UNIVERSITI SAINS MALAYSIA

Second Semester Examination
2014/2015 Academic Session

June 2015

MSG 162 – Applied Statistical Methods
[Kaedah Statistik Gunaan]

Duration : 3 hours
[Masa : 3 jam]

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

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

Instructions: Answer **FOUR** (4) questions.

*[**Arahan**: Jawab **semua empat** (4) 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.(a) A manufacturer frequently sends small packages to a customer in another city via air freight, and in many cases, it is important for a package to reach the customer as soon as possible. Three different firms offer air freight service, including pickup and delivery on a 24-hour basis. The head of the manufacturer's shipping department would like to know whether the firms differ in speed of service at two different periods of the day, morning and afternoon. An experiment is designed to investigate this issue. Packages are sent by the air freight firms at different times of the period. The customer records the time that each package arrives, so that the time elapsed during shipment can be determined. These times are rounded to the nearest hour.

Firm	Morning		Afternoon		Total
A	8, 6, 6, 12, 7, 8	47	7, 10, 8, 11, 9, 11	56	103
B	11, 11, 9, 10, 8, 11	60	10, 13, 10, 12, 11, 10	66	126
C	7, 4, 6, 4, 9, 7	37	10, 8, 6, 5, 8, 6	43	80
Total	144		165		309

$$\sum_{i=1}\sum_{j=1}\sum_{k=1} Y_{ijk}^2 = 2{,}843$$

(i) State a model and its assumptions.
(ii) Determine the effects that influence the response. Test at $\alpha = 0.05$.
(ii) Perform an appropriate follow-up analysis using Duncan multiple range test.

1.(b) A study suggested that coffee drinking may be linked to heart disease. The report states that coffee drinking affects the levels of apolipoprotein B, a cholesterol protein linked to heart disease. An experiment was conducted in which a sample of eight adult men over age 35 who drink from one to five cups of coffee per day were selected and their levels of apolipoprotein B were recorded.

Subject	1	2	3	4	5	6	7	8
Apolipoprotein B Level	37	13	37	28	33	38	29	32
Cups of coffee per day	4	1	5	2	3	5	2	3

(i) State a model and its assumptions.
(ii) Perform an appropriate statistical analysis at 0.01 significance level.

[25 marks]

1.(a) Sebuah kilang sering menghantar bungkusan kecil kepada seorang pelanggan di bandar lain melalui pengangkutan udara, dan dalam beberapa kes, adalah penting suatu bungkusan tiba ke pelanggan secepat mungkin. Tiga firma berlainan menawarkan khidmat pengangkutan udara, termasuk pengambilan dan penghantaran secara 24-jam. Ketua jabatan penghantaran bagi kilang tersebut ingin mengetahui sama ada firma tersebut berbeza dalam kepantasan perkhidmatan pada dua tempoh berlainan bagi suatu hari, pagi dan petang. Suatu ujikaji direka bentuk untuk mengkaji isu ini. Bungkusan dihantar oleh firma pengangkutan udara tersebut pada waktu berlainan bagi suatu tempoh hari. Pelanggan mencatat masa setiap bungkusan tiba, supaya masa berlalu semasa penghantaran dapat ditentukan. Masa tersebut dibundar kepada jam terdekat.

Firma	*Pagi*		*Petang*		*Jumlah*
A	*8, 6, 6, 12, 7, 8*	*47*	*7, 10, 8, 11, 9, 11*	*56*	*103*
B	*11, 11, 9, 10, 8, 11*	*60*	*10, 13, 10, 12, 11, 10*	*66*	*126*
C	*7, 4, 6, 4, 9, 7*	*37*	*10, 8, 6, 5, 8, 6*	*43*	*80*
Jumlah	*144*		*165*		*309*

$$\sum_{i=1}\sum_{j=1}\sum_{k=1} Y_{ijk}^2 = 2{,}843$$

(i) Nyatakan suatu model serta anggapannya.
(ii) Tentukan kesan yang mempengaruhi respon. Uji pada $\alpha = 0.05$.
(iii) Lakukan suatu analisis lanjutan yang sesuai menggunakan ujian julat berganda Duncan.

1.(b) Suatu kajian mencadangkan bahawa meminum kopi mungkin dikaitkan dengan penyakit jantung. Laporan tersebut menyatakan bahawa meminum kopi memberi kesan kepada tahap apolipoprotein B, suatu protein kolesterol yang dikaitkan dengan penyakit jantung. Suatu ujikaji dilakukan iaitu suatu sampel lapan lelaki dewasa berumur 35 tahun ke atas yang minum satu hingga lima cawan kopi sehari dipilih dan tahap apolipoprotein B mereka dicatat.

Subjek	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Aras apolipoprotein B	*37*	*13*	*37*	*28*	*33*	*38*	*29*	*32*
Cawan kopi sehari	*4*	*1*	*5*	*2*	*3*	*5*	*2*	*3*

(i) Nyatakan suatu model serta anggapannya.
(ii) Lakukan suatu analisis statistik yang sesuai pada aras signifikan 0.01.

[25 markah]

2.(a) Biscuits lose their crispness in damp climates unless they are packaged in containersthat protect them from humidity. A bakery firm wishes to investigate two different methods of packaging, a metal foil and a plastic wrapping material, including a cardboard box control. Six boxes are wrapped with each method of packaging, and placed in a chamber in which the humidity is maintained at 80% for 24 hours. The boxes are opened and a biscuit were selected at random from each box to be measured for moisture content. The measurements on the biscuits are given in milligrams.

Box	Metal Foil	Plastic	Cardboard	Metal Foil and Plastic	Total
1	46	60	73	38	217
2	49	66	75	36	226
3	46	60	77	40	223
4	59	52	67	50	228
5	53	55	69	47	224
6	53	49	62	44	208
Total	306	342	423	255	1,326

$$\sum_{i=1}\sum_{j=1} Y_{ij}^2 = 76{,}380$$

(i) State a model and its assumptions.
(ii) Perform an appropriate analysis at $\alpha = 0.01$.
(iii) Perform a follow-up analysis using Tukey HSD procedure.
(iv) Determine the following effects on the response:
  a. the metal foil packaging effect
  b. the combined effect of different methods of packaging

2.(b) The use of preservatives by food processors has become a controversial issue. Two preservatives, A and B, are extensively tested and determined safe for use in meats. A processor wants to compare the preservatives for their effects on retarding spoilage. Seven cuts of fresh meat were used in an experiment, three pieces from each cut were randomly assigned to the different preservatives, and the number of hours until spoilage begins were recorded.

	Cuts of fresh meat							
Preservative	1	2	3	4	5	6	7	Total
None	40	44	41	37	36	38	35	271
A	103	102	103	99	85	101	100	693
B	87	98	111	120	85	83	90	674
Total	230	244	255	256	206	222	225	1,638

$$\sum_{i=1}\sum_{j=1} Y_{ij}^2 = 145{,}508$$

(i) State a model and its assumptions.
(ii) Perform an appropriate statistical analysis at $\alpha = 0.01$.
(iii) Perform a follow-up analysis.

[30 marks]

2.(a) Biskut hilang kerangupan di dalam iklim lembap kecuali ia dibungkus di dalam bekas yang melindunginya daripada kelembapan. Suatu firma membuat roti ingin mengkaji dua kaedah pembungkusan berlainan, suatu bahan pembungkus kerajang logam dan plastik, termasuk kawalan kotak kadbod. Enam kotak dibungkus dengan setiap kaedah pembungkusan dan diletak di dalam suatu ruang yang kelembapan dikekalkan pada 80% selama 24 jam. Kotak dibuka dan satu biskut dipilih secara rawak daripada setiap kotak untuk diukur kandungan kelembapan. Ukuran pada biskut diberikan dalam miligram.

Kotak	*Kerajang Logam*	*Plastik*	*Kadbod*	*Kerajang Logam dan Plastik*	*Jumlah*
1	*46*	*60*	*73*	*38*	*217*
2	*49*	*66*	*75*	*36*	*226*
3	*46*	*60*	*77*	*40*	*223*
4	*59*	*52*	*67*	*50*	*228*
5	*53*	*55*	*69*	*47*	*224*
6	*53*	*49*	*62*	*44*	*208*
Jumlah	*306*	*342*	*423*	*255*	*1,326*

$$\sum_{i=1}\sum_{j=1} Y_{ij}^2 = 76{,}380$$

(i) Nyatakan suatu model serta anggapannya.
(ii) Lakukan suatu analisis statistik sesuai pada $\alpha = 0.01$.
(iii) Lakukan suatu analisis lanjutan menggunakan tatacara Tukey HSD.
(iv) Tentukan kesan berikut keatas respon:
a. kesan pembungkusan kerajang logam
b. kesan gabungan bagi kaedah pembungkusan berlainan

2.(b) Penggunaan bahan pengawet oleh pemproses makanan telah menjadi suatu isu kontroversi. Dua bahan pengawet, A dan B, diuji secara meluas dan ditentukan selamat untuk digunakan di dalam daging. Suatu syarikat pemproses ingin membandingkan pengawet tersebut bagi kesan memperlahankan kerosakan. Tujuh potongan daging segar digunakan dalam suatu ujikaji, tiga keping daripada setiap potongan diumpuk secara rawak kepada pengawet yang berlainan, dan bilangan jam sehingga kerosakan mula dicatat.

	Potongan daging segar							
Bahan pengawet	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*Jumlah*
Tiada	*40*	*44*	*41*	*37*	*36*	*38*	*35*	*271*
A	*103*	*102*	*103*	*99*	*85*	*101*	*100*	*693*
B	*87*	*98*	*111*	*120*	*85*	*83*	*90*	*674*
Jumlah	*230*	*244*	*255*	*256*	*206*	*222*	*225*	*1,638*

$$\sum_{i=1}\sum_{j=1} Y_{ij}^2 = 145{,}508$$

(i) Nyatakan suatu model serta anggapannya.
(ii) Lakukan suatu analisis statistik yang sesuai pada $\alpha = 0.01$.
(iii) Lakukan suatu analisis lanjutan.

[30 markah]

3.(a) A power plant which uses water from the surrounding bay for cooling its condensers is required by the environment protection agency (EPA) to determine whether discharging its heated water into the bay has a detrimental effect on the flora (plant life) in the water. The EPA requests that the power plant makes its investigation at three strategically chosen locations. Location A and B are located near the plant's discharge tubes while location C is located farther out in the bay. During one randomly selected day in each of six months, a diver descends to each of the locations, randomly samples a square meter area of the bottom, and counts the number of blades of the different types of grasses present.

	Month						
Location	March	April	May	June	July	August	Total
A	28	25	37	20	27	28	165
B	31	22	30	26	29	25	163
C	53	61	56	48	54	55	327
Total	112	108	123	94	110	108	655

$$\sum_i \sum_j Y_{ij}^2 = 27{,}089$$

(i) State a model and its assumptions.
(ii) Perform an appropriate statistical analysis at $\alpha = 0.10$.
(iii) Perform a follow-up analysis.

3.(b) A study was designed to evaluate the effectiveness of different chemicals developed to control fire ants. The type of environmental conditions in which the chemical is placed might have an effect on the effectiveness of the treatment to kill fire ants. Thus the researcher randomly assigned five locations to each chemical chosen at random. To reduce the effect of different colonies of fire ants and the type of mound they inhibit, the researcher created artificial fire ant mounds and populated them with 50,000 ants having similar ancestry. The number of fire ants killed (in thousands) during a 1-week period was recorded.

	Location					
Chemical	1	2	3	4	5	Total
A	7.2	8.5	9.1	8.2	7.8	40.8
B	9.5	8.8	7.6	7.3	9.2	42.4
C	5.4	6.3	6.1	5.0	6.5	29.3
Total	22.1	23.6	22.8	20.5	23.5	112.5

$$\sum_{i=1} \sum_{j=1} Y_{ij}^2 = 871.670$$

(i) State a model and its assumptions.
(ii) Perform an appropriate analysis at $\alpha = 0.05$.
(iii) Perform a follow-up analysis.

[25 marks]

3.(a) Suatu loji kuasa yang menggunakan air daripada teluk sekitarnya untuk menyejukkan kondensernya dikehendaki oleh agensi perlindungan persekitaran (EPA) untuk menentukan sama ada melepaskan air panasnya ke dalam teluk memberi kesan mudarat ke atas flora (kehidupan tumbuhan) di dalam air. Pihak EPA mengarahkan loji kuasa tersebut membuat penyelidikan pada tiga lokasi yang dipilih secara strategik. Lokasi A dan B berdekatan dengan pelepasan loji manakala lokasi C berada jauh keluar tetapi masih dalam kawasan teluk. Pada satu hari yang dipilih secara rawak dalam setiap enam bulan, seorang penyelam meyelam ke setiap lokasi, mengambil secara rawak sampel satu meter persegi bahagian bawah, dan mengira bilangan dedaun bagi berlainan jenis rumpai yang ada.

	Bulan						
Lokasi	*Mac*	*April*	*Mei*	*Jun*	*Julai*	*Ogos*	*Jumlah*
A	*28*	*25*	*37*	*20*	*27*	*28*	*165*
B	*31*	*22*	*30*	*26*	*29*	*25*	*163*
C	*53*	*61*	*56*	*48*	*54*	*55*	*327*
Jumlah	*112*	*108*	*123*	*94*	*110*	*108*	*655*

$$\sum_i \sum_j Y_{ij}^2 = 27{,}089$$

(i) Nyatakan suatu model serta anggapannya.
(ii) Lakukan suatu analisis statistik yang sesuai pada $\alpha = 0.10$.
(iii) Lakukan suatu analisis lanjutan.

3.(b) Suatu kajian dirangka untuk menilai keberkesanan bahan kimia berlainan yang dirumus untuk mengawal semut api. Jenis keadaan persekitaran untuk diletakkan bahan kimia tersebut mungkin mempunyai kesan keatas keberkesanan rawatan untuk menghapus semut api. Oleh itu penyelidik mengumpukkan secara rawak lima lokasi kepada setiap bahan kimia yang dipilih secara rawak. Untuk mengurangkan kesan koloni semut api yang berlainan dan jenis busut yang mereka kerumuni, penyelidik mencipta busut semut api tiruan dan diisi dengan 50,000 ekor semut yang berketurunan serupa. Bilangan semut api yang mati (dalam ribu) semasa suatu tempoh 1-minggu dicatat.

	Lokasi					
Bahan kimia	*1*	*2*	*3*	*4*	*5*	*Jumlah*
A	*7.2*	*8.5*	*9.1*	*8.2*	*7.8*	*40.8*
B	*9.5*	*8.8*	*7.6*	*7.3*	*9.2*	*42.4*
C	*5.4*	*6.3*	*6.1*	*5.0*	*6.5*	*29.3*
Jumlah	*22.1*	*23.6*	*22.8*	*20.5*	*23.5*	*112.5*

$$\sum_{i=1} \sum_{j=1} Y_{ij}^2 = 871.670$$

(i) Nyatakan suatu model serta anggapannya.
(ii) Lakukan suatu analisis statistik yang sesuai pada $\alpha = 0.05$.
(iii) Lakukan suatu analisis lanjutan.

[25 markah]

4. A psychologist is interested in examining the effects of sleep deprivation on a person's ability to perform simple arithmetic tasks. To do this, prospective subjects are screened to obtain individuals whose daily sleep patterns were closely matched. Twenty subjects are randomly chosen and each individual selected is randomly assigned to each of five different groups of sleep deprivation: 0, 2, 4, 6 and 8 hours of sleep.

All subjects are then placed on a standard routine for the next 24 hours. The following day after breakfast, each individual is tested to determine the number of arithmetic additions done correctly in a 10-minute period. That evening the amount of sleep each person is allowed depends on the group to which he or she had been assigned. The following morning after breakfast, each person is again tested using a different but equally difficult set of additions. The difference in the number of correct answers between the first test day and the second test day were recorded.

Group	Sleep (in hours)	Difference in correct answers
1	0	39
	0	33
	0	41
	0	40
2	2	25
	2	29
	2	34
	2	26
3	4	10
	4	18
	4	14
	4	17
4	6	4
	6	6
	6	-1
	6	9
5	8	-5
	8	0
	8	-3
	8	-8
Total	80	328

$\sum(\text{Sleep}) \text{ x } (\text{Difference in correct answers}) = 444$

$\sum(\text{Sleep})^2 = 480$

$\sum(\text{Difference in correct answers})^2 = 10{,}330$

(i) State a model and its assumptions.
(ii) Perform an appropriate statistical analysis at 0.01 significance level.
(iii) Obtain an estimation interval and prediction interval for an individual having three hours of sleep.

[20 marks]

4. *Seorang ahli psikologi berminat untuk mengkaji kesan kurang tidur kepada keupayaan seseorang untuk melakukan pengiraan aritmetik mudah. Untuk melakukan ini, bakal subjek ditapis untuk mendapatkan individu yang mempunyai sepadan corak tidur setiap hari. Dua puluh subjek dipilih secara rawak dan setiap orang yang dipilih diumpukkan secara rawak kepada setiap lima kumpulan kurang tidur berlainan: 0, 2, 4, 6 dan 8 jam tidur. Kesemua subjek kemudiannya ditempatkan pada suatu rutin piawai bagi 24 jam berikutnya. Pada hari berikutnya selepas sarapan pagi, setiap individu diuji untuk menentukan bilangan penambahan aritmetik yang dilakukan dengan betul dalam tempoh 10 minit. Pada petangnya jumlah tidur yang dibenarkan bagi setiap orang bersandar kepada kumpulan yang dia diumpukkan. Pada esok pagi selepas sarapan pagi, setiap orang sekali lagi diuji menggunakan set penambahan yang berbeza tetapi sama sukar. Perbezaan bilangan jawapan yang betul antara hari ujian pertama dan hari kedua ujian dicatat.*

Kumpulan	*Tidur (dalam jam)*	*Perbezaan jawapan betul*
1	*0*	*39*
	0	*33*
	0	*41*
	0	*40*
2	*2*	*25*
	2	*29*
	2	*34*
	2	*26*
3	*4*	*10*
	4	*18*
	4	*14*
	4	*17*
4	*6*	*4*
	6	*6*
	6	*-1*
	6	*9*
5	*8*	*-5*
	8	*0*
	8	*-3*
	8	*-8*
Jumlah	*80*	*328*

$$\sum(Tidur) \text{ x } (Perbezaan\ jawapan\ betul) = 444$$

$$\sum(Tidur)^2 = 480$$

$$\sum(Perbezaan\ jawapan\ betul)^2 = 10{,}330$$

(i) Nyatakan suatu model serta anggapannya.
(ii) Lakukan suatu analisis statistik yang sesuai pada aras signigikan 0.01.
(iii) Dapatkan suatu selang anggaran dan selang ramalan bagi seorang individu yang tidur tiga jam.

[20 markah]

APPENDIX: FORMULAS

1. Completely Randomized Design

$$SST = \sum_i \sum_j Y_{ij}^2 - \frac{Y_{..}^2}{N}$$

$$SSA = \sum_i \frac{Y_{i.}^2}{n_i} - \frac{Y_{..}^2}{N}$$

For any contrast : $L = \sum_i c_i \bar{Y}_{i.}$

$$SSL = \frac{\left(\sum_i c_i \bar{Y}_{i.} \right)^2}{\sum_i \frac{c_{i.}^2}{n_i}}$$

2. Completely Randomized Block Design

$$SST = \sum_i \sum_j Y_{ij}^2 - \frac{Y_{..}^2}{N}$$

$$SSA = \sum_i \frac{Y_{i.}^2}{b} - \frac{Y_{..}^2}{N}$$

$$SSB = \sum_j \frac{Y_{.j}^2}{a} - \frac{Y_{..}^2}{N}$$

3. Latin Square Design

$$SST = \sum_i \sum_j \sum_k Y_{ijk}^2 - \frac{Y_{...}^2}{N}$$

$$SSR = \sum_i \frac{Y_{i..}^2}{a} - \frac{Y_{...}^2}{N}$$

$$SSC = \sum_j \frac{Y_{.j.}^2}{a} - \frac{Y_{...}^2}{N}$$

$$SSA = \sum_k \frac{Y_{..k}^2}{a} - \frac{Y_{...}^2}{N}$$

4. Two-way Factorial Design

$$SST = \sum_i \sum_j \sum_k Y_{ijk}^2 - \frac{Y_{...}^2}{N}$$

$$SSA = \sum_i \frac{Y_{i..}^2}{bn} - \frac{Y_{...}^2}{N}$$

$$SSB = \sum_j \frac{Y_{.j.}^2}{an} - \frac{Y_{...}^2}{N}$$

$$SSE = \sum_i \sum_j \sum_k Y_{ijk}^2 - \frac{Y_{ij.}^2}{n}$$

5. Multiple Comparison Procedures:

Duncan: $r_{\alpha, p, df}$, p = range df = degrees of freedom

Tukey: $\frac{1}{\sqrt{2}} q(\alpha, a, df)$, a = number of treatments df = degrees of freedom

Scheffe': $\sqrt{(a-1) F_{\alpha, a-1, df}}$, a = number of treatments df = degrees of freedom

6. Regression

$$b_1 = \frac{SS_{XY}}{SS_X} \quad , \quad b_o = \bar{Y} - b_1 \bar{X}$$

$$SSE = SS_Y - \frac{\left[SS_{XY}\right]^2}{SS_X}$$

$$SS_{XY} = \sum X_i Y_i - \frac{\left(\sum X_i\right)\left(\sum Y_i\right)}{n}$$

$$SS_X = \sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n}$$

$$SS_Y = \sum Y_i^2 - \frac{\left(\sum Y_i\right)^2}{n}$$

$$Var(b_1) = \frac{\sigma^2}{SS_X}$$

$$Var\left(b_o\right) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{SS_X}\right]$$

$$Var(\hat{Y}_h) = \sigma^2 \left[\frac{1}{n} + \frac{\left(X_h - \bar{X}\right)^2}{SS_X}\right]$$

…13/-

7. Correlation

$$r = \frac{SS_{XY}}{\sqrt{SS_{XX}\,SS_{YY}}}$$

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

8. Kruskal-Wallis Test

$$T = \frac{12}{N(N+1)}\sum_i \frac{R_i^2}{n_i} - 3(N+1)$$

9. Friedman Test

$$T = \frac{12}{ab(a+1)}\sum_i R_i^2 - 3b(a+1)$$

10. Spearman Test

$$r_s = 1 - \frac{6\sum_{i=1}^{n}\left[R(X_i) - R(Y_i)\right]^2}{n(n^2-1)} = 1 - \frac{6\sum_i d_i^2}{n(n^2-1)}$$

APPENDIX: TABLES

Duncan Multiple Range Table

Spearman Table

Turkey HSD table