UNIVERSITI SAINS MALAYSIA

Second Semester Examination
Academic Session 2006/2007

April 2007

MST 565 – Linear Models
[Model Linear]

Duration : 3 hours
[Masa : 3 jam]

Please check that this examination paper consists of NINE pages of printed material before you begin the examination.

[Sila pastikan bahawa kertas peperiksaan ini mengandungi SEMBILAN muka surat yang bercetak sebelum anda memulakan peperiksaan ini.]

Instructions: Answer **all five** [5] questions.

*[**Arahan:** Jawab **semua lima** [5] soalan.]*

...2/-

1. (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 $Y_1' = (X_1, X_2)$.

[10 marks]

(ii) Find the conditional distribution of $(X_1, X_2 | X_3)$.

[10 marks]

(iii) Find $\sigma_{11.3}, \sigma_{12.3}$.

[10 marks]

(iv) Find $\rho_{12}, \rho_{31}, \rho_{12.3}$.

[10 marks]

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

[15 marks]

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

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

[15 marks]

(b) If $\sum_{i=1}^{n} (y_i - \bar{y})^2 = y'(I - \frac{1}{n}J)y$, then show that the sample variance

$$s^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}$$ is an unbiased estimator of σ^2.

[30 marks]

.../3-

1. (a) *Andaikan* $X' = (X_1, X_2, X_3)$ *sebagai suatu vektor 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* $Y_1' = (X_1, X_2)$.

[10 markah]

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

[10 markah]

(iii) *Dapatkan* $\sigma_{11.3}, \sigma_{12.3}$.

[10 markah]

(iv) *Dapatkan* $\rho_{12}, \rho_{31}, \rho_{12.3}$.

[10 markah]

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

[15 markah]

(vi) *Dapatkan kovarians bagi* Z_1 and Z_2, *dengan*

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

[15 markah]

(b) *Jika* $\sum_{i=1}^{n}(y_i - \bar{y})^2 = y'(I - \frac{1}{n}J)y$, *maka tunjukkan bahawa varians sampel* $s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$ *adalah suatu penganggar saksama bagi* σ^2.

[30 markah]

2. (a) If $Y = X\beta + \varepsilon$, where X is $n \times (k+1)$ of rank k+1, then the least squares estimators of β are $\widehat{\beta} = (X'X)^{-1}(X'Y)$ and $\text{cov}(\widehat{\beta}) = \sigma^2 (X'X)^{-1}$.

[30 marks]

(b) Using (a) find the estimators of parameters and covariance of the estimators for the model $Y = \beta_0 + \beta_1 X_1 + \varepsilon$.

[20 marks]

(c) The following data represents heat evolved in calories per gram of cement (Y) as a function of the amount of each of four ingredients in the mix: tricalcium aluminate (X_1), tricalcium silicate (X_2), tetracalcium alumino ferrite (X_3), and dicalcium silicate (X_4). The data are shown below.

Observation	Y	X_1	X_2	X_3	X_4
1	78.5	7	26	6	60
2	74.3	1	29	15	52
3	104.3	11	56	8	20
4	87.6	11	31	8	47
5	95.9	7	52	6	33
6	109.2	11	55	9	22
7	102.7	3	71	17	6
8	72.5	1	31	22	44
9	93.1	2	54	18	22
10	115.9	21	47	4	26
11	83.8	1	40	23	34
12	113.3	11	66	9	12
13	109.4	10	68	8	12

Consider variables Y, X_1, and X_2 to answer the following questions.

(i) Obtain $\hat{\beta}$ and standard error of $\hat{\beta}$ and test for the parameters.

(ii) Comment on the overall fitting of the model.

(iii) Construct 95% confidence intervals for the parameters.

(iv) Obtain R^2 and interpret your result.

[50 marks]

2. (a) *Jika $Y = X\beta + \varepsilon$, dengan* **X** *merupakan suatu vector $n\times(k+1)$ berpangkat $k+1$, maka penganggar-penganggar kuasa dua terkecil bagi β adalah $\hat{\beta} = (X'X)^{-1}(X'Y)$ dan $\text{cov}(\hat{\beta}) = \sigma^2(X'X)^{-1}$.*

[30 markah]

(b) *Dengan menggunakan bahagian (a), dapatkan penganggar parameter dan kovarians penganggar model $Y = \beta_0 + \beta_1 X_1 + \varepsilon$.*

[20 markah]

(c) *Data berikut mewakili haba yang dikeluarkan, dalam kalori bagi setiap gram simen (Y) sebagai suatu fungsi bagi setiap dari empat campuran dalam bancuhan: tricalcium aluminate (X_1), tricalcium silicate (X_2), tetracalcium alumino ferrite (X_3), dan dicalcium silicate (X_4). Data tersebut ditunjukkan di bawah:*

Cerapan	Y	X_1	X_2	X_3	X_4
1	78.5	7	26	6	60
2	74.3	1	29	15	52
3	104.3	11	56	8	20
4	87.6	11	31	8	47
5	95.9	7	52	6	33
6	109.2	11	55	9	22
7	102.7	3	71	17	6
8	72.5	1	31	22	44
9	93.1	2	54	18	22
10	115.9	21	47	4	26
11	83.8	1	40	23	34
12	113.3	11	66	9	12
13	109.4	10	68	8	12

Pertimbangkan pemboleh ubah Y, X_1, dan X_2 untuk menjawab soalan-soalan berikut.

(i) *Dapatkan $\hat{\beta}$ dan ralat piawai bagi $\hat{\beta}$ dan jalankan ujian bagi parameter-parameter tersebut.*

(ii) *Berikan komen penyuaian keseluruhan bagi model.*

(iii) *Bina selang keyakinan 95% bagi parameter-parameter tersebut.*

(iv) *Dapatkan nilai R^2 dan tafsirkan keputusan anda.*

[50 markah]

.../6-

3. (a) Let $Y = X\boldsymbol{\beta} + \varepsilon$ where Y is $n\times 1$, X is $n\times p$, $\boldsymbol{\beta}$ is $p\times 1$, ε is $n\times 1$, and p=k+1. Let the partitioned vector of regression coefficients be

$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \hline \boldsymbol{\beta}_2 \end{pmatrix}$ where $\boldsymbol{\beta}_1$ is $(p\text{-}r)\times 1$ and $\boldsymbol{\beta}_2$ is $r\times 1$. Then describe the procedure for testing the hypotheses $H_0 : \boldsymbol{\beta}_2 = 0$, against $H_1 : \boldsymbol{\beta}_2 \neq 0$.

[40 marks]

(b) Fit the regression equation for independent variables X_1, X_2, X_3 and X_4 on dependent variable, Y, using the data in 2(c).

Consider the partitioned vector of regression coefficients $\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \hline \boldsymbol{\beta}_2 \end{pmatrix}$,

where $\boldsymbol{\beta}_1 = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ and $\boldsymbol{\beta}_2 = \begin{pmatrix} \beta_3 \\ \beta_4 \end{pmatrix}$. Then test the hypotheses $H_0 : \boldsymbol{\beta}_2 = 0$, against $H_1 : \boldsymbol{\beta}_2 \neq 0$ and comment on your findings.

[60 marks]

4. (a) Consider the following analysis of variance model:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i=1,2,\ldots,k, \quad j=1,2,\ldots,n.$$

Discuss the estimation procedure for the parameters of the model and also show the procedure for testing the hypothesis $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$.

[40 marks]

(b) At a certain university eight graduate programs are grouped together for administrative purposes. Enrollments in the programs for three consecutive years are shown below:

	Program							
Year	A	B	C	D	E	F	G	H
1	1	4	7	2	9	2	3	3
2	3	6	12	1	13	6	6	3
3	5	9	7	3	14	4	3	4

Write the analysis of variance model to represent the enrollments in eight graduate programs. Use the analysis of variance to test for the null hypothesis of equal program means and comment on the findings.

[60 marks]

.../7-

3. (a) *Andaikan* $Y = X\boldsymbol{\beta} + \varepsilon$ *dengan Y merupakan suatu vektor* $n \times 1$, *X merupakan suatu vektor* $n \times p$, $\boldsymbol{\beta}$ *merupakan suatu vektor* $p \times 1$, ε *merupakan suatu vektor* $n \times 1$, *dan* $p = k+1$. *Andaikan vektor pekali regresi yang telah dibahagikan adalah*

$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \hline \boldsymbol{\beta}_2 \end{pmatrix}$ *dengan* $\boldsymbol{\beta}_1$ *merupakan* $(p\text{-}r) \times 1$ *dan* $\boldsymbol{\beta}_2$ *adalah* $r \times 1$. *Terangkan prosedur bagi menguji hipotesis* $H_0 : \boldsymbol{\beta}_2 = 0$, lawan $H_1 : \boldsymbol{\beta}_2 \neq 0$.

[40 markah]

(b) *Dengan menggunakan data di bahagian 2(c), suaikan suatu persamaan regresi bagi pemboleh ubah tak bersandar* X_1, X_2, X_3 dan X_4 *terhadap pemboleh ubah bersandar Y. Pertimbangkan vektor pekali regresi yang telah dibahagikan* $\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \hline \boldsymbol{\beta}_2 \end{pmatrix}$, *dengan* $\boldsymbol{\beta}_1 = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ *dan* $\boldsymbol{\beta}_2 = \begin{pmatrix} \beta_3 \\ \beta_4 \end{pmatrix}$.

Jalankan suatu ujian hipotesis bagi $H_0 : \beta_2 = 0$, lawan $H_1 : \beta_2 \neq 0$ *dan berikan komen penemuan anda.*

[60 markah]

4. (a) *Pertimbangkan model analisis varians berikut:*

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i=1,2,\ldots,k, \quad j=1,2,\ldots,n.$$

Bincangkan prosedur penganggaran parameter-parameter model dan juga tunjukkan prosedur bagi pengujian hipotesis $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$.

[40 markah]

(b) *Dalam sebuah universiti, lapan program siswazah telah dikumpulkan untuk tujuan pentadbiran. Pendaftaran program bagi tiga tahun berturut-turut diberikan di bawah :*

	Program							
Tahun	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
1	1	4	7	2	9	2	3	3
2	3	6	12	1	13	6	6	3
3	5	9	7	3	14	4	3	4

Tulis model analisis varians untuk mewakili pendaftaran dalam lapan program siswazah tersebut. Guna analisis varians untuk menguji hipotesis nol persamaan min program dan berikan komen terhadap hasilnya.

[60 markah]

.../8-

5. (a) What are the three components of a generalized linear model? Illustrate the likelihood function for generalized linear models. What is the logit link? If $\pi(X)=\dfrac{e^{g(X)}}{1+e^{g(X)}}$, where $g(X)=\beta_0+\beta_1X_1+\ldots+\beta_pX_p$, then find the logit function, and construct the likelihood function.

[30 marks]

(b) The following data represent the birth weights of 20 infants. The variables are defined as follows:

Low Birth Weight (0=Birth Weight $\geq$ 2500 g, 1=Birth Weight <2500 g) =LOW
Age of Mothers in Years = AGE
History of Premature Labor (0=None, 1=One, etc.)=PTL
History of Hypertension (1=Yes, 0=No)=HT.

The data are shown below:

LOW	AGE	PTL	HT
0	19	0	0
0	20	0	0
0	20	0	0
0	21	0	0
0	22	0	1
0	21	1	0
0	22	0	1
0	26	1	0
0	19	0	1
0	25	0	1
1	27	1	0
1	29	0	0
1	34	0	1
1	25	1	1
1	25	0	0
1	27	0	0
1	23	0	0
1	24	1	0
1	23	1	0
1	24	1	0

Fit a linear logistic regression model in order to analyze the data on birth weight (LOW). Test for the significance of the parameters. Interpret your findings.

[70 marks]

.../9-

5. (a) *Apakah tiga komponen bagi suatu model linear teritlak? Tunjukkan fungsi kebolehjadian bagi model linear teritlak. Apakah yang dimaksudkan dengan logit link? Jika* $\pi(X) = \dfrac{e^{g(X)}}{1+e^{g(X)}}$, *dengan* $g(X) = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p$, *maka dapatkan fungsi logitnya dan bina fungsi kebolehjadian*

[30 markah]

(b) *Data berikut mewakili ukuran berat kelahiran 20 orang bayi. Pemboleh ubah-pemboleh ubah ditakrifkan seperti berikut:*
Ukuran berat kelahiran rendah (0=Ukuran Berat Kelahiran $\geq$ *2500 g, 1= Ukuran Berat Kelahiran <2500 g) =LOW*
Umur Ibu dalam Tahun = AGE
Sejarah History of Premature Labor (0=None, 1=One, etc.)=PTL
Sejarah Hipertensi (1=Yes, 0=No)=HT.

Data diberikan di bawah:

LOW	AGE	PTL	HT
0	19	0	0
0	20	0	0
0	20	0	0
0	21	0	0
0	22	0	1
0	21	1	0
0	22	0	1
0	26	1	0
0	19	0	1
0	25	0	1
1	27	1	0
1	29	0	0
1	34	0	1
1	25	1	1
1	25	0	0
1	27	0	0
1	23	0	0
1	24	1	0
1	23	1	0
1	24	1	0

Suaikan suatu model regresi logistik linear untuk menganalisis data ukuran berat kelahiran ini (LOW). Uji keertian parameter-parameter. Tafsirkan penemuan anda.

[70 markah]

- ooo O ooo -