

UNIVERSITI SAINS MALAYSIA

Peperiksaan Semester Kedua
Sidang Akademik 1996/97

April 1997

MSG 162/MSG 261 - Kaedah Statistik Gunaan

Masa: [3 jam]

ARAHAN KEPADA CALON:

Sila pastikan bahawa kertas peperiksaan ini mengandungi EMPAT soalan di dalam ENAM halaman yang bercetak sebelum anda memulakan peperiksaan ini.

Jawab SEMUA soalan. Soalan-soalan mesti dijawab di dalam Bahasa Malaysia. Sifir New Cambridge Statistical Tables disediakan. Satu set lampiran diberikan. Mesin kira non-programmable boleh digunakan dan ia disediakan oleh pelajar.

1. (a) Di sebuah stesen minyak, dua jenis gasoline A dan B dijual. Pada suatu hari, 50 pembelian pertama adalah seperti

AABAABABBAAABBABBABBABBA
BBABBAABBBBAABABABAAABAA
BB

Adakah jujukan ini rawak? Gunakan $\alpha = 0.05$.

(25 markah)

- (b) Seorang penyelidik memberi satu ujian tentang isu ekologi kepada 15 orang lelaki dan 13 orang perempuan, dan skor mereka adalah seperti berikut:

Skor lelaki (x)		Skor perempuan (y)	
10.0	16.0	15.0	18.4
11.0	17.0	16.2	18.5
12.0	18.5	16.3	18.7
12.4	19.0	17.3	19.1
12.5	19.3	17.5	19.2
12.5	19.5	17.8	
14.0	20.0	18.0	
15.2		18.3	

Ujikan hipotesis bahawa 2 sampel ini dari dua populasi yang mediannya sama. Gunakan ujian Mann-Whitney $\alpha = 0.05$.

(25 markah)

..2/-

- (c) Sepuluh orang pelajar matematik tahun satu dipilih secara rawak, dan pencapaian mereka di dalam kursus Kalkulus dan kursus Statistik dicatatkan.

Kalkulus	30	42	47	50	50	51	65	65	70	80
Statistik	30	70	40	65	50	65	70	78	82	41

- (i) Kira pekali korelasi Spearman, r_s .
 (ii) Berdasarkan data ini, bolehkah kita menyimpul bahawa pencapaian Kalkulus dan Statistik ada korelasi? Gunakan Spearman r_s . $\alpha = 0.05$.
 (25 markah)
- (d) Di bandar Kuala Lumpur kadar kemalangan kenderaan automobil per 10,000 kenderaan automobil untuk 21 tahun yang lalu adalah seperti yang ditunjukkan:

16.9 16.9 16.7 17.1 17.3 17.9 18.4 18.1 18.3 19.6 18.6
 19.2 17.7 20.0 19.0 18.8 19.3 20.2 19.3 19.4 19.5

Adakah kadar kemalangan untuk bandar itu menokok? Gunakan paras keertian $\alpha = 0.05$

(25 markah)

2. (a) Empat cara yang berlainan digunakan untuk menghasilkan besi keluri. Kekuatan tensil besi keluri dari setiap cara diukur. Datanya adalah seperti yang berikut: (data telah disusun mengikut magnitud)

A	B	C	D
80	91	80	71
85	92	81	73
86	96	82	75
87	98	84	75
88	98	86	76
88	98	86	76
90	99	86	77
	99	89	78
			80
			80

Bolehkah kita menyatakan bahawa kekuatan tensil besi keluri dari empat cara ini adalah sama? Gunakan $\alpha = 0.05$.

(25 markah)

- (b) Syarikat Top Dog Puppy Food ingin mengetahui sama ada tiga Formula A, B, C adalah sama berkesan kepada kesihatan anjing. Di dalam satu eksperimen, 15 kumpulan (anak-beranak) digunakan. Setiap kumpulan ada tiga anjing. Secara rawak satu diberi Formula A, satu diberi Formula B dan satu lagi diberi Formula C. Selepas sesuatu tempoh masa, seorang doktor haiwan memberi pangkat kepada tiga anjing bagi setiap kumpulan.

Kumpulan	Formula		
	A	B	C
1	3	2	1
2	3	1	2
3	3	2	1
4	2.5	2.5	1
5	1	3	2
6	3	2	1
7	3	2	1
8	2	3	1
9	2.5	2.5	1
10	3	1	2
11	2	3	1
12	3	1	2
13	3	1	2
14	2	3	1
15	2	3	1

Ujikan hipotesis bahawa tiga Formula adalah sama berkesan. $\alpha = 0.05$.

(25 markah)

- (c) Seorang penganalisis pemasaran ingin mengetahui sama ada min pendapatan tahunan keluarga adalah kurang dari RM15,000.00 di Bandar Georgetown. Dari sampel sebanyak 16 keluarga, didapati data:

8900 9300 10100 18000 10300 1200 7500 9900
 11200 15300 17200 23000 15100 14900 14300 10200

Ujikan hipotesis ini. Gunakan ujian pangkat bertanda Wilcoxon. $\alpha = 0.05$.

(25 markah)

...4/-

- (d) Untuk membandingkan kuasa relatif tiga ujian statistik, 120 set data buatan dijanakan komputer. Bagi setiap set data, ketiga-tiga ujian statistik itu dijalankan dengan paras keertian $\alpha = 0.05$, dan keputusan untuk menerima atau menolak hipotesis H_0 dicatat. Maklumatnya adalah seperti yang berikut:

Ujian I	Ujian II	Ujian III	Bilangan set data
terima	terima	terima	28
terima	terima	tolak	8
terima	tolak	terima	15
tolak	terima	terima	9
tolak	tolak	terima	19
tolak	terima	tolak	6
terima	tolak	tolak	8
tolak	tolak	tolak	27

Berdasarkan maklumat ini, bolehkan kita menyatakan bahawa tiga ujian statistik ini sama berkuasa? Gunakan $\alpha = 0.05$.

(25 markah)

3. (a) (i) Namakan dua model ANOVA dan terangkan secara ringkas bagi setiap satu.
(ii) Apakah perbezaan di antara model ANOVA dan model regresi.

(30 markah)

- (b) Lapan subjek A, B, C, D, E, F, G dan H akan diberi rawatan A atau B secara rawak, iaitu 4 subjek bagi setiap rawatan. Terangkan bagaimana anda akan menjalankan suatu ujikaji rekabentuk rawakan lengkap dengan menunjukkan subjek mana yang akan menerima rawatan mana dan pada masa bila. Gunakan nombor-nombor rawak berikut. (Mulakan dengan nombor pertama).

11, 17, 49, 96, 91, 81, 41, 78, 82, 59, 52, 06,
58, 75, 47, 36, 89, 26, 70, 66, 13, 69, 80, 40.

- (c) Andaikan data berikut diperolehi dari suatu rekabentuk rawakan lengkap.

Indek	n	min	s
1	5	0.48	0.12
2	5	0.69	0.30
3	5	0.51	0.17
4	5	0.50	0.14

- (i) Dapatkan suatu set kontras untuk membandingkan purata bagi indek 1 dan 2 dengan purata bagi indek 3 dan 4, purata bagi indek 1 dan 3 dengan purata bagi indek 2 dan 4, dan purata bagi indek 2 dan 3 dengan purata bagi indek 1 dan 4. Adakah set kontras ini berortogon? Terangkan.

...5/-

- (ii) Dapatkan jadual ANOVA bagi menguji set kontras di (i) dan berikan kesimpulannya. Gunakan $\alpha = .05$.
- (iii) Dengan kaedah Scheffe, dapatkan suatu selang keyakinan serentak 95% bagi kontras yang pertama, iaitu membandingkan purata bagi indeks 1 dan 2 dengan purata bagi indeks 3 dan 4.
- (iv) Jalankan ujian kesamaan varians dengan menggunakan ujian Bartlett.

(50 markah)

4. (a) Terangkan secara ringkas mengenai dua perkara berikut:

- (i) plot kebarangkalian normal
(ii) reja piawai.

(20 markah)

- (b) 50 siput yang diperoleh dari setiap satu dari 4 kawasan yang berlainan dicatatkan berat keringnya (dalam gram). Ringkasan bagi min dan (sisihan piawai) bagi data yang dikumpulkan adalah seperti berikut:

	ombak rendah	ombak tinggi
Tak tercemar	0.48 (0.12)	0.69 (0.30)
Tercemar	0.51 (0.17)	0.50 (0.14)

dengan $SS_{\text{kecemaran}} = 0.3200$
 $SS_{\text{ombak}} = 0.5000$
 $SS_{\text{kecemaran*ombak}} = 0.6050$
 $SS_{\text{ralat}} = 7.492$

- (i) Dapatkan jadual ANOVA dan berikan kesimpulannya.
(ii) Dapatkan plot salingtindak bagi min-minnya. Apakah yang anda lihat? Bandingkan dengan kesimpulan anda di (i)

(30 markah)

(c) Berikut ialah data dari suatu rekabentuk segiempat sama Latin.

Baris	Lajur			
	1	2	3	4
1	44 (C)	41 (A)	30 (B)	40 (D)
2	41 (B)	42 (C)	49 (D)	49 (A)
3	59 (A)	41 (D)	59 (C)	34 (B)
4	58 (D)	37 (B)	53 (A)	59 (C)

Diberi $SS_{\text{baris}} = 388.75$
 $SS_{\text{lajur}} = 204.25$
 $SS_{\text{abjad}} = 584.75$
 $SS_{\text{jumlah}} = 1297.75$

...6/-

Dapatkan suatu set selang keyakinan serentak 95% bagi perbandingan $\mu_i = \mu_4$ ($i = 1, 2, 3$), membandingkan setiap min dengan min kawalan dan gunakan selang keyakinan ini untuk menguji $H_0: \mu_i = \mu_4$ ($i = 1, 2, 3$).

(30 markah)

- (d) Suatu eksperimen dijalankan untuk mengkaji faktor-faktor yang mempengaruhi populasi bakteria dalam air mandian. Faktor-faktor yang diambil kira ialah masa (dalam jam) semenjak kali terakhir seseorang itu mandi, aktiviti semasa mandi dan jantina. Empat respon diukur: paras '*fecal coliform*' dan '*total coliform*' (dalam organisma per 100 ml.) bagi 15 dan 30 minit selepas mandi.

Berikut ialah data yang dikumpulkan:

Masa semenjak kali terakhir mandi (A)	Aktiviti mandi (B)	Jantina (C)	'Fecal coliform'		'Total coliform'	
			15 min	30 min	15 min	30 min
1	berehat	Perempuan	1, 2	1, 4	3, 10	7, 27
1	berehat	Lelaki	153, 96	170, 67	426, 147	590, 193
1	bergerak cergas	Perempuan	16, 21	10, 21	323, 33	360, 53
1	bergerak cergas	Lelaki	143, 300	170, 377	580, 665	450, 810
24	berehat	Perempuan	12, 37	15, 39	57, 280	80, 250
24	berehat	Lelaki	129, 390	148, 360	280, 1470	243, 1560
24	bergerak cergas	Perempuan	4, 2	6, 5	183, 10	193, 87
24	bergerak cergas	Lelaki	113, 280	217, 250	650, 675	735, 795

berserta dengan maklumat-maklumat berikut:

Sumber	dk	SS
A	1	2.409
B	1	2.277
A*B	1	1.905
C	1	15.687
A*C	1	0.511
B*C	1	0.505
A*B*C	1	1.030
Ralat	8	6.242

Jalankan satu analisis terhadap maklumat di atas.

(20 markah)

-ooo0ooo-

Rumus-Rumus

1. Bagi sebarang kontras, $C = \sum_i C_i y_i$.

$$SS_C = \frac{\left(\sum_i (C_i y_i) \right)^2}{n \sum C_i^2}$$

2. Kaedah Scheffe

Kontras, $C_u = \sum_i C_{iu} \bar{y}_i$

$$S_{C_u} = \sqrt{MSE \sum_i \frac{C_{iu}^2}{n_i}}$$

$$S_{\alpha, u} = S_{C_u} \sqrt{(a-1) F_{\alpha, a-1, N-a}}$$

3. Ujian Bartlett

$$C = 1 + \frac{1}{3(a-1)} \left[\left(\sum_{i=1}^a \frac{1}{dk_i} \right) - \frac{1}{dk_T} \right]$$

$$B = \frac{1}{C} \left[(dk_T) \log_e MSE - \sum_{i=1}^a (dk_i) \log_e S_i^2 \right]$$

Lampiran 2

Nilai Genting bagi Ujian Dunnett bagi membandingkan Rawatan dengan Suatu Kawalan^a

$d_{.05}(a - 1, f)$
Perbandingan Dua Bahagian

f	$a - 1 =$ Bilangan min rawatan (tanpa kawalan)								
	1	2	3	4	5	6	7	8	9
5	2.57	3.03	3.29	3.48	3.62	3.73	3.82	3.90	3.97
6	2.45	2.86	3.10	3.26	3.39	3.49	3.57	3.64	3.71
7	2.36	2.75	2.97	3.12	3.24	3.33	3.41	3.47	3.53
8	2.31	2.67	2.88	3.02	3.13	3.22	3.29	3.35	3.41
9	2.26	2.61	2.81	2.95	3.05	3.14	3.20	3.26	3.32
10	2.23	2.57	2.76	2.89	2.99	3.07	3.14	3.19	3.24
11	2.20	2.53	2.72	2.84	2.94	3.02	3.08	3.14	3.19
12	2.18	2.50	2.68	2.81	2.90	2.98	3.04	3.09	3.14
13	2.16	2.48	2.65	2.78	2.87	2.94	3.00	3.06	3.10
14	2.14	2.46	2.63	2.75	2.84	2.91	2.97	3.02	3.07
15	2.13	2.44	2.61	2.73	2.82	2.89	2.95	3.00	3.04
16	2.12	2.42	2.59	2.71	2.80	2.87	2.92	2.97	3.02
17	2.11	2.41	2.58	2.69	2.78	2.85	2.90	2.95	3.00
18	2.10	2.40	2.56	2.68	2.76	2.83	2.89	2.94	2.98
19	2.09	2.39	2.55	2.66	2.75	2.81	2.87	2.92	2.96
20	2.09	2.38	2.54	2.65	2.73	2.80	2.86	2.90	2.95
24	2.06	2.35	2.51	2.61	2.70	2.76	2.81	2.86	2.90
30	2.04	2.32	2.47	2.58	2.66	2.72	2.77	2.82	2.86
40	2.02	2.29	2.44	2.54	2.62	2.68	2.73	2.77	2.81
60	2.00	2.27	2.41	2.51	2.58	2.64	2.69	2.73	2.77
120	1.98	2.24	2.38	2.47	2.55	2.60	2.65	2.69	2.73
∞	1.96	2.21	2.35	2.44	2.51	2.57	2.61	2.65	2.69

f = darjah kebebasan

^a Dikeluarkan dengan kebenaran daripada C. W. Dunnett, "New Tables for Multiple Comparison with a Control," *Biometrics*, Jil. 20, No. 3, 1964, dan daripada C. W. Dunnett, "A Multiple Comparison Procedure for Comparing Several Treatments with a Control," *Journal of the American Statistical Association*, Jil. 50, 1955.

Lampiran 3

Table A.8 Quantiles of the Mann-Whitney test statistic

n_1	p	$n_2=2$	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
13	.001	0	0	2	4	6	9	12	15	18	21	24	27	30	33	36	39	43	46	49	
	.005	0	2	4	8	11	14	18	21	25	28	32	35	39	43	46	50	54	58	61	
	.01	1	3	6	10	13	17	21	24	28	32	36	40	44	48	52	56	60	64	68	
	.025	2	5	9	13	17	21	25	29	34	38	42	46	51	55	60	64	68	73	77	
	.05	3	7	11	16	20	25	29	34	38	43	48	52	57	62	66	71	76	81	85	
	.10	5	10	14	19	24	29	34	39	44	49	54	59	64	69	75	80	85	90	95	
	14	.001	0	0	2	4	7	10	13	16	20	23	26	30	33	37	40	44	47	51	55
		.005	0	2	5	8	12	16	19	23	27	31	35	39	43	47	51	55	59	64	68
		.01	1	3	7	11	14	18	23	27	31	35	39	44	48	52	57	61	66	70	74
		.025	2	6	10	14	18	23	27	32	37	41	46	51	56	60	65	70	75	79	84
.05		4	8	12	17	22	27	32	37	42	47	52	57	62	67	72	78	83	88	93	
.10		5	11	16	21	26	32	37	42	48	53	59	64	70	75	81	86	92	98	103	
15		.001	0	0	2	5	8	11	15	18	22	25	29	33	37	41	44	48	52	56	60
		.005	0	3	6	9	13	17	21	25	30	34	38	43	47	52	56	61	65	70	74
		.01	1	4	8	12	16	20	25	29	34	38	43	48	52	57	62	67	71	76	81
		.025	2	6	11	15	20	25	30	35	40	45	50	55	60	65	71	76	81	86	91
	.05	4	8	13	19	24	29	34	40	45	51	56	62	67	73	78	84	89	95	101	
	.10	6	11	17	23	28	34	40	46	52	58	64	69	75	81	87	93	99	105	111	
	16	.001	0	0	3	6	9	12	16	20	24	28	32	36	40	44	49	53	57	61	66
		.005	0	3	6	10	14	19	23	28	32	37	42	46	51	56	61	66	71	75	80
		.01	1	4	8	13	17	22	27	32	37	42	47	52	57	62	67	72	77	83	88
		.025	2	7	12	16	22	27	32	38	43	48	54	60	65	71	76	82	87	93	99
.05		4	9	15	20	26	31	37	43	49	55	61	66	72	78	84	90	96	102	108	
.10		6	12	18	24	30	37	43	49	55	62	68	75	81	87	94	100	107	113	120	

T

Table 7 QUANTILES OF THE WILCOXON SIGNED RANKS TEST STATISTIC^a

	$w_{.005}$	$w_{.01}$	$w_{.025}$	$w_{.05}$	$w_{.10}$	$w_{.20}$	$w_{.30}$	$w_{.40}$	$w_{.50}$	$\frac{n(n+1)}{2}$
$n = 4$	0	0	0	0	1	3	3	4	5	10
5	0	0	0	1	3	4	5	6	7.5	15
6	0	0	1	3	4	6	8	9	10.5	21
7	0	1	3	4	6	9	11	12	14	28
8	1	2	4	6	9	12	14	16	18	36
9	2	4	6	9	11	15	18	20	22.5	45
10	4	6	9	11	15	19	22	25	27.5	55
11	6	8	11	14	18	23	27	30	33	66
12	8	10	14	18	22	28	32	36	39	78
13	10	13	18	22	27	33	38	42	45.5	91
14	13	16	22	26	32	39	44	48	52.5	105
15	16	20	26	31	37	45	51	55	60	120
16	20	24	30	36	43	51	58	63	68	136
17	24	28	35	42	49	58	65	71	76.5	153
18	28	33	41	48	56	66	73	80	85.5	171
19	33	38	47	54	63	74	82	89	95	190
20	38	44	53	61	70	82	91	98	105	210

For n larger than 20, the p th quantile w_p of the Wilcoxon signed ranks test statistic may be approximated by $w_p = [n(n+1)/4] + x_p \sqrt{n(n+1)(2n+1)/24}$, where x_p is the p th quantile of a standard normal random variable, obtained from Table 1.

SOURCE. Adapted from Table 1, McCormack (1965).

^a The entries in this table are quantiles w_p of the Wilcoxon signed ranks test statistic T , given by Equation (5.1.4), for selected values of $p \leq .50$. Quantiles w_p for $p > .50$ may be computed from the equation

$$w_p = n(n+1)/2 - w_{1-p}$$

where $n(n+1)/2$ is given in the right hand column in the table. Note that $P(T < w_p) \leq p$ and $P(T > w_p) \leq 1 - p$ if H_0 is true. Critical regions correspond to values of T less than (or greater than) but not including the appropriate quantile.

Table 10 QUANTILES OF THE SPEARMAN TEST STATISTIC^a

<i>n</i>	<i>p</i> = .900	.950	.975	.990	.995	.999
4	.8000	.8000				
5	.7000	.8000	.9000	.9000		
6	.6000	.7714	.8286	.8857	.9429	
7	.5357	.6786	.7450	.8571	.8929	.9643
8	.5000	.6190	.7143	.8095	.8571	.9286
9	.4667	.5833	.6833	.7667	.8167	.9000
10	.4424	.5515	.6364	.7333	.7818	.8667
11	.4182	.5273	.6091	.7000	.7455	.8364
12	.3986	.4965	.5804	.6713	.7273	.8182
13	.3791	.4780	.5549	.6429	.6978	.7912
14	.3626	.4593	.5341	.6220	.6747	.7670
15	.3500	.4429	.5179	.6000	.6536	.7464
16	.3382	.4265	.5000	.5824	.6324	.7265
17	.3260	.4118	.4853	.5637	.6152	.7083
18	.3148	.3994	.4716	.5480	.5975	.6904
19	.3070	.3895	.4579	.5333	.5825	.6737
20	.2977	.3789	.4451	.5203	.5684	.6586
21	.2909	.3688	.4351	.5078	.5545	.6455
22	.2829	.3597	.4241	.4963	.5426	.6318
23	.2767	.3518	.4150	.4852	.5306	.6186
24	.2704	.3435	.4061	.4748	.5200	.6070
25	.2646	.3362	.3977	.4654	.5100	.5962
26	.2588	.3299	.3894	.4564	.5002	.5856
27	.2540	.3236	.3822	.4481	.4915	.5757
28	.2490	.3175	.3749	.4401	.4828	.5660
29	.2443	.3113	.3685	.4320	.4744	.5567
30	.2400	.3059	.3620	.4251	.4665	.5479

For *n* greater than 30 the approximate quantiles of *p* may be obtained from

$$w_p \cong \frac{x_p}{\sqrt{n-1}}$$

where x_p is the *p* quantile of a standard normal random variable obtained from Table 1.

SOURCE. Adapted from Glasser and Winter (1961), with corrections.

^a The entries in this table are selected quantiles w_p of the Spearman rank correlation coefficient *p* when used as a test statistic. The lower quantiles may be obtained from the equation

$$w_p = -w_{1-p}$$

The critical region corresponds to values of *p* smaller than (or greater than) but not including the appropriate quantile. Note that the median of *p* is 0.