UNIVERSITI SAINS MALAYSIA

Second Semester Examination
2014/2015 Academic Session

June 2015

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 **EIGHT** (8) questions.

*[**<u>Arahan</u> :** Jawab **LAPAN** (8) soalan.]*

In the event of any discrepancies, the English version shall be used.

[Sekiranya terdapat sebarang percanggahan pada soalan peperiksaan, versi Bahasa Inggeris hendaklah diguna pakai].

1. Show that matrix **A** below is positive definite and find a matrix **P** such that **P′P=A**.

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}$$

[8 marks]

*1. Tunjukkan matrik **A** di bawah adalah tentu positif dan cari matriks **P** supaya* **P′P=A**.

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}$$

[8 markah]

2. Show that the system below is inconsistent.

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 - x_2 + 2x_3 &= -3 \\ 3x_1 - x_2 + 5x_3 &= -2 \\ 2x_1 + x_2 + x_3 &= 4 \end{aligned}$$

[8 marks]

2. Tunjukkan sistem di bawah adalah tak konsisten.

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 - x_2 + 2x_3 &= -3 \\ 3x_1 - x_2 + 5x_3 &= -2 \\ 2x_1 + x_2 + x_3 &= 4 \end{aligned}$$

[8 markah]

3. Let $Y_1,\ldots,Y_n$ be uncorrelated random variable with equal means μ and $Var(Y_i)=\sigma_i^2$.

(a) For $\bar{Y}=\frac{1}{n}\sum_1^n Y_i$, what is $Var(\bar{Y})$?

(b) Find a constant K_n such that $Q=K_n\sum_1^n (Y_i-\bar{Y})^2$ is an unbiased estimator of $Var(\bar{Y})$.

[8 marks]

3. Biarkan $Y_1,\ldots,Y_n$ pemboleh ubah rawak tak berkolerasi dengan min yang sama μ dan $Var(Y_i)=\sigma_i^2$.

(a) Untuk $\bar{Y}=\frac{1}{n}\sum_1^n Y_i$, apakah $Var(\bar{Y})$?

(b) Cari pemalar K_n supaya $Q=K_n\sum_1^n (Y_i-\bar{Y})^2$ adalah penganggar tak pincang bagi $Var(\bar{Y})$.

[8 markah]

4. Consider the random vector $\mathbf{X}\sim N\left(\begin{bmatrix}1\\2\end{bmatrix},\begin{bmatrix}2&1\\1&2\end{bmatrix}\right)$,

(a) Find the joint distribution of X_1 and $(X_2-X_1)/2$.

(b) Are X_1 and $(X_2-X_1)/2$ independent? Explain your answer.

[10 marks]

4. Pertimbangkan vektor rawak $\mathbf{X}\sim N\left(\begin{bmatrix}1\\2\end{bmatrix},\begin{bmatrix}2&1\\1&2\end{bmatrix}\right)$,

(a) Cari taburan tercantum X_1 dan $(X_2-X_1)/2$.

(b) Adakah X_1 and $(X_2-X_1)/2$ tak bersandar? Terangkan jawapan anda.

[10 markah]

5. If $\mathbf{y}$ is $N_n\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\right)$ and the likelihood function is given by

$$L\left(\boldsymbol{\beta}, \sigma^2\right) = 1 \Big/ \left(2\pi\sigma^2\right)^{n/2} e^{-(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})/2\sigma^2},$$

(a) Differentiate $\ln L\left(\boldsymbol{\beta}, \sigma^2\right)$ with respect to $\boldsymbol{\beta}$ to obtain $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}\mathbf{y}$.

(b) Differentiate $\ln L\left(\boldsymbol{\beta}, \sigma^2\right)$ with respect to σ^2 to obtain $\hat{\sigma}^2 = \left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right) \Big/ n$.

[10 marks]

5. Jika $\mathbf{y}$ *adalah* $N_n\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\right)$ *dan fungsi kebolehjadian diberikan sebagai*

$$L\left(\boldsymbol{\beta}, \sigma^2\right) = 1 \Big/ \left(2\pi\sigma^2\right)^{n/2} e^{-(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})/2\sigma^2},$$

(a) Bezakan $\ln L\left(\boldsymbol{\beta}, \sigma^2\right)$ *terhadap* $\boldsymbol{\beta}$ *untuk mendapatkan* $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}\mathbf{y}$.

(b) Bezakan $\ln L\left(\boldsymbol{\beta}, \sigma^2\right)$ *terhadap* σ^2 *untuk mendapatkan* $\hat{\sigma}^2 = \left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right) \Big/ n$.

[10 markah]

6. Table below shows hotel room rates for several hotels. The variables are defined as follows: y= daily rate of the room, x_1= population (in hundreds) of the city in which the hotel is located, x_2= rating of the hotel (1,2 or 3) and x_3= number of rooms (units) in the hotel.

Table 1

x_1	x_2	x_3	y
385	2	59	56.0
385	1	62	38.0
1256	2	139	56.0
1256	2	51	49.0
1256	3	150	67.0
1256	3	151	60.5
350	3	89	50.5
292	2	44	49.0
185	3	101	55.0
290	2	101	65.0
3545	3	317	94.0
3645	3	159	57.0
3645	3	257	67.0
820	2	98	66.0
360	2	116	54.0
200	2	130	63.0
131	2	32	46.5
515	3	128	59.0
550	3	118	65.0
260	3	117	57.0
650	2	53	53.0
580	2	147	44.0
363	2	61	42.0
250	2	170	47.0
462	3	117	66.0
2140	2	190	43.0
2140	1	49	51.0
5750	3	175	65.0
5750	2	169	41.0
5840	3	160	78.0
5840	3	201	83.5
5840	3	117	71.0
5840	2	148	46.0
3954	3	250	94.0
3954	3	246	78.0

(a) Calculate R^2 for each the following model; $y = \beta_0 + \beta x_1 + \varepsilon$, $y = \beta_0 + \beta x_2 + \varepsilon$, and $y = \beta_0 + \beta x_3 + \varepsilon$. Which model explains the highest proportion of the total variation in the observed room rates?

(b) Which of the above models best fits the data? Justify your answers.

(c) Using the best fitted model, compute a 95% prediction interval for an individual room rate for a hotel having 100 units of room, given a rate 3 and located in a city with a population 150000.

(d) Using the best fitted model, compute a 95% confidence interval for the average room rate for a hotel having 100 units of room, given a rate 3 and located in a city with a population 150000.

[20 marks]

6. *Jadual di bawah menunjukkan kadar sewa bilik bagi beberapa hotel. Pembolehubah adalah ditakrifkan seperti berikut: y = kadar sewa bilik, x_1 = jumlah penduduk (dalam ratusan) bandar di mana hotel itu terletak, x_2 = penarafan hotel (1,2 atau 3) dan x_3 = bilangan bilik (unit) di hotel.*

Jadual 1

x_1	x_2	x_3	y
385	2	59	56.0
385	1	62	38.0
1256	2	139	56.0
1256	2	51	49.0
1256	3	150	67.0
1256	3	151	60.5
350	3	89	50.5
292	2	44	49.0
185	3	101	55.0
290	2	101	65.0
3545	3	317	94.0
3645	3	159	57.0
3645	3	257	67.0
820	2	98	66.0
360	2	116	54.0
200	2	130	63.0
131	2	32	46.5
515	3	128	59.0
550	3	118	65.0
260	3	117	57.0
650	2	53	53.0
580	2	147	44.0

...7/-

363	2	61	42.0
250	2	170	47.0
462	3	117	66.0
2140	2	190	43.0
2140	1	49	51.0
5750	3	175	65.0
5750	2	169	41.0
5840	3	160	78.0
5840	3	201	83.5
5840	3	117	71.0
5840	2	148	46.0
3954	3	250	94.0
3954	3	246	78.0

(a) *Kira R^2 bagi setiap model berikut* $y = \beta_0 + \beta x_1 + \varepsilon$, $y = \beta_0 + \beta x_2 + \varepsilon$, *dan* $y = \beta_0 + \beta x_3 + \varepsilon$. *Model yang manakah menerangkan kadar tertinggi daripada jumlah variasi dalam cerapan sewa bilik?*

(b) *Model yang manakah di atas menyuai data paling baik? Tentusahkan jawapan anda.*

(c) *Dengan menggunakan model penyuaian terbaik, kirakan selang ramalan 95% untuk sewa bilik individu untuk sebuah hotel yang mempunyai 100 unit, dengan taraf 3 dan terletak di bandar dengan jumlah penduduk 150,000.*

(d) *Dengan menggunakan model penyuaian terbaik, kirakan selang keyakinan 95% bagi purata sewa bilik untuk hotel yang mempunyai 100 unit, dengan taraf 3 dan terletak di bandar dengan jumlah penduduk 150,000.*

[20 markah]

7. For the general linear model, $\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$ where $\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$, the least square estimate of $\boldsymbol{\beta}$ is given as solution(s) of the normal equation, $\left(\mathbf{X}'\mathbf{X}\right)\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$. For a given situation with $\boldsymbol{\beta}' = \left(\beta_1 \quad \beta_2 \quad \beta_3 \quad \beta_4 \quad \beta_5 \quad \beta_6\right)$ it is found that $\hat{\boldsymbol{\beta}}_1' = \left(31 \quad 15 \quad 25 \quad 18 \quad 19 \quad 16\right)$ and $\hat{\boldsymbol{\beta}}_2' = \left(20 \quad 4 \quad 36 \quad 7 \quad 30 \quad 5\right)$ are two solutions of the normal equations.

(a) Are there likely to be more solutions of the normal equations and, if so, how many?

(b) The rank of the design matrix, **X**, is 5, 6 or 7. State, with reasons, which rank should it be.

(c) Exactly one of the following three parametric functions is estimable. Which is it, and why? (i) β_3 (ii) $\beta_3 + \beta_4$ (iii) $\beta_3 - \beta_4$.

[14 marks]

7. *Untuk suatu model linear am,* $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ *di mana* $\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$, *penganggar kuasa dua terkecil untuk* $\boldsymbol{\beta}$ *diberikan sebagai penyelesaian persamaan normal,* $\left(\mathbf{X}'\mathbf{X}\right)\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$. *Dalam keadaan tertentu dengan* $\boldsymbol{\beta}' = \left(\beta_1 \quad \beta_2 \quad \beta_3 \quad \beta_4 \quad \beta_5 \quad \beta_6\right)$, *didapati bahawa* $\hat{\boldsymbol{\beta}}_1' = \left(31 \quad 15 \quad 25 \quad 18 \quad 19 \quad 16\right)$ *dan* $\hat{\boldsymbol{\beta}}_2' = \left(20 \quad 4 \quad 36 \quad 7 \quad 30 \quad 5\right)$ *adalah dua penyelesaian persamaan normal tersebut.*

   (a) *Adakah kemungkinan persaman normal ini mempunyai penyelesaian yang lebih dan jika demikian, berapa banyak?*

   (b) *Pangkat bagi matrik rekabentuk,* ***X*** *adalah 5, 6 atau 7. Terang dengan sebab pangkat yang mana sepatutnya.*

   (c) *Dengan tepat salah satu fungsi berparameter berikut adalah teranggarkan. Yang manakah dan kenapa? (i)* β_3 *(ii)* $\beta_3 + \beta_4$ *(iii)* $\beta_3 - \beta_4$.

[14 markah]

8. Petronas wishes to examine the content of two different RON 95 petrol additives in their six gas stations. They randomly select six gas stations in Penang and randomly assign three gas stations to each additive type (Type 1 and Type 2). The results of the experiment are given in the table below

Table 2

RON 95 Petrol additive (ml per liter petrol)	
Type 1	Type 2
1.51	1.69
1.92	0.64
1.08	0.9

The linear model that was used is $y_{ij} = \mu + \tau_i + \varepsilon_{ij}, i = 1,2 \; j = 1,2,3$ where τ_i is the additive Type *i* and ε_{ij} is the error for the *j*th sample from additive Type *i*.

   (a) Write down the linear model in a matrix form.

   (b) Find two different conditional inverses for $\mathbf{X}'\mathbf{X}$, where $\mathbf{X}$ is the design matrix.

   (c) Under what conditions does a matrix have 0, 1, or an infinite number of conditional inverses?

   (d) Estimate $\tau_1 - \tau_2$.

   (e) Find a 95% confidence interval for $\tau_1 - \tau_2$.

[22 marks]

8. *Petronas ingin memeriksa kandungan dua pemangkin berbeza untuk petrol RON 95 di enam stesen minyak mereka. Mereka memilih secara rawak enam stesen minyak di Pulau Pinang dan membahagi secara rawak tiga stesen minyak untuk setiap jenis bahan pemangkin (Jenis 1 dan Jenis 2). Keputusan eksperimen diberikan dalam jadual di bawah*

Pemangkin petrol RON 95 (ml per liter petrol)	
Jenis 1	*Jenis 2*
1.51	*1.69*
1.92	*0.64*
1.08	*0.9*

Model linear yang digunakan adalah $y_{ij} = \mu + \tau_i + \varepsilon_{ij},\ i = 1,2\ \ j = 1,2,3$ *yang mana* τ_i *adalah pemangkin Jenis i and* ε_{ij} *adalah ralat bagi jth sampel dari pemangkin Jenis i.*

(a) *Tuliskan model linear dalam bentuk matriks.*

(b) *Cari dua songsangan bersyarat berbeza untuk* $\mathbf{X}'\mathbf{X}$, *yang mana* $\mathbf{X}$ *adalah matriks reka bentuk.*

(c) *Dalam keadaan apakah matriks mempunyai 0, 1, atau bilangan tak terhingga songsangan bersyarat?*

(d) *Anggarkan* $\tau_1 - \tau_2$.

(e) *Cari selang keyakinan 95% untuk* $\tau_1 - \tau_2$.

[22 markah]

FORMULAE

$\mathbf{AA}^{-}\mathbf{A} = \mathbf{A}$

$\mathbf{c} = \dfrac{\mathbf{b}}{\sqrt{\mathbf{b}'\mathbf{b}}}$

$\mathbf{Ax} = \lambda \mathbf{x}$

$\mathbf{P}_\rho = \mathbf{D}_\sigma^{-1}\boldsymbol{\Sigma}\mathbf{D}_\sigma^{-1}$, $\boldsymbol{\Sigma} = \mathbf{D}_\sigma \mathbf{P}_\rho \mathbf{D}_\sigma$

$\mathrm{cov}(\mathbf{z}) = \mathrm{cov}(\mathbf{Ay}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$,

$\mathrm{cov}(\mathbf{z},\mathbf{w}) = \mathrm{cov}(\mathbf{Ay},\mathbf{By}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}'$,

$\mathrm{cov}(\mathbf{Ay}+\mathbf{b}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$,

$M_y(t) = E\left(e^{t'y}\right)$

$E(\mathbf{y}\,|\,\mathbf{x}) = \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x}-\boldsymbol{\mu}_x)$,

$E(\mathbf{y}'\mathbf{Ay}) = tr(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$

$\mathrm{cov}(\mathbf{By},\mathbf{y}'\mathbf{Ay}) = 2\mathbf{B}\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\mu}$

$\mathrm{cov}(\mathbf{y}\,|\,\mathbf{x}) = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}$

$r^2 = \dfrac{SSR}{SST} = \dfrac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$

$\hat{\boldsymbol{\beta}}_1 = \left(\dfrac{\mathbf{X}_c'\mathbf{X}_c}{n-1}\right)^{-1}\dfrac{\mathbf{X}_c'\mathbf{y}}{n-1} = \mathbf{S}_{xx}^{-1}\mathbf{s}_{yx}$

$E(y\,|\,\mathbf{x}) = \mu_y + \boldsymbol{\sigma}_{yx}'\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x}-\boldsymbol{\mu}_x)$,

$F = \dfrac{SSR/k}{SSE/(n-k-1)}$

$F = \dfrac{\left(\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \hat{\boldsymbol{\beta}}_1^{*\prime}\mathbf{X}_1'\mathbf{y}\right)/h}{\left(\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}\right)/(n-k-1)}$

$\mathrm{var}(y\,|\,\mathbf{x}) = \sigma_y^2 - \boldsymbol{\sigma}_{yx}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xy}$

$\dfrac{(n-k-1)s^2}{\chi^2_{\alpha/2,n-k-1}} \le \sigma^2 \le \dfrac{(n-k-1)s^2}{\chi^2_{1-\alpha/2,n-k-1}}$

$\mathbf{x}_0'\hat{\boldsymbol{\beta}} \pm t_{\alpha/2,n-k-1}\, s\sqrt{1+\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}$

$\mathrm{var}(\mathbf{y}'\mathbf{Ay}) = 2tr\left[(\mathbf{A}\boldsymbol{\Sigma})^2\right] + 4\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\mu}$

$\hat{\beta}_j \pm t_{\alpha/2,n-k-1}\, s\sqrt{g_{jj}}$

$\mathbf{x}_0'\hat{\boldsymbol{\beta}} \pm t_{\alpha/2,n-k-1}\, s\sqrt{\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}$

$F = \dfrac{(\mathbf{C}\hat{\boldsymbol{\beta}})'\left[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\right]^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}/q}{SSE/(n-k-1)}$

$\mathbf{a}'\hat{\boldsymbol{\beta}} \pm t_{\alpha/2,n-k-1}\, s\sqrt{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}$

$(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \le (k+1)s^2 F_{\alpha,k+1,n-k-1}$

$\mathbf{A}^{1/2} = \mathbf{C}\mathbf{D}^{1/2}\mathbf{C}'$,

$\mathbf{D}_\sigma = \left[diag(\boldsymbol{\Sigma})\right]^{1/2}$

$\mathbf{s}_{yx} = \dfrac{\mathbf{X}_c'\mathbf{y}}{n-1}$

$\boldsymbol{\Sigma} = E\left[(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})'\right] = E(\mathbf{y}\mathbf{y}') - \boldsymbol{\mu}\boldsymbol{\mu}'$

$\mathbf{S}_{xx} = \dfrac{\mathbf{X}_c'\mathbf{X}_c}{n-1}$

$R^2 = \dfrac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{y}^2}{\mathbf{y}'\mathbf{y} - n\bar{y}^2}$

$t_j = \dfrac{\hat{\beta}_j}{s\sqrt{g_{jj}}}$

$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\alpha/2,n-2}\, s\sqrt{1+\dfrac{1}{n}+\dfrac{(x_0-\bar{x})^2}{\sum_{i=1}^n (x_0-\bar{x})^2}}$

$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\alpha/2,n-2}\, s\sqrt{\dfrac{1}{n}+\dfrac{(x_0-\bar{x})^2}{\sum_{i=1}^n (x_i-\bar{x})^2}}$

$F = \dfrac{(R^2 - R_r^2)/h}{(1-R^2)/(n-k-1)}$