Second Semester Examination
2017/2018 Academic Session

May / June 2018

MSG162 - Applied Statistical Methods
(Kaedah Statistik Gunaan)

Duration : 3 hours
[Masa : 3 jam]

Please check that this examination paper consists of **TWELVE (12)** pages of printed material before you begin the examination.

[*Sila pastikan bahawa kertas peperiksaan ini mengandungi **DUA BELAS** **(12)** 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.]

Question 1

An investigation was conducted to determine the source of reduction in yield of a certain chemical product. It was thought that four different brands of a material may result in different yield reductions. The following are the results of the percent of yield reductions for three batches, for each of the four brands:

Percent of yield reductions			
Brand 1	Brand 2	Brand 3	Brand 4
25.6	25.2	20.8	31.6
24.3	28.6	26.7	29.8
27.9	24.7	22.2	34.3

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

(a) Test the hypothesis that the four brands have the same yield reduction at the significance level, $\alpha = 0.05$.

(b) Compute a 95% interval estimate of the mean yield reduction of Brand 3.

(c) Compute a 99% interval estimate of the mean difference between the yield reduction of Brands 4 and 3.

(d) Use the Fisher's LSD method to find significant differences in the brands at $\alpha = 0.05$.

[25 marks]

Soalan 1

Suatu siasatan telah dijalankan untuk menentukan punca pengurangan hasil suatu produk kimia. Adalah difikirkan bahawa empat jenama yang berlainan untuk suatu bahan akan menyebabkan pengurangan hasil yang berbeza. Berikut adalah keputusan peratusan pengurangan hasil untuk 3 kelompok, bagi setiap daripada empat jenama:

Peratusan pengurangan hasil			
Jenama 1	*Jenama 2*	*Jenama 3*	*Jenama 4*
25.6	*25.2*	*20.8*	*31.6*
24.3	*28.6*	*26.7*	*29.8*
27.9	*24.7*	*22.2*	*34.3*

Andaikan bahawa semua andaian untuk menjalankan ujian ANOVA dipenuhi.

(a) Uji hipotesis bahawa keempat-empat jenama mempunyai pengurangan hasil yang sama pada aras keertian, $\alpha=0.05$.

…3/-

(b) Kirakan anggaran selang 95% untuk min pengurangan hasil bagi Jenama 3.

(c) Kirakan anggaran selang 99% untuk perbezaan dalam min pengurangan hasil antara Jenama 4 dan 3.

(d) Gunakan kaedah Fisher LSD untuk mencari perbezaan bererti antara jenama-jenama pada α=0.05.

[25 markah]

Question 2

A study was conducted to study the effects of four types of fertilizers, A, B, C and D on the yield of beans. Beans were grown in three blocks of soil (each block contains 4 homogeneous plots). It was thought that different blocks of soil might influence the yield of beans for a given fertilizer. The yields in kilograms per lot are as follows:

Soil Block	Fertilizer A	Fertilizer B	Fertilizer C	Fertilizer D
1	42.7	39.3	48.5	32.8
2	50.0	38.0	50.9	40.2
3	51.9	46.3	53.5	51.1

(a) State an appropriate statistical model and its assumptions.

(b) By using the ANOVA test, test whether there are significant differences in the yields of beans for the four types of fertilizers. Subsequently, conduct a pair-wise comparison of the treatment means using the Duncan's multiple range test. Use $\alpha = 0.05$.

(c) By using the ANOVA test, test whether there are significant differences in the yields of beans among the soil blocks? Use $\alpha = 0.05$.

[25 marks]

Soalan 2

Suatu kajian telah dijalankan untuk mengkaji kesan empat jenis baja, A, B, C dan D ke atas hasil kacang. Kacang ditanam dalam tiga blok tanah (setiap blok mengandungi empat plot homogen). Adalah difikirkan bahawa blok tanah yang berlainan mungkin mempengaruhi hasil kacang untuk suatu baja tertentu. Hasil dalam kilogram per lot adalah seperti berikut:

Blok Tanah	*Baja A*	*Baja B*	*Baja C*	*Baja D*
1	*42.7*	*39.3*	*48.5*	*32.8*
2	*50.0*	*38.0*	*50.9*	*40.2*
3	*51.9*	*46.3*	*53.5*	*51.1*

(a) Nyatakan suatu model berstatistik yang sesuai dan andaiannya.

(b) Dengan menggunakan ujian ANOVA, uji sama ada terdapat perbezaan bererti dalam hasil kacang bagi empat jenis baja tersebut. Seterusnya, jalankan perbandingan pasangan demi pasangan untuk min rawatan dengan menggunakan ujian julat berganda Duncan. Gunakan $\alpha = 0.05$.

(c) Dengan menggunakan ujian ANOVA, uji sama ada terdapat perbezaan bererti dalam hasil kacang antara blok-blok tanah? Gunakan $\alpha = 0.05$.

[25 markah]

Question 3

Gaskets were produced by a company from two different machines and three different material types. Three replicates were considered for each of the machine and material type combination. The following data on the number of gaskets (in thousands) produced per hour were obtained.

Machine	Material Type		
	I	II	III
A	4.31, 4.27, 4.40	3.36, 3.42, 3.48	4.01, 3.94, 3.89
B	3.94, 3.81, 3.99	3.91, 3.80, 3.85	3.48, 3.53, 3.42

(a) State an appropriate statistical model and its assumptions.

(b) By using the ANOVA test, test for

(i) equality in treatment means.

(ii) the significance of interaction effect between material type and machine.

(iii) the significance of machine effects.

(iv) the significance of material type effects.

Use $\alpha = 0.05$.

[25 marks]

Soalan 3

Gasket dikeluarkan oleh sebuah syarikat daripada dua mesin berbeza dan tiga jenis bahan berlainan. Tiga ulangan dipertimbangkan untuk setiap gabungan mesin dan jenis bahan. Data berikut tentang bilangan gasket (dalam ribu) yang dihasilkan setiap jam telah diperoleh.

Mesin	*Jenis Bahan*		
	I	*II*	*III*
A	*4.31, 4.27, 4.40*	*3.36, 3.42, 3.48*	*4.01, 3.94, 3.89*
B	*3.94, 3.81, 3.99*	*3.91, 3.80, 3.85*	*3.48, 3.53, 3.42*

(a) Nyatakan suatu model berstatistik yang sesuai dan andaiannya.

(b) Dengan menggunakan ujian ANOVA, uji

(i) kesamaan dalam min rawatan.

(ii) keertian kesan interaksi antara jenis bahan dan mesin.

(iii) keertian kesan mesin.

(v) keertian kesan jenis bahan.

Gunakan $\alpha = 0.05$.

[25 markah]

Question 4

A placement test was given to 20 new students at a college. At the end of the year, these 20 students sat for the final examination. The placement test and final examination marks for these 20 students were recorded in the following table:

Marks in placement test	Marks in final examination
50	53
35	41
35	61
40	56
55	68
65	36
35	11
60	70
90	79
35	59
90	54
80	91
60	48
60	71
60	71
40	47
55	53
50	68
65	57
50	79

(a) State the estimated regression model.

(b) Are the two variables correlated? Use $\alpha = 0.1$.

(c) Compute an interval estimate for the marks in final examination if a student gets 35 marks in the placement test. Use $\alpha = 0.1$.

(d) Compute a prediction interval for the marks in final examination if a student gets 35 marks in the placement test. Use $\alpha = 0.1$.

[25 marks]

Soalan 4

Suatu ujian penempatan telah diberikan kepada 20 pelajar baru di sebuah maktab. Pada akhir tahun, 20 pelajar ini menduduki peperiksaan akhir. Markah ujian penempatan dan peperiksaan akhir untuk 20 pelajar ini direkodkan dalam jadual berikut:

Markah ujian penempatan	*Markah peperiksaan akhir*
50	*53*
35	*41*
35	*61*
40	*56*
55	*68*
65	*36*
35	*11*
60	*70*
90	*79*
35	*59*
90	*54*
80	*91*
60	*48*
60	*71*
60	*71*
40	*47*
55	*53*
50	*68*
65	*57*
50	*79*

(a) *Nyatakan model regresi anggaran.*

(b) *Adakah dua pembolehubah itu berkorelasi? Gunakan $\alpha = 0.1$.*

(c) *Kirakan suatu anggaran selang untuk markah peperiksaan akhir jika seorang pelajar mendapat 35 markah dalam ujian penempatan. Gunakan $\alpha = 0.1$.*

(d) *Kirakan suatu selang ramalan untuk markah peperiksaan akhir jika seorang pelajar mendapat 35 markah dalam ujian penempatan. Gunakan $\alpha = 0.1$.*

[25 markah]

APPENDIX: FORMULAE

1. Completely Randomized Design

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

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

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

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

2. Completely Randomized Block Design

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

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

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

3. Latin Square Design

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

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

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

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

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: $r_{\alpha,\, p,\, df}$, 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

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

6. Regression

$$b_1 = \frac{SS_{XY}}{SS_X} \quad , \qquad b_0 = \bar{Y} - b_1\bar{X}$$

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

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

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

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

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

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

7. Correlation

$$r = \frac{SS_{XY}}{\sqrt{SS_X\ SS_Y}}$$

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

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

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

APPENDIX: 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

- oooOOooo -