distribution of $Y_1 = (X_1, X_2)$.

(ii) Find the conditional distribution of $(X_1, X_2 \mid X_3)$

(iii) Find the distribution of Z, where $Z = 2X_1 - X_2 + X_3$

(iv) Find the covariance of Z_1 and Z_2, where

$$Z_1 = 2X_1 - X_2 + X_3$$
$$Z_2 = X_1 + X_2 - X_3$$

(b) Let $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ where X is an $n \times (k+1)$ matrix of full rank, $\boldsymbol{\beta}$ is a vector of unknown parameters, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ random vector with mean $\mathbf{0}$ and variance $\sigma^2 I$. Show that the least squares estimator (LSE) for $\boldsymbol{\beta}$ is given by

$$\mathbf{b} = (X'X)^{-1} X'\mathbf{y}.$$

[25 marks]

3. (a) The following data represents weekly fuel consumption (y) as a function of average hourly temperature (x_1) and the chill index (x_2). Here the chill index represents other weather related factors that influence fuel consumption. The data are shown below:

Week, i	Average hourly temperature, x_1	Chill index, x_2	Fuel consumption, y (tons)
1	28.0	18	12.4
2	28.0	14	11.7
3	32.5	24	12.4
4	39.0	22	10.8
5	45.9	8	9.4
6	57.8	16	9.5
7	58.1	1	8.0
8	62.5	0	7.5

(i) Suggest a model to represent the fuel consumption.

(ii) What is the X matrix for such a model?

(iii) Find $X'X, (X'X)^{-1}$ and $X'y$.

2. (a) *Andaikan* $X' = (X_1, X_2, X_3)$ *sebagai suatu vector rawak tertabur secara* $N_3(\mu, \Sigma)$, *dengan*

$$\mu = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 3 \end{pmatrix}.$$

(i) *Dapatkan taburan sut bagi* $Y_1 = (X_1, X_2)$.

(ii) *Dapatkan taburan bersyarat bagi* $(X_1, X_2 \mid X_3)$.

(iii) *Dapatkan taburan bagi* Z, *dengan* $Z = 2X_1 - X_2 + X_3$.

(iv) *Dapatkan kovarians bagi* Z_1 *dan* Z_2, *dengan*

$$Z_1 = 2X_1 - X_2 + X_3$$
$$Z_2 = X_1 + X_2 - X_3$$

(b) *Andaikan* $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, *dengan* X *sebagai matriks* $n \times (k+1)$ *berpangkat penuh,* $\boldsymbol{\beta}$ *sebagai vektor parameter yang tak diketahui, dan* $\boldsymbol{\varepsilon}$ *merupakan* $n \times 1$ *vector rawak dengan min* $\mathbf{0}$ *dan varian* $\sigma^2 I$. *Tunjukkan bahawa penganggar kuasa dua terkecil bagi* $\boldsymbol{\beta}$ *adalah*

$$\mathbf{b} = (X'X)^{-1} X'\mathbf{y}.$$

[25 markah]

3. (a) *Berikut merupakan data yang menunjukkan penggunaan mingguan bahan api* (y) *sebagai fungsi terhadap purata suhu sejam* (x_1) *dan indeks chill* (x_2). *Indeks chill mewakili factor lain berkaitan cuaca yang mempengaruhi penggunaan bahan api. Data yang diperolehi adalah:*

Minggu, i	*Purata suhu sejam,* x_1	*Indeks chill,* x_2	*Penggunaan bahan api,* y *(tons)*
1	28.0	18	12.4
2	28.0	14	11.7
3	32.5	24	12.4
4	39.0	22	10.8
5	45.9	8	9.4
6	57.8	16	9.5
7	58.1	1	8.0
8	62.5	0	7.5

(i) *Cadangkan model untuk mewakili penggunaan bahan api tersebut.*

(ii) *Apakah matriks* X *untuk model yang dicadangkan tersebut.*

(iii) *Carikan* $X'X, (X'X)^{-1}$ *dan* $X'y$.

(iv) Using (iii), find the least squares estimator for $\boldsymbol{\beta}$ and what is the estimated model?

(v) Using PROC REG in SAS , find s^2.

(b) Using data in (a) test the hypotheses $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$ by setting $\alpha = 0.05$ and comment on your finding.

[25 marks]

4. (a) A company want to study the effect of shelf display height, at supermarkets, which has levels B (bottom), M (middle), and T (Top) on monthly demand for its bakery product. For $l = B,M,T$ we define μ_l to be the mean monthly demand at supermarkets using display height l. To compare μ_B, μ_M, μ_T, the company will employ a completely randomized design. Specifically, for $l = B,M,T$ the company will randomly select a sample of n_l supermarkets of equal sales potential. These supermarket will sell the product for one month using display height l. For $k = 1,2,\cdots,n_l$, we let

$y_{l,k}$ = the monthly demand observed at the kth supermarket using display height l

These data are obtained:

Display height		
B	M	T
$y_{B,1} = 58.2$	$y_{M,1} = 73.0$	$y_{T,1} = 52.5$
$y_{B,2} = 53.7$	$y_{M,2} = 78.1$	$y_{T,2} = 49.8$
$y_{B,3} = 56.6$	$y_{M,3} = 76.2$	$y_{T,3} = 56.0$
	$y_{M,4} = 82.0$	$y_{T,4} = 51.9$
	$y_{M,5} = 78.4$	$y_{T,5} = 53.3$

Figure 1 presents the SAS output obtained by using PROC GLM to perform a one-way ANOVA of these data.

(i) From the output, identify, report, and interpret F(means) and its associated prob-value.

(ii) Use the output to calculate 95% (individual) confidence intervals for $\mu_M - \mu_B, \mu_T - \mu_B$ and $\mu_T - \mu_M$.

(iii) Can we analyze these data using PROC ANOVA instead of PROC GLM, why?

(iv) *Menggunakan (iii), apakah penganggar kuasa dua terkecil bagi* $\boldsymbol{\beta}$ *dan apakah model yang dianggar?*

(v) *Menggunakan arahan PROC REG dalam SAS, cari* s^2.

(b) *Menggunakan data (a) di atas, lakukan pengujian hipotesis* $H_0 : \beta_2 = 0$ *melawan* $H_1 : \beta_2 \neq 0$ *dengan* $\alpha = 0.05$ *dan nyatakan kesimpulan anda.*

[25 markah]

4. (a) *Sebuah syarikat hendak mengkaji kesan ketinggian rak pameran di pasaraya iaitu rak* B *(bawah), rak* M *(tengah) dan rak* T *(atas) terhadap permintaan barangan membuat kek dan roti. Andaikan* μ_l *mewakili purata bulanan permintaan di pasaraya yang mempamerkan barangan tersebut di ketinggian rak* l. *Syarikat ini menjalankan rekabentuk ujikaji rawak untuk membandingkan* μ_B, μ_M, μ_T *dengan memilih secara rawak sample* n_l *pasaraya yang menunjukkan potensi jualan yang sama. Pasaraya ini akan menjual barangan tersebut selama sebulan yang dipamerkan di rak pada ketinggian* l. *Andaikan bagi* $k = 1, 2, \cdots, n_l$,

$y_{l,k}$ = permintaan bulanan di pasaraya ke-k yang menggunakan rak pameran di ketinggian l

Data yang diperolehi adalah seperti berikut:

Ketinggian rak		
B	*M*	*T*
$y_{B,1} = 58.2$	$y_{M,1} = 73.0$	$y_{T,1} = 52.5$
$y_{B,2} = 53.7$	$y_{M,2} = 78.1$	$y_{T,2} = 49.8$
$y_{B,3} = 56.6$	$y_{M,3} = 76.2$	$y_{T,3} = 56.0$
	$y_{M,4} = 82.0$	$y_{T,4} = 51.9$
	$y_{M,5} = 78.4$	$y_{T,5} = 53.3$

Rajah 1 mewakili output SAS yang diperolehi menggunakan PROC GLM untuk menjalankan analisis varian (ANOVA) sehala terhadap data tersebut.

(i) *Berdasarkan output tersebut, kenalpasti, tuliskan dan tafsirkan* F *(min) dan nilai* $-P$ *yang berkaitan.*

(ii) *Gunakan output tersebut untuk mengira selang keyakinan 95% (individu) bagi* $\mu_M - \mu_B, \mu_T - \mu_B$ *dan* $\mu_T - \mu_M$.

(iii) *Bolehkah data tersebut dianalisis menggunakan PROC ANOVA selain daripada PROC GLM, kenapa?*

(b) Three different treatment methods for removing organic carbon from an industrial waste are to be compared. The methods are airflotation (AF), foam separation (FS) and ferric-chloride coagulation (FCC). These data are obtained:

AF(I)	FS(II)	FCC(III)
34.6	38.8	26.7
35.1	39.0	26.7
35.3	40.1	27.0
35.8	40.9	27.1
36.1	41.0	27.5
36.5	43.2	28.1
36.8	44.9	28.1
37.2	46.9	28.7
37.4	51.6	30.7
37.7	53.6	31.2

Assuming that the one-way classification model with fixed effects is appropriate, use these data to test $H_0 : \tau_1 = \tau_2 = \tau_3$. What is your conclusion?

[25 marks]

(b) *Tiga kaedah rawatan untuk menyah karbon organic dari pembuangan sisa kilang hendak dibandingkan. Kaedah tersebut ialah gelembungan udara (AF), pemisahan buih (FS) dan pemendakan ferik-klorida (FCC). Data yang diperolehi adalah:*

AF(I)	FS(II)	FCC(III)
34.6	38.8	26.7
35.1	39.0	26.7
35.3	40.1	27.0
35.8	40.9	27.1
36.1	41.0	27.5
36.5	43.2	28.1
36.8	44.9	28.1
37.2	46.9	28.7
37.4	51.6	30.7
37.7	53.6	31.2

Andaikan bahawa model pengkelasan sehala dengan kesan tetap adalah sesuai. Gunakan data ini untuk menguji $H_0 : \tau_1 = \tau_2 = \tau_3$. *Nyatakan kesimpulan anda?*

[25 markah]

Figure 1: SAS output of a one-way ANOVA of the bakery data in Question 4(a)

The GLM Procedure

Class Level Information

Class	Levels	Values
height	3	B M T

Number of Observations Read 13
Number of Observations Used 13

Dependent Variable: demand

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1741.584410	870.792205	117.84	<.0001
Error	10	73.898667	7.389867		
Corrected Total	12	1815.483077			

R-Square	Coeff Var	Root MSE	demand Mean
0.959295	4.316551	2.718431	62.97692

Source	DF	Type I SS	Mean Square	F Value	Pr > F	DF	Type III SS
height	2	1741.584410	870.792205	117.84	<.0001	2	1741.584410

Parameter	Estimate	Standard Error	t Value	Pr > |t|
MUM-MUB	21.7066667	1.98526125	10.93	<.0001
MUT-MUB	-3.1333333	1.98526125	-1.58	0.1456
MUT-MUM	-24.8400000	1.71928667	-14.45	<.0001

Observation	Observed	Predicted	Residual	95% Confidence Limits for Individual Predicted Value	
1	58.20000000	55.83333333	2.36666667	48.83926420	62.82740247
2	53.70000000	55.83333333	-2.13333333	48.83926420	62.82740247
3	55.60000000	55.83333333	-0.23333333	48.83926420	62.82740247
4	73.00000000	77.54000000	-4.54000000	70.90484342	84.17515658
5	78.10000000	77.54000000	0.56000000	70.90484342	84.17515658
6	76.20000000	77.54000000	-1.34000000	70.90484342	84.17515658
7	82.00000000	77.54000000	4.46000000	70.90484342	84.17515658
8	78.40000000	77.54000000	0.86000000	70.90484342	84.17515658
9	52.50000000	52.70000000	-0.20000000	46.06484342	59.33515658
10	49.80000000	52.70000000	-2.90000000	46.06484342	59.33515658
11	56.00000000	52.70000000	3.30000000	46.06484342	59.33515658
12	51.90000000	52.70000000	-0.80000000	46.06484342	59.33515658
13	53.30000000	52.70000000	0.60000000	46.06484342	59.33515658

Observation	Observed	Predicted	Residual	95% Confidence Limits for Mean Predicted Value	
1	58.20000000	55.83333333	2.36666667	52.33629876	59.33036790
2	53.70000000	55.83333333	-2.13333333	52.33629876	59.33036790
3	55.60000000	55.83333333	-0.23333333	52.33629876	59.33036790
4	73.00000000	77.54000000	-4.54000000	74.83120867	80.24879133
5	78.10000000	77.54000000	0.56000000	74.83120867	80.24879133
6	76.20000000	77.54000000	-1.34000000	74.83120867	80.24879133
7	82.00000000	77.54000000	4.46000000	74.83120867	80.24879133
8	78.40000000	77.54000000	0.86000000	74.83120867	80.24879133
9	52.50000000	52.70000000	-0.20000000	49.99120867	55.40879133
10	49.80000000	52.70000000	-2.90000000	49.99120867	55.40879133
11	56.00000000	52.70000000	3.30000000	49.99120867	55.40879133
12	51.90000000	52.70000000	-0.80000000	49.99120867	55.40879133
13	53.30000000	52.70000000	0.60000000	49.99120867	55.40879133

- ooo O ooo -