UNIVERSITI SAINS MALAYSIA

Second Semester Examination
2016/2017 Academic Session

June 2017

MSG162 – Applied Statistical Methods
[Kaedah Statistik Gunaan]

Duration : 3 hours
[Masa : 3 jam]

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

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

Instructions: Answer **ALL 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. The effect of storage time on the mineral content of cereal grains was studied by a manufacturer. Five different storage times were considered. The following mineral content data were obtained:

Storage Time	Mineral Content (in mg)					Total
1 month	80	88	83	85	79	415
2 months	75	68	72	79	80	374
3 months	74	73	76	77	75	375
4 months	67	72	74	73	76	362
5 months	62	64	67	69	66	328

Assume that all assumptions to perform an ANOVA test are satisfied.

(a) Is there any difference in mineral content due to storage time? Use the significance level, $\alpha = 0.05$.

(b) Compute a 99% interval estimate of the mean of mineral content for the 3 months storage time.

(c) Compute a 95% interval estimate of the difference in the mean of mineral content between the 2 months and 4 months storage time.

(d) Test all pairs of means of mineral content using the Duncan's multiple range test. Use $\alpha = 0.05$.

[30 marks]

1. Kesan masa simpanan ke atas kandungan mineral bijirin telah dikaji oleh pengeluar. Lima masa simpanan yang berlainan dipertimbangkan. Data kandungan mineral berikut telah diperoleh:

Masa Simpanan	*Kandungan Mineral (dalam mg)*					*Jumlah*
1 bulan	*80*	*88*	*83*	*85*	*79*	*415*
2 bulan	*75*	*68*	*72*	*79*	*80*	*374*
3 bulan	*74*	*73*	*76*	*77*	*75*	*375*
4 bulan	*67*	*72*	*74*	*73*	*76*	*362*
5 bulan	*62*	*64*	*67*	*69*	*66*	*328*

Andaikan bahawa semua andaian untuk menjalankan ujian ANOVA dipenuhi.

(a) Adakah terdapat perbezaan dalam kandungan mineral akibat masa simpanan? Gunakan aras keertian, $\alpha = 0.05$.

(b) Kirakan anggaran selang 99% untuk min kandungan mineral bagi masa simpanan 3 bulan.

(c) Kirakan anggaran selang 95% untuk perbezaan dalam min kandungan mineral antara masa simpanan 2 bulan dan 4 bulan.

(d) Uji semua pasangan min kandungan mineral dengan menggunakan ujian julat berganda Duncan. Gunakan $\alpha = 0.05$.

[30 markah]

2. A plant biologist conducted an experiment to compare the yields (in kg) due to 4 types of raw materials (A, B, C, D). Four types of peanuts were used on four plots of land. Each plot of land was divided into four subplots of equal sizes. The yields (in kg) are given in the table below.

Type of peanut	Plot				Row Total
	1	2	3	4	
1	5 (A)	10 (B)	11 (C)	9 (D)	35
2	7 (B)	18 (C)	11 (D)	8 (A)	44
3	10 (C)	14 (D)	7 (A)	6 (B)	37
4	10 (D)	12 (A)	12 (B)	14 (C)	48
Column Total	32	54	41	37	164

(a) State the model and its assumptions.

(b) By using the ANOVA test, is there any difference in yields due to raw materials? Use $\alpha = 0.05$.

(c) If there is a significant difference in yields, perform a pair-wise comparison of the means due to raw materials using the Fisher LSD procedure. Use $\alpha = 0.05$.

(d) Is there any difference in yields among the four types of peanut? Use $\alpha = 0.01$.

(e) Estimate the following quantities using the Scheffe' procedure and draw conclusions from the results obtained: Use $\alpha = 0.05$.

(i) The effect of raw materials B and D with raw material A.

(ii) The effect of raw materials A and D with raw materials B and C.

[30 marks]

2. *Seorang ahli biologi tumbuhan menjalankan suatu eksperimen untuk membandingkan hasil (dalam kg) yang disebabkan oleh 4 jenis bahan mentah (A, B, C, D). Empat jenis kacang digunakan pada empat plot tanah. Setiap plot tanah dibahagikan kepada empat subplot dengan saiz yang sama. Hasil (dalam kg) diberikan dalam jadual di bawah.*

Jenis kacang	*Plot*				*Jumlah baris*
	1	*2*	*3*	*4*	
1	*5 (A)*	*10 (B)*	*11 (C)*	*9 (D)*	*35*
2	*7 (B)*	*18 (C)*	*11 (D)*	*8 (A)*	*44*
3	*10 (C)*	*14 (D)*	*7 (A)*	*6 (B)*	*37*
4	*10 (D)*	*12 (A)*	*12 (B)*	*14 (C)*	*48*
Jumlah lajur	*32*	*54*	*41*	*37*	*164*

(a) Nyatakan model dan andaiannya.

(b) Dengan menggunakan ujian ANOVA, adakah terdapat perbezaan dalam hasil yang disebabkan oleh bahan mentah? Gunakan $\alpha = 0.05$.

(c) Jika terdapat perbezaan bererti dalam hasil, jalankan perbandingan pasangan demi pasangan untuk min yang disebabkan oleh bahan mentah dengan menggunakan prosedur Fisher LSD. Gunakan $\alpha = 0.05$.

(d) Adakah terdapat perbezaan dalam hasil antara empat jenis kacang itu? Gunakan $\alpha = 0.01$.

(e) Anggarkan kuantiti berikut dengan menggunakan prosedur Scheffe' dan buat kesimpulan daripada keputusan yang diperoleh: Gunakan $\alpha = 0.05$.

(i) Kesan bahan mentah B dan D dengan bahan mentah A.

(ii) Kesan bahan mentah A dan D dengan bahan mentah B dan C.

[30 markah]

3. The quality control department of a fabric factory was studying the effect of two factors, i.e. operator and cycle time (in minutes), on the dyeing of a cloth used in manufacturing shirts. Three operators and four cycle times were selected. Two small specimens of cloths were dyed under each combination of cycle time and operator. The three operators were selected at random but only four cycle times were used in the experiment. The finished cloth was compared to a standard and a numerical score was given. The following results were obtained:

Cycle Time	Operator		
	1	2	3
40	17, 20	16, 21	24, 22
50	12, 9	18, 13	17, 12
60	16, 12	18, 21	25, 23
70	21, 17	23, 21	23, 22

Assume that all assumptions to perform an ANOVA test are satisfied.

(a) State the model and its assumptions.

(b) By using ANOVA, test for

(i) the significance of interaction effects between cycle time and operator. Use $\alpha = 0.05$.

(ii) the significance of cycle time effects. Use $\alpha = 0.05$.

(iii) the significance of operator effects. Use $\alpha = 0.05$.

[20 marks]

3. *Jabatan kawalan kualiti suatu kilang kain sedang mengkaji kesan dua faktor, iaitu operator dan masa kitaran (dalam minit), ke atas pencelupan suatu kain yang digunakan dalam pengeluaran pakaian. Tiga orang operator dan empat masa kitaran dipilih. Dua spesimen kecil kain telah dicelup di bawah setiap kombinasi masa kitaran dan operator. Tiga orang operator tersebut telah dipilih secara rawak tetapi hanya empat masa kitaran digunakan dalam eksperimen itu. Kain yang siap dibandingkan dengan suatu piawaian dan suatu skor berangka diberikan. Keputusan berikut diperoleh:*

Masa Kitaran	*Operator*		
	1	*2*	*3*
40	*17, 20*	*16, 21*	*24, 22*
50	*12, 9*	*18, 13*	*17, 12*
60	*16, 12*	*18, 21*	*25, 23*
70	*21, 17*	*23, 21*	*23, 22*

Andaikan bahawa semua andaian untuk menjalankan ujian ANOVA dipenuhi.

(a) Nyatakan model dan andaiannya.

(b) Dengan menggunakan ANOVA, uji untuk

(i) keertian kesan interaksi antara masa kitaran dan operator. Gunakan $\alpha = 0.05$.

(ii) keertian kesan masa kitaran. Gunakan $\alpha = 0.05$.

(iii) keertian kesan operator. Gunakan $\alpha = 0.05$.

[20 markah]

4. Ten workers were randomly selected from a production line. The number of hours they attended a training and the number of production units assembled correctly were recorded as follows:

Number of hours attended training	17	20	10	14	6	22	7	11	15	21
Number of units assembled correctly	88	105	82	94	70	108	70	92	90	108

(a) State the estimated regression model.

(b) Are the two variables correlated? Use $\alpha = 0.05$.

(c) Find an interval estimate for the number of units assembled correctly if a worker attends 9 hours of training. Use $\alpha = 0.05$.

(d) Find a prediction interval for the number of units assembled correctly if a worker attends 9 hours of training. Use $\alpha = 0.05$.

[20 marks]

4. Sepuluh orang pekerja telah dipilih secara rawak dari suatu barisan pengeluaran. Bilangan jam mereka menghadiri suatu latihan dan bilangan unit pengeluaran yang dipasang dengan betul telah direkodkan seperti berikut:

Bilangan jam menghadiri latihan	*17*	*20*	*10*	*14*	*6*	*22*	*7*	*11*	*15*	*21*
Bilangan unit yang dipasang dengan betul	*88*	*105*	*82*	*94*	*70*	*108*	*70*	*92*	*90*	*108*

(a) Nyatakan model regresi anggaran.

(b) Adakah dua pembolehubah itu berkorelasi? Gunakan $\alpha = 0.05$.

(c) Cari suatu anggaran selang untuk bilangan unit yang dipasang dengan betul jika seorang pekerja menghadiri 9 jam latihan. Gunakan $\alpha = 0.05$.

(d) Cari suatu selang ramalan untuk bilangan unit yang dipasang dengan betul jika seorang pekerja menghadiri 9 jam latihan. Gunakan $\alpha = 0.05$.

[20 markah]

APPENDIX: FORMULAE

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

$$\text{SST} = \sum_{\text{i}}\sum_{\text{j}}\sum_{\text{k}} Y_{\text{ijk}}^2 - \frac{Y_{\cdots}^2}{\text{N}}$$

$$\text{SSA} = \sum_{\text{i}} \frac{Y_{\text{i}\cdot\cdot}^2}{\text{bn}} - \frac{Y_{\cdots}^2}{\text{N}}$$

$$\text{SSB} = \sum_{\text{j}} \frac{Y_{\cdot\text{j}\cdot}^2}{\text{an}} - \frac{Y_{\cdots}^2}{\text{N}}$$

$$\text{SSE} = \sum_{\text{i}}\sum_{\text{j}}\sum_{\text{k}} Y_{\text{ijk}}^2 - \sum_{\text{i}}\sum_{\text{j}} \frac{Y_{\text{ij}\cdot}^2}{\text{n}}$$

5. Multiple Comparison Procedures:

Duncan: $\text{r}_{\alpha,\, \text{p},\, \text{df}}$, p = range df = degrees of freedom

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

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

Fisher LSD: $\text{t}_{\frac{\alpha}{2},\, \text{df}}$, df = degrees of freedom

6. Regression

$$b_1 = \frac{\text{SS}_{XY}}{\text{SS}_X} \quad , \qquad b_0 = \bar{Y} - b_1\bar{X}$$

$$\text{SSE} = \text{SS}_Y - \frac{[\text{SS}_{XY}]^2}{\text{SS}_X}$$

$$\text{SS}_{XY} = \sum X_{\text{i}} Y_{\text{i}} - \frac{(\sum X_{\text{i}})(\sum Y_{\text{i}})}{\text{n}}$$

$$\text{SS}_X = \sum X_{\text{i}}^2 - \frac{(\sum X_{\text{i}})^2}{\text{n}}$$

$$\text{SS}_Y = \sum Y_{\text{i}}^2 - \frac{(\sum Y_{\text{i}})^2}{\text{n}}$$

$$\mathrm{Var}\left(b_1\right)=\frac{\sigma^2}{\mathrm{SS}_X}$$

$$\mathrm{Var}\left(\hat{\mathrm{Y}}_\mathrm{h}\right)=\sigma^2\left[\frac{1}{\mathrm{n}}+\frac{\left(\mathrm{X}_\mathrm{h}-\bar{X}\right)^2}{\mathrm{SS}_X}\right]$$

7. Correlation

$$\mathrm{r}=\frac{\mathrm{SS}_{XY}}{\sqrt{\mathrm{SS}_X\ \mathrm{SS}_Y}}$$

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

8. Kruskal-Wallis Test

$$\mathrm{T}=\frac{12}{\mathrm{N}\left(\mathrm{N}+1\right)}\sum_\mathrm{i}\frac{\mathrm{R}_\mathrm{i}^2}{\mathrm{n}_\mathrm{i}}-3\left(\mathrm{N}+1\right)$$

9. Friedman Test

$$\mathrm{T}=\frac{12}{\mathrm{ab}\left(\mathrm{a}+1\right)}\sum_\mathrm{i}\mathrm{R}_\mathrm{i}^2-3\mathrm{b}\left(\mathrm{a}+1\right)$$

10. Spearman Test

$$\mathrm{r}\ =\ 1-\frac{6\sum_{\mathrm{i}=1}^{\mathrm{n}}\left[\mathrm{R}\left(X_\mathrm{i}\right)-\mathrm{R}\left(Y_\mathrm{i}\right)\right]^2}{\mathrm{n}\left(\mathrm{n}^2-1\right)}=1-\frac{6\sum_\mathrm{i}\mathrm{d}_\mathrm{i}^2}{\mathrm{n}\left(\mathrm{n}^2-1\right)}$$

APPENDIX: TABLE

Duncan's Multiple Range Table

$r_{0.05, p, df}$ (df = v)

$\alpha = 0.05$									
	p								
v	2	3	4	5	6	7	8	9	10
1	17.97	17.97	17.97	17.97	17.97	17.97	17.97	17.97	17.97
2	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085
3	4.501	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516
4	3.927	4.013	4.033	4.033	4.033	4.033	4.033	4.033	4.033
5	*3.635*	*3.749*	*3.797*	*3.814*	*3.814*	*3.814*	*3.814*	*3.814*	*3.814*
6	3.461	3.587	3.649	3.68	3.694	3.697	3.697	3.697	3.697
7	3.344	3.477	3.548	3.588	3.611	3.622	3.626	3.626	3.626
8	3.261	3.399	3.475	3.521	3.549	3.566	3.575	3.579	3.579
9	3.199	3.339	3.420	3.470	3.502	3.523	3.536	3.544	3.547
10	3.151	3.293	3.376	3.430	3.465	3.489	3.505	3.516	3.522
11	3.113	3.256	3.342	3.397	3.435	3.462	3.48	3.493	3.501
12	3.082	3.225	3.313	3.370	3.410	3.439	3.459	3.474	3.484
13	3.055	3.200	3.289	3.348	3.389	3.419	3.442	3.458	3.470
14	3.033	3.178	3.268	3.329	3.372	3.403	3.426	3.444	3.457
15	3.014	3.160	3.25	3.312	3.356	3.389	3.413	3.432	3.446
16	2.998	3.144	3.235	3.298	3.343	3.376	3.402	3.422	3.437
17	2.984	3.130	3.222	3.285	3.331	3.366	3.392	3.412	3.429
18	2.971	3.118	3.210	3.274	3.321	3.356	3.383	3.405	3.421
19	2.960	3.107	3.199	3.264	3.311	3.347	3.375	3.397	3.415
20	2.950	3.097	3.190	3.255	3.303	3.339	3.368	3.391	3.409
24	2.919	3.066	3.160	3.226	3.276	3.315	3.345	3.370	3.390
30	2.888	3.035	3.131	3.199	3.250	3.290	3.322	3.349	3.371
40	2.858	3.006	3.102	3.171	3.224	3.266	3.300	3.328	3.352
60	2.829	2.976	3.073	3.143	3.198	3.241	3.277	3.307	3.333
120	2.800	2.947	3.045	3.116	3.172	3.217	3.254	3.287	3.314
∞	2.772	2.918	3.017	3.089	3.146	3.193	3.232	3.265	3.294

- ooo 0 ooo -