UNIVERSITI SAINS MALAYSIA

Second Semester Examination
2015/2016 Academic Session

June 2016

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 large food products company conducted an experiment to investigate the effect of package wrapper material on sales of one of the company's products. Three types of wrapping material were employed, a waxed paper, plastic and metal foil in three colors. Three supermarkets were randomly assigned to each of the different wrapper material. After the product had been in the supermarkets for one week, the company recorded the percentage change in weekly sales.

Package Color

Material	No color		Yellow		Red		Total
Waxed paper	6 -2 4	8	5 2 5	12	7 -3 8	12	32
Plastic	-3 7 -2	2	3 6 4	13	-4 3 5	4	19
Metal Foil	7 4 10	21	12 8 10	30	10 5 14	29	80
Total	31		55		45		131

$$\sum_{i=1}\sum_{j=1}\sum_{k=1} Y_{ijk}^2 = 1{,}199$$

(i) State a model and its assumptions.
(ii) Determine the effects that influence the response. Test at $\alpha = 0.01$.
(iii) Perform an appropriate follow-up analysis using Duncan multiple range test.

1.(b) A study was conducted to determine automobile gasoline mileage for different formulations of gasoline. Three formulations were selected and automobiles of the same make and model were used in this experiment. Four automobiles were randomly selected and assigned to each of the formulation. The data, in miles per gallon were recorded.

	Automobile				
Formulation	1	2	3	4	Total
Non-leaded 87 octane	25.7	27.0	27.3	26.1	106.1
Non-leaded 91 octane	27.2	28.1	27.9	27.7	110.9
Non-leaded 87 octane with 15% ethanol	26.1	27.5	26.8	27.8	108.2
Total	79.0	82.6	82.0	81.6	325.2

$$\sum_{i=1}\sum_{j=1} Y_{ij}^2 = 8{,}819.68$$

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

[25 marks]

1.(a) Sebuah syarikat besar produk makanan melakukan suatu eksperimen untuk mengkaji kesan bahan pakej pembungkus ke atas jualan salah satu produk syarikat. Tiga jenis bahan bungkusan digunakan, kertas lilin, plastik dan logam kerajang dalam tiga warna. Tiga pasar raya diumpukkan secara rawak kepada setiap bahan bungkusan yang berbeza. Selepas produk berada di pasar raya selama seminggu, syarikat itu mencatatkan perubahan peratusan dalam jualan seminggu.

Warna pakej

Bahan	*Tiada warna*		*Kuning*		*Merah*		*Jumlah*
Waxed paper	*6* *-2* *4*	*8*	*5* *2* *5*	*12*	*7* *-3* *8*	*12*	*32*
Plastik	*-3* *7* *-2*	*2*	*3* *6* *4*	*13*	*-4* *3* *5*	*4*	*19*
Logam kerajang	*7* *4* *10*	*21*	*12* *8* *10*	*30*	*10* *5* *14*	*29*	*80*
Jumlah	*31*		*55*		*45*		*131*

$$\sum_{i=1}\sum_{j=1}\sum_{k=1} Y_{ijk}^2 = 1{,}199$$

(i) Nyatakan suatu model serta anggapannya.
(ii) Tentukan kesan yang mempengaruhi respon. Uji pada $\alpha = 0.01$.
(iii) Lakukan suatu analisis lanjutan menggunakan ujian julat berganda Duncan.

1.(b) Satu kajian telah dilakukan untuk menentukan penggunaan petrol kereta bagi formulasi petrol yang berbeza. Tiga rumusan telah dipilih dan kereta daripada buatan dan model yang sama telah digunakan dalam eksperimen ini. Empat kereta telah dipilih secara rawak dan diumpukkan kepada setiap formulasi. Data, dalam batu segelen telah direkodkan.

Kereta

Formulasi	*1*	*2*	*3*	*4*	*Jumlah*
Tanpa plumbum 87 oktan	*25.7*	*27.0*	*27.3*	*26.1*	*106.1*
Tanpa plumbum 91 oktan	*27.2*	*28.1*	*27.9*	*27.7*	*110.9*
Tanpa plumbum 87 oktan dengan 15% ethanol	*26.1*	*27.5*	*26.8*	*27.8*	*108.2*
Jumlah	*79.0*	*82.6*	*82.0*	*81.6*	*325.2*

$$\sum_{i=1}\sum_{j=1} Y_{ij}^2 = 8{,}819.68$$

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

[25 markah]

2.(a) In screening for compounds useful in treating hypertension, researchers assign three rats to each of three groups. The rats in group 1 receive 0.2 milligram per kilogram of a test compound, those in group 2 and 3 receive 0.2 and 0.4 milligram per kilogram respectively. The decrease in blood pressure two hours post-dose, compared to the corresponding pre-dose blood pressure was recorded.

Group	Dose (in mg / kg)	Blood pressure drop
1	0.1	12
1	0.1	10
1	0.1	13
2	0.2	24
2	0.2	25
2	0.2	22
3	0.4	49
3	0.4	30
3	0.4	32

(i) State a model and its assumptions.
(ii) Perform an appropriate statistical analysis at 0.05 significance level.

2.(b) An entomologist is investigating the effectiveness of two types of fumigants with different concentrations in controlling parasites in citrus plants.

Treatment I: Low concentration of fumigant A
Treatment II: Low concentration of both fumigants
Treatment III: High concentration of fumigant A
Treatment IV: High concentration of fumigant B

To evaluate the fumigants, six fields of differing soil characteristics, drainage and amount of wind shield were planted with citrus. The fumigants were randomly assigned to a plot in each field. One hundred plants were randomly selected from each field, 25 from each plot, and the number of parasites were counted.

	Field						
Treatment	1	2	3	4	5	6	Total
I	35	23	26	33	28	31	176
II	70	53	47	77	77	71	395
III	77	50	21	41	38	60	287
IV	76	48	20	39	37	58	278
Total	258	174	114	190	180	220	1,136

$$\sum_{i=1}\sum_{j=1} Y_{ij}^2 = 62{,}410$$

(i) State an appropriate model and its assumptions.
(ii) Perform an appropriate analysis at $\alpha = 0.05$.
(iii) Perform a follow-up analysis using Tukey HSD procedure.
(iv) Evaluate the following effects on the response:
  a. the combination effect of both fumigant
  b. the concentration effect

[30 marks]

2.(a) Dalam saringan sebatian yang berguna dalam merawat tekanan darah tinggi, penyelidik mengumpukkan tiga tikus untuk setiap tiga kumpulan. Tikus dalam kumpulan 1 menerima 0.2 miligram per kilogram sebatian ujian, kumpulan 2 dan 3 menerima 0.2 dan 0.4 miligram per kilogram masing-masing. Penurunan tekanan darah dua jam selepas dos, berbanding tekanan darah pra-dos yang berpadan telah direkodkan.

Kumpulan	*Dos (dalam mg / kg)*	*Penurunan tekanan darah*
1	*0.1*	*12*
1	*0.1*	*10*
1	*0.1*	*13*
2	*0.2*	*24*
2	*0.2*	*25*
2	*0.2*	*22*
3	*0.4*	*49*
3	*0.4*	*30*
3	*0.4*	*32*

(i) Nyatakan suatu model serta anggapannya.
(ii) Lakukan suatu analisis statistik yang sesuai pada aras signifikan 0.05.

2.(b) Seorang ahli entomologi sedang menyiasat keberkesanan dua jenis fumigan dengan kepekatan yang berbeza dalam mengawal parasit dalam tumbuhan sitrus.

Rawatan I: Kepekatan rendah fumigan A
Rawatan II: Kepekatan rendah kedua-dua fumigann
Rawatan III: Kepekatan tinggi fumigan A
Rawatan IV: Kepekatan tinggi fumigan B

Untuk menilai fumigan tersebut, enam ladang yang berbeza ciri tanah, saliran dan jumlah perlindungan angin telah ditanam dengan limau. Fumigan tersebut diumpukkan secara rawak kepada setiap plot dalam setiapladang.Seratus tumbuhan telah dipilih secara rawak daripada setiap ladang, 25 dari setiap plot, dan bilangan parasit telah dikira.

	Ladang						
Rawatan	*1*	*2*	*3*	*4*	*5*	*6*	*Jumlah*
I	*35*	*23*	*26*	*33*	*28*	*31*	*176*
II	*70*	*53*	*47*	*77*	*77*	*71*	*395*
III	*77*	*50*	*21*	*41*	*38*	*60*	*287*
IV	*76*	*48*	*20*	*39*	*37*	*58*	*278*
Jumlah	*258*	*174*	*114*	*190*	*180*	*220*	*1,136*

$$\sum_{i=1}\sum_{j=1} Y_{ij}^2 = 62{,}410$$

(i) Nyatakan suatu model serta anggapannya.
(ii) Lakukan suatu analisis yang sesuai pada $\alpha = 0.05$.
(iii) Lakukan suatu analisis lanjutan menggunakan tatacara Tukey HSD.
(iv) Nilaikan kesan berikut ke atas respon:
a. kesan gabungan kedua-dua fumigan
b. kesan kepekatan

[30 markah]

3.(a) A large body of evidence shows that soy has health benefits for most people. Some of these benefits come largely from isoflavones, plant compounds that have estrogen-like properties. The amount of isoflavones varies widely depending on the type of food processing. A consumer group purchased various soy products and conducted laboratory tests to determine the amount of isoflavones in each product. Three major categories of soy products: cereals and snacks, energy bars and veggie burgers were considered. Five different products from each of the three categories were randomly selected and the amount of isoflavones (in mg) was determined for an adult serving of the product.

Product	Cereals and snacks	Energy bars	Veggie burgers	Total
1	5	36	35	76
2	17	19	49	85
3	35	10	25	70
4	12	9	22	43
5	10	7	19	36
6	4	5	18	27
Total	83	86	168	337

$$\sum_{i=1}\sum_{j=1} Y_{ij}^2 = 9{,}131$$

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

3.(b) Various agents are used to control weeds in crops. Of particular concern is the overusage of chemical agents. Although effective in controlling weeds, these agents may also drain into the underground water system and cause health problem. Thus, several new biological weed agents have been proposed to eliminate the contamination problem present in chemical agents. Researchers conducted a study of biological agents to assess their effectiveness in comparison to the chemical weed agents. This study consisted of two biological agents (Bio1 and Bio2), and a commonly used chemical agent. One-acre plots of land were planted with hay and each weed agent were applied to five plots. The hay was harvested and total yield in tons per acre was recorded.

	Plot					
Agent type	1	2	3	4	5	Total
Bio1	1.40	1.75	1.38	1.65	1.55	7.73
Bio2	1.70	1.85	1.96	2.05	1.80	9.36
Chemical	2.10	1.95	1.65	1.88	2.00	9.58
Total	5.20	5.55	4.99	5.58	5.35	26.67

$$\sum_{i=1}\sum_{j=1} Y_{ij}^2 = 48.1179$$

(i) State an appropriate model and its assumptions.
(ii) Perform an appropriate analysis at $\alpha = 0.01$.
(iii) Perform a follow-up analysis. Which biological agent would be a suitable alternative to the chemical agent?

[25 marks]

3.(a) Banyak bukti menunjukkan bahawa soya mempunyai manfaat kesihatan bagi kebanyakan orang. Sebahagian besar daripada manfaat ini didapati daripada isoflavon, sebatian tumbuhan yang mempunyai ciri seperti estrogen. Jumlah isoflavon berbeza bersandar kepada jenis pemprosesan makanan. Sekumpulan pengguna membeli pelbagai produk soya dan melakukan ujian makmal untuk menentukan jumlah isoflavon dalam setiap produk. Tiga kategori utama produk soya: bijirin dan snek, bar bertenaga dan burger sayuran dipertimbangkan. Lima produk berbeza daripada setiap satu tiga kategori telah dipilih secara rawak dan jumlah isoflavon (dalam mg) telah ditentukan bagi hidangan dewasa sesuatu produk.

Produk	*Bijirin dan snek*	*Bar bertenaga*	*Burger sayuran*	*Jumlah*
1	*5*	*36*	*35*	*76*
2	*17*	*19*	*49*	*85*
3	*35*	*10*	*25*	*70*
4	*12*	*9*	*22*	*43*
5	*10*	*7*	*19*	*36*
6	*4*	*5*	*18*	*27*
Jumlah	*83*	*86*	*168*	*337*

$$\sum_{i=1}\sum_{j=1} Y_{ij}^2 = 9{,}131$$

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

3.(b) Pelbagai agen digunakan untuk mengawal rumpai dalam tanaman. Suatu kebimbangan tertentu adalah terlalu berlebihan penggunaan agen kimia. Walaupun berkesan dalam mengawal rumpai, agen ini juga boleh mengalir ke sistem sistem air bawah tanah dan menyebabkan masalah kesihatan. Oleh itu, beberapa agen rumpai biologi baru telah dicadangkan untuk menghapuskan masalah pencemaran yang wujud di dalam agen kimia. Penyelidik melakukan suatu kajian agen biologi untuk menilai keberkesanannya berbanding dengan agen kimia rumpai. Kajian ini terdiri daripada dua agen biologi (Bio 1 dan Bio 2), dan suatu agen kimia yang biasa digunakan. plot satu ekar tanah telah ditanam dengan rumput kering dan setiap agen rumpai telah digunakan ke atas lima plot. Rumput kering itu telah dituai dan jumlah hasil dalam tan seekar direkodkan.

	Plot					
Jenis agen	*1*	*2*	*3*	*4*	*5*	*Jumlah*
Bio1	*1.40*	*1.75*	*1.38*	*1.65*	*1.55*	*7.73*
Bio2	*1.70*	*1.85*	*1.96*	*2.05*	*1.80*	*9.36*
Kimia	*2.10*	*1.95*	*1.65*	*1.88*	*2.00*	*9.58*
Jumlah	*5.20*	*5.55*	*4.99*	*5.58*	*5.35*	*26.67*

$$\sum_{i=1}\sum_{j=1} Y_{ij}^2 = 48.1179$$

(i) Nyatakan suatu model serta anggapannya.
(ii) Lakukan suatu analisis yang sesuai pada $\alpha = 0.01$.
(iii) Lakukan suatu analisis lanjutan. Agen biologi yang manakah akan menjadi alternatif sesuai kepada agen kimia?

[25 markah]

...8/-

4. Athletes are constantly seeking measures of the degree of their cardiovascular fitness prior to a major race. They wished to know when their training is at a level which will produce a peak performance. One such measure of fitness is the time to exhaustion from running on a treadmill at a specified angle and speed. The important question is "Does this measure of cardiovascular fitness translate into performance in a 10-km running race?" Twenty experienced distance runners who professed to be at top condition were evaluated on the treadmill and then had their times (in minutes) recorded in a 10-km race.

10-km time	7.5	7.8	7.9	8.1	8.3	8.7	8.9	9.2	9.4	9.8
Treadmill time	43.5	45.2	44.9	41.1	43.8	44.4	42.7	43.1	41.8	43.7

10-km time	10.1	10.3	10.5	10.7	10.8	10.9	11.2	11.5	11.7	11.8
Treadmill time	39.5	38.2	43.9	37.1	37.7	39.2	35.7	37.2	34.8	38.5

$\sum(\text{10-km time}) = 195.10$ $\sum(\text{Treadmill time}) = 816.00$

$\sum(\text{10-km time})^2 = 1{,}940.05$ $\sum(\text{Treadmill time})^2 = 33{,}500.80$

$\sum(\text{10-km time x Treadmill time}) = 7{,}887.85$

(i) State a model and its assumptions.
(ii) Perform an appropriate statistical analysis at 0.05 significance level.
(iii) Obtain an estimation interval and prediction interval for the amount of time performed in a 10-km running race if the time to exhaustion from running on a treadmill is 10.6 minutes.

[20 marks]

4. *Para atlet sentiasa mengukur tahap kecergasan kardiovaskular mereka sebelum suatu larian utama. Mereka ingin mengetahui masa latihan yang berada pada suatu tahap yang akan menghasilkan prestasi tinggi. Salah satu ukuran kecergasan adalah masa sehingga keletihan berlari di atas suatu treadmil pada suatu sudut dan kelajuan yang tertentu. Suatu persoalan penting ialah "Adakah ukuran kecergasan kardiovaskular ini dapat diterjemahkan kepada prestasi dalam perlumbaan berlari 10-km?" Dua puluh pelari jarak jauh berpengalaman yang mengaku berada pada keadaan tahap terbaik telah dinilai di atas suatu treadmil dan kemudiannya masa dalam suatu larian 10km (dalam minit) mereka dicatat.*

Masa 10km	*7.5*	*7.8*	*7.9*	*8.1*	*8.3*	*8.7*	*8.9*	*9.2*	*9.4*	*9.8*
Masa treadmil	*43.5*	*45.2*	*44.9*	*41.1*	*43.8*	*44.4*	*42.7*	*43.1*	*41.8*	*43.7*

Masa 10km	*10.1*	*10.3*	*10.5*	*10.7*	*10.8*	*10.9*	*11.2*	*11.5*	*11.7*	*11.8*
Masa treadmil	*39.5*	*38.2*	*43.9*	*37.1*	*37.7*	*39.2*	*35.7*	*37.2*	*34.8*	*38.5*

$\sum(\textit{Masa 10-km}) = 195.10$ $\sum(\textit{Masa Treadmil}) = 816.00$

$\sum(\textit{Masa 10-km})^2 = 1{,}940.05$ $\sum(\textit{Masa Treadmil})^2 = 33{,}500.80$

$\sum(\textit{Masa 10-km} \times \textit{Masa treadmil}) = 7{,}887.85$

i) Nyatakan suatu model serta anggapannya.
ii) Lakukan suatu analisis statistik yang sesuai pada aras signifikan 0.05.
iii) Dapatkan suatu selang anggaran dan selang ramalan bagi jumlah masa yang dilakukan dalam suatu perlumbaan berlari 10-km jika masa keletihan daripada berlari di atas treadmil adalah 10.6 minit.

[20 markah]

APPENDIX: FORMULAS/*LAMPIRAN: FORMULA*

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

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

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

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

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

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

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

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

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/*LAMPIRAN : JADUAL*

Duncan Multiple Range Table

Spearman Table

Tukey HSD Table

-ooo0ooo-