

---

**UNIVERSITI SAINS MALAYSIA**

Supplementary Examination

Academic Session 2008/2009

June 2009

**BOI 109/4 – Biostatistics  
[Biostatistik]**

Duration: 3 hours  
[Masa : 3 jam]

---

Please ensure that this examination paper contains TWENTY ONE printed pages before you begin the examination.

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

**Instructions:** Answer **FIVE** (5) out of **SIX** (6) questions, in English or Bahasa Malaysia. Each question carries 20 marks.

**Arahan:** Jawab **LIMA** (5) daripada **ENAM** (6) soalan yang diberikan dalam Bahasa Inggeris atau Bahasa Malaysia. Tiap-tiap soalan bernilai 20 markah.]

- 2 -

1. [a] Recently, there have been concerns about the effects of phthalates on the development of male reproductive system in mammals. Phthalates are common ingredients in many plastics. In a pilot study, a researcher gave pregnant rats daily doses of 750 mg/kg of body weight of DEHP (di-2-ethylhexyl phthalate) throughout the period of sexual organ development. The newly born male rat pups were sacrificed and their seminal vesicles were dissected and weighed. Below are the weights of the seminal vesicles (in mg):

1630 1580 1670 1350 1650 1600 1710.

If mean weight of seminal vesicles of untreated newborn males is 1700 mg, can you say that rats exposed to the DHEP have a lighter seminal vesicles?

(10 marks)

- [b] As part of a benthic community survey of the Marine Research Centre in Muka Head, 15 sea stars, *Asteropectens* sp., were collected and their longest arm was measured (cm) and presented as below. Assume normality for this population.

10.3 11.1 10.5 10.0 11.3 14.5 13.0 12.1 9.4 11.3 12.0

11.5 9.3 10.1 7.6

- [i] Calculate the sample mean and variance.

(4 marks)

- [ii] Calculate the 95% confidence interval for the population mean.

(4 marks)

- [iii] Explain what is the meaning of "confidence interval".

(2 marks)

- 3 -

2. An experiment was conducted to compare the yield of two rice varieties, A and B. Fourteen rice plots were randomly chosen in one rice farm in Kedah, and the yields in metric tons per hectare in each plot are as follows:

Plot	1	2	3	4	5	6	7
Variety A	4.6	4.8	3.2	4.7	4.3	3.7	4.1
Variety B	4.1	4.0	3.5	4.1	4.5	3.3	3.8

- [a] Why do you think both varieties were tested in 14 different plots in the same farm rather than testing variety A in 7 plots in one farm while the 7 plots for variety B in different farms?

(5 marks)

- [b] Carry out a hypothesis test to decide whether the mean yields of the two varieties are the same.

(10 marks)

- [c] Explain the meaning of the words "random samples".

(5 marks)

3. As part of a review of Malaysian healthcare websites, a researcher examined 60 different websites. Of these, 46 websites contained information intended for general public. The Department of Health in Malaysia recommends a reading age at or below 12 - 14 years for health information websites. The following table provides the reading age and frequency of the websites.

Reading age	8	9	10	11	12	13	14	15	16	17
Frequency	1	0	1	2	1	4	2	7	4	24

- [a] Find the mean reading age and standard deviation.

- [b] Make a bar graph.

- 4 -

- [c] From your graph, does it appear to you that the reading age for these websites is at or below 14 years of age?
- [d] Calculate 95% confidence interval for the reading age. Does the confidence interval support your impression from part (c) above?

(20 marks)

4. A student has carried out an experiment to study the effect of two types of bacteria on lymphocyte counts on rats. The student has obtained the rats of the same age from seven different sources. He has randomly taken three rats from each source and each rat was assigned each treatment, that is either K (control), A (treated with bacteria A) and B (treated with bacteria B). Results obtained (millions units of lymphocytes per cm<sup>3</sup> blood) are stated below.

Treatment	Rat from different Sources						
	1	2	3	4	5	6	7
K	5.1	5.5	6.9	4.0	3.9	5.6	7.0
A	5.4	5.4	7.0	3.9	3.5	5.8	7.6
B	6.0	6.8	6.5	4.8	5.5	6.4	9.0

Analyse the data to determine whether the number of lymphocytes in rats are affected significantly by the type of bacteria. Does the rats taken from different sources response differently to the bacteria treatment? Carry out mean comparison using LSD method on the mean lymphocytes number for the different treatment.

(20 marks)

- 5 -

5. The following data are results obtained from a study on the yield of paddy after application of a fertilizer.

Amount of fertilizer (kg/ha)	Paddy yield (kg/ha)
0	154
50	278
75	341
100	402
150	514

- [a] What is the difference between correlation and regression?

(4 marks)

- [b] If you are interested to know the relationship between the amount of fertilizer application and the yield of paddy, what statistical method will you use?

(2 marks)

- [c] If you are interested to know the paddy yield with application of 125kg/ha of fertilizer, which statistical method will you use?

(2 marks)

- [d] Determine the linear regression equation with the data given above.

(8 marks)

- [e] Do you expect the paddy yield to increase linearly if the amount of fertilizer is continuously increased, say to a higher level of 200, 500, 1000 kg/ha? Give your reason.

(4 marks)

- 6 -

6. [a] A study was carried out to determine whether the tail length (cm) of 6 male cats is similar with the tail length of 6 female cats. The following data were found to be not normally distributed. Analyse the data and determine whether there is any difference in the tail length between the male and female cats.

Male	Female
74	80
77	78
73	77
75	84
72	82
71	79

(10 marks)

- [b] When pink *Melastoma malabatricum* are crossed, it is expected that the F<sub>2</sub> frequencies will be in the order of 1 red: 2 pink: 1 white. A researcher obtained an actual ratio of 30 red: 48 pink: 27 white. Are these results consistent with the genetic model?

(10 marks)

1. [a] Baru-baru ini terdapat keperihatinan tentang kesan ftalat terhadap perkembangan sistem pembiakan organ jantan pada haiwan mamalia. Ftalat ialah sebatian yang sering terdapat pada banyak bahan plastik. Dalam satu kajian rintis, seorang pengkaji memberikan tikus hamil dengan satu dos harian sebanyak 750 mg DEHP (di-2-etilheksil ftalat) per kg berat badan selama perkembangan organ seks pada janin anak tikus. Pembedahan dilakukan terhadap anak tikus yang baru lahir dengan mengambil seminal vesikel dan ditimbang. Berikut ialah berat seminal vesikel anak tikus jantan tersebut (dalam mg):

1630 1580 1670 1350 1650 1600 1710.

Jika min berat seminal vesikel anak tikus jantan yang tidak diperlakukan ialah 1700 mg, apakah anda boleh mengatakan bahawa tikus yang didedahkan dengan DHEP menghasilkan seminal vesikel yang lebih ringan?

(10 markah)

- [b] Sebagai sebahagian daripada kajian tinjauan komuniti bentik di Pusat Penyelidikan Marin di Muka Head, 15 tapak Sulaiman, *Asteropectens* sp., dikutip dan lengan terpanjangnya diukur (cm) seperti yang dikemukakan di bawah ini. Anggap populasinya bertaburan normal.

10.3 11.1 10.5 10.0 11.3 14.5 13.0 12.1 9.4 11.3 12.0

11.5 9.3 10.1 7.6

- [i] Hitung min dan varian sampel tersebut.

(4 markah)

- [ii] Hitung selang keyakinan 95% untuk min populasi.

(4 markah)

- [iii] Terangkan apakah yang dimaksudkan dengan "selang keyakinan".

(2 markah)

- 8 -

2. Satu eksperimen telah dilakukan untuk membandingkan hasil dua varieti padi, A dan B. Tujuh petak sawah padi dipilih secara rawak dan hasil setiap petak untuk setiap varieti dalam ton metrik adalah seperti berikut:

Petak	1	2	3	4	5	6	7
Varieti A	4.6	4.8	3.2	4.7	4.3	3.7	4.1
Varieti B	4.1	4.0	3.5	4.1	4.5	3.3	3.8

- [a] Menurut pendapat anda, mengapakah kedua-dua varieti tersebut diuji dalam 14 petak yang berbeza pada sawah yang sama tetapi tidak pula varieti A dalam 7 petak pada sawah yang sama manakala 7 petak untuk varieti B pula pada sawah yang berbeza?

(5 markah)

- [b] Lakukan suatu ujian hipotesis untuk menentukan sama ada min hasil padi kedua varieti tersebut adalah sama?

(10 markah)

- [c] Huraikan erti perkataan "sampel rawak".

(5 markah)

3. Sebagai sebahagian daripada penilaian terhadap Laman Web Kesihatan di Malaysia, seorang pengkaji memeriksa 60 laman web yang berbeza. Daripada bilangan ini, 46 laman web mengandungi keterangan yang ditujukan kepada khalayak ramai. Jabatan Kesihatan Malaysia telah membuat perakuan bahawa umur pembaca pada atau di bawah 12 - 14 tahun adalah sasaran laman web tersebut. Jadual berikut mengandungi umur pembaca dan kekerapan laman web yang berkenaan.

Umur membaca	8	9	10	11	12	13	14	15	16	17
Kekerapan	1	0	1	2	1	4	2	7	4	24

- [a] Hitung min umur pembaca dan sisihan piawai.
- [b] Buat satu graf palang.
- [c] Daripada graf tersebut, apakah ada bukti yang menunjukkan bahawa umur pembaca untuk laman web tersebut yang bersesuaian ialah pada atau di bawah umur 14 tahun?
- [d] Hitung selang keyakinan 95% untuk umur pembaca. Apakah selang keyakinan ini menyokong pemerhatian anda pada (c) di atas?

(20 markah)

-10-

4. Seorang pelajar telah menjalankan satu eksperimen untuk mengkaji kesan dua jenis bakteria terhadap bilangan limfosit pada tikus. Pelajar tersebut mengambil tikus yang sama umur daripada tujuh sumber yang berlainan. Dia mengambil secara rawak tiga ekor tikus daripada setiap sumber dan setiap ekor diagih satu olahan, iaitu sama ada K (Kawalan), A (dirawat dengan bakteria A) dan B (dirawat dengan bakteria B). Keputusan (ribu unit limfosit per cm<sup>3</sup> darah) yang diperoleh adalah seperti berikut.

Olahan	Tikus dari sumber berlainan						
	1	2	3	4	5	6	7
K	5.1	5.5	6.9	4.0	3.9	5.6	7.0
A	5.4	5.4	7.0	3.9	3.5	5.8	7.6
B	6.0	6.8	6.5	4.8	5.5	6.4	9.0

Jalankan analisis data untuk menetukan sama ada bilangan limfosit dalam tikus dipengaruhi oleh bakteria secara bererti. Adakah tikus yang diambil daripada sumber yang berlainan memberi kesan yang berbeza kepada rawatan bakteria? Lakukan perbandingan min bilangan limfosit bagi rawatan yang berbeza dengan kaedah LSD.

(20 markah)

5. Data berikut adalah keputusan yang diperolehi daripada satu kajian tentang hasil padi selepas aplikasi baja.

Kadar pembajaan (kg/ha)	Hasil padi (kg/ha)
0	154
50	278
75	341
100	402
150	514

- [a] Apakah bezanya antara korelasi dan regresi linear?

(4 markah)

- [b] Sekiranya anda ingin mengetahui pertalian antara jumlah baja yang diguna dan hasil padi, apakah kaedah statistik yang anda akan gunakan?

(2 markah)

-11-

- [c] Sekiranya anda ingin mengetahui hasil padi pada 125 kg/ha baja yang diguna, apakah kaedah statistik yang anda akan gunakan?

(2 markah)

- [d] Dapatkan persamaan regresi linear dengan data yang diberi di atas.

(8 markah)

- [e] Adakah anda jangka bahawa hasil padi akan turut meningkat secara linear jika kadar pembajaan ditingkatkan kepada aras yang lebih tinggi, 200, 500, 1000 kg/ha? Berikan alasan anda.

(4 markah)

6. [a] Satu kajian telah dijalankan untuk menentu sama ada panjang ekor (cm) bagi 6 ekor kucing jantan sama dengan 6 ekor kucing betina. Data berikut yang dikutip didapati tidak bertabur secara normal. Lakukan analisis data dan tentukan sama ada terdapat perbezaan dalam panjang ekor antara kucing jantan dan betina.

Jantan	Betina
74	80
77	78
73	77
75	84
72	82
71	79

(10 markah)

- [b] Apabila *Melastoma malabatricum* berwarna merah jambu dikacukkan, telah dijangka bahawa frekuensi F2 akan berada dalam nisbah 1 merah: 2 merah jambu: 1 putih. Seorang ahli penyelidik mendapatkan nisbah sebenar iaitu 30 merah: 48 merah jambu: 27 putih. Adakah keputusan ini selaras dengan model genetik itu?

(10 markah)

Useful formula

A.

$$r = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{\left[ n \sum_i X_i^2 - \left( \sum_i X_i \right)^2 \right] \left[ n \sum_i Y_i^2 - \left( \sum_i Y_i \right)^2 \right]}}$$

B.

$$t = \frac{(\bar{y}_1 - \bar{y}_2)}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

C.

$$= \frac{\bar{y}_1 - \bar{y}_2}{s_{y_1 - y_2}}$$

D.

$$t = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{\left( \frac{s_1^2}{n_1} \right) + \left( \frac{s_2^2}{n_2} \right)}}$$

E.

$$z = \frac{\bar{y} - 0.5n}{\sqrt{0.25n}}$$

F.

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} = z$$

G.

$$F = \frac{s_1^2}{s_2^2}$$

H.

$$z = \frac{\bar{y} - \mu}{s_y}$$





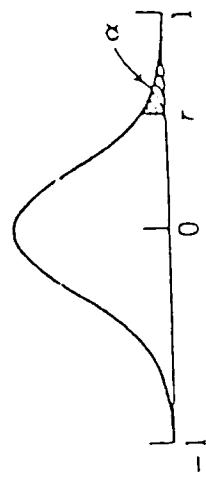






## Critical values Table for Pearson correlation constant, r

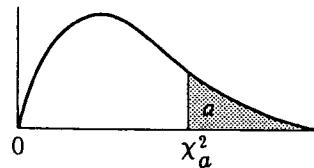
For two-tail test at  $\alpha = 0.05$ , read at  $\alpha = 0.025$



$\alpha \backslash n$	0.05	0.025	0.010	0.005
5	0.805	0.878	0.934	0.959
6	0.729	0.811	0.882	0.917
7	0.669	0.754	0.833	0.875
8	0.621	0.707	0.789	0.834
9	0.582	0.666	0.750	0.798
10	0.549	0.632	0.716	0.765
11	0.521	0.602	0.685	0.735
12	0.497	0.576	0.658	0.708
13	0.476	0.553	0.634	0.684
14	0.457	0.532	0.612	0.661
15	0.441	0.514	0.592	0.641
16	0.426	0.497	0.574	0.623

$\alpha \backslash n$	0.05	0.025	0.010	0.005
17	0.412	0.482	0.558	0.606
18	0.400	0.468	0.542	0.590
19	0.389	0.456	0.528	0.575
20	0.378	0.444	0.516	0.561
25	0.337	0.396	0.462	0.505
30	0.306	0.361	0.423	0.463
40	0.264	0.312	0.366	0.402
50	0.235	0.279	0.328	0.361
60	0.214	0.254	0.300	0.330
80	0.185	0.220	0.260	0.286
100	0.165	0.196	0.232	0.256

*Percentage Points of the Chi-Square  
Distribution*



<i>df</i>	<i>a</i> = .995	<i>a</i> = .990	<i>a</i> = .975	<i>a</i> = .950	<i>a</i> = .900
1	0.0000393	0.0001571	0.0009821	0.0039321	0.0157908
2	0.0100251	0.0201007	0.0506356	0.102587	0.210720
3	0.0717212	0.114832	0.215795	0.351846	0.584375
4	0.206990	0.297110	0.484419	0.710721	1.063623
5	0.411740	0.554300	0.831211	1.145476	1.61031
6	0.675727	0.872085	1.237347	1.63539	2.20413
7	0.989265	1.239043	1.68987	2.16735	2.83311
8	1.344419	1.646482	2.17973	2.73264	3.48954
9	1.734926	2.087912	2.70039	3.32511	4.16816
10	2.15585	2.55821	3.24697	3.94030	4.86518
11	2.60321	3.05347	3.81575	4.57481	5.57779
12	3.07382	3.57056	4.40379	5.22603	6.30380
13	3.56503	4.10691	5.00874	5.89186	7.04150
14	4.07468	4.66043	5.62872	6.57063	7.78953
15	4.60094	5.22935	6.26214	7.26094	8.54675
16	5.14224	5.81221	6.90766	7.96164	9.31223
17	5.69724	6.40776	7.56418	8.67176	10.0852
18	6.26481	7.01491	8.23075	9.39046	10.8649
19	6.84398	7.63273	8.90655	10.1170	11.6509
20	7.43386	8.26040	9.59083	10.8508	12.4426
21	8.03366	8.89720	10.28293	11.5913	13.2396
22	8.64272	9.54249	10.9823	12.3380	14.0415
23	9.26042	10.19567	11.6885	13.0905	14.8479
24	9.88623	10.8564	12.4011	13.8484	15.6587
25	10.5197	11.5240	13.1197	14.6114	16.4734
26	11.1603	12.1981	13.8439	15.3791	17.2919
27	11.8076	12.8786	14.5733	16.1513	18.1138
28	12.4613	13.5648	15.3079	16.9279	18.9392
29	13.1211	14.2565	16.0471	17.7083	19.7677
30	13.7867	14.9535	16.7908	18.4926	20.5992
40	20.7065	22.1643	24.4331	26.5093	29.0505
50	27.9907	29.7067	32.3574	34.7642	37.6886
60	35.5346	37.4848	40.4817	43.1879	46.4589
70	43.2752	45.4418	48.7576	51.7393	55.3290
80	51.1720	53.5400	57.1532	60.3915	64.2778
90	59.1963	61.7541	65.6466	69.1260	73.2912
100	67.3276	70.0648	74.2219	77.9295	82.3581

*Percentage Points of the Chi-Square Distribution*

Table 3 (continued)

$\alpha = .10$	$\alpha = .05$	$\alpha = .025$	$\alpha = .010$	$\alpha = .005$	$df$
2.70554	3.84146	5.02389	6.63490	7.87944	1
4.60517	5.99147	7.37776	9.21034	10.5966	2
6.25139	7.81473	9.34840	11.3449	12.8381	3
7.77944	9.48773	11.1433	13.2767	14.8602	4
9.23635	11.0705	12.8325	15.0863	16.7496	5
10.6446	12.5916	14.4494	16.8119	18.5476	6
12.0170	14.0671	16.0128	18.4753	20.2777	7
13.3616	15.5073	17.5346	20.0902	21.9550	8
14.6837	16.9190	19.0228	21.6660	23.5893	9
15.9871	18.3070	20.4831	23.2093	25.1882	10
17.2750	19.6751	21.9200	24.7250	26.7569	11
18.5494	21.0261	23.3367	26.2170	28.2995	12
19.8119	22.3621	24.7356	27.6883	29.8194	13
21.0642	23.6848	26.1190	29.1413	31.3193	14
22.3072	24.9958	27.4884	30.5779	32.8013	15
23.5418	26.2962	28.8454	31.9999	34.2672	16
24.7690	27.5871	30.1910	33.4087	35.7185	17
25.9894	28.8693	31.5264	34.8053	37.1564	18
27.2036	30.1435	32.8523	36.1908	38.5822	19
28.4120	31.4104	34.1696	37.5662	39.9968	20
29.6151	32.6705	35.4789	38.9321	41.4010	21
30.8133	33.9244	36.7807	40.2894	42.7956	22
32.0069	35.1725	38.0757	41.6384	44.1813	23
33.1963	36.4151	39.3641	42.9798	45.5585	24
34.3816	37.6525	40.6465	44.3141	46.9278	25
35.5631	38.8852	41.9232	45.6417	48.2899	26
36.7412	40.1133	43.1944	46.9630	49.6449	27
37.9159	41.3372	44.4607	48.2782	50.9933	28
39.0875	42.5569	45.7222	49.5879	52.3356	29
40.2560	43.7729	46.9792	50.8922	53.6720	30
51.8050	55.7585	59.3417	63.6907	66.7659	40
63.1671	67.5048	71.4202	76.1539	79.4900	50
74.3970	79.0819	83.2976	88.3794	91.9517	60
85.5271	90.5312	95.0231	100.425	104.215	70
96.5782	101.879	106.629	112.329	116.321	80
107.565	113.145	118.136	124.116	128.299	90
118.498	124.342	129.561	135.807	140.169	100

From "Tables of the Percentage Points of the  $\chi^2$ -Distribution." *Biometrika*, Vol. 32 (1941), pp. 188-189, by Catherine M. Thompson. Reproduced by permission of the *Biometrika* Trustees.

