UNIVERSITI SAINS MALAYSIA

Second Semester Examination
Academic Session 2008/2009

April/Mei 2009

BOI 109/4 – Biostatistics
[Biostatistik]

Duration: 3 hours
[Masa : 3 jam]

Please ensure that this examination paper contains <u>SIXTEEN</u> printed pages before you begin the examination.

[Sila pastikan bahawa kertas peperiksaan ini mengandungi <u>ENAM BELAS</u> muka surat yang bercetak sebelum anda memulakan peperiksaan ini.]

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

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

1. [a] A new systemic fungicide was tested against sigatoka disease on mature banana plants. Nine mature banana plants were selected in Balik Pulau and the fungicide was sprayed twice in two weeks intervals. Disease severities from a scale of 0 (no disease symptoms) to 10 (nearly 100% infection) were measured before and after the sprayings. Based on these data, conduct a statistical analysis to determine whether the fungicide is effective in curing the disease ($\alpha = 0.01$).

(11 marks)

Spraying	Disease severity scale on 9 mature banana plants								
	1	2	3	4	5	6	7	8	9
Before	6.5	6.2	5.9	5.6	6.5	6.9	5.8	6.3	6.5
After	6.3	6.1	6.1	5.7	6.4	6.7	5.9	6.1	6.3

[b] Write a short note on each of the following:

[i] Poisson Distribution.

[ii] Binomial Distribution.

[iii] Stratified Random Sampling.

(9 marks)

2. A researcher from Agriculture Institute conducted a project in an exotica papaya farm in Pulau Pinang to determine the size of marketable papaya fruits. He sampled 99 marketable papaya fruits from the farm and found that the mean weight and standard deviation were 314.7 and 52.8 g, respectively. If we randomly pick up a marketable papaya fruit from the farm,

[i] What is the probability of having the weight of the papaya fruit between 305-325 g?

(5 marks)

[ii]   Calculate 95% confidence interval of the min weight of papaya fruit.

(6 marks)

[iii]   His data also showed that the probability in having a papaya fruit with more than 300 g is 53%. If we randomly pick 5 marketable fruits from the farm, what is the probability in having 4 fruits with more than 300 g each?

(4 marks)

[iv]   Write a short note on "Standard Normal Curve". What is the main difference between the Standard Normal Curve and other normal curves?

(5 marks)

3.   [a]   Data from a research conducted earlier showed that the organic content in the alluvial soil in Kuala Juru was 268 mg/kg with a standard deviation of 75 mg/kg. If another researcher is interested to know the interval estimate of the mean organic content in the alluvial soil from that area with 50 mg/kg accuracy, how many samples are required? ($\alpha = 0.05$).

(6 marks)

[b] In a study, a researcher wanted to know if patients with primary hypertension (PH) have higher total cholesterol levels than normotensive (NT) patients. For his study, he was able to get 21 PH patients and 23 NT patients. Total cholesterol level (mg/dl) for these patients are as follows. Run an appropriate statistical test to show whether or not the patients with primary hypertension (PH) have higher total cholesterol levels than normotensive (NT) patients.

(14 marks)

Total cholesterol level (mg/dl)					Normotensive Patients (NT)			
Primary Hypertensive Patients (PH)								
207	208	226	202		286	196	203	163
172	199	221	218		226	168	189	196
191	185	223	216		187	229	196	189
221	235	181			204	184	142	142
203	214	217			203	186	179	168
241	134	208			206	211	212	

4. An experiment was carried out to study the influence of bleeding time, factor A, and diethylstilbestrol treatment (an estrogenic compound), factor B, on plasma phospholipid in goats. Five goats were assigned to each of the treatment groups. The treatment combinations are for morning and afternoon bleeding with and without diethylstilbestrol treatment. The data obtained are as follows:

Diethylstilbestrol factor	Bleeding time	
	morning	afternoon
Untreated	8.53	39.14
	20.53	26.20
	12.53	31.33
	14.00	45.80
	10.80	40.20
treated	17.53	32.00
	21.07	23.80
	20.80	28.87
	17.33	25.06
	20.07	29.33

[a] List all treatment combinations.

(2 marks)

[b] Analyze the data and conclude your findings.

(12 marks)

[c] Graphically show the interaction response between the two factors.

(6 marks)

5. A plant physiologist studied the relationship between the photosynthesis rate of sentang (*Azadirachta excelsa*) and its relation to stomata size (μm). Use the data given below :

Plant	Stomata size (μm.)	Photosynthesis rate
1	31	280
2	34	331
3	35	348
4	38	360
5	41	402
6	48	439
7	52	498
8	58	526

[a] Name the dependent and independent variables

(2 marks)

[b] Compute the correlation coefficient of the two variables.

(5 marks)

[c] Determine the regression equation relating photosynthetic rate and stomata size.

(10 marks)

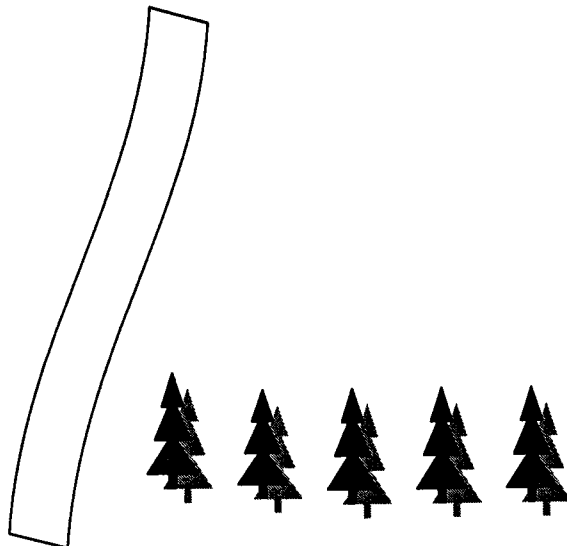[d] Determine the photosynthesis rate when the leaf stomata size is 45μm.

(3 marks)

6. [a] In order to determine the mode of distribution of a certain species of cricket, a total of 200 quadrates were randomly thrown on a field. Each quadrate was examined for crickets, and the following data were obtained in terms of the cricket count.

No. of Quadrates	Number of crickets/quadrate
98	0
52	1
28	2
22	3

Analyze these data to determine whether the crickets are randomly dispersed throughout the area.

(10 marks)

[b] Four varieties of corn (*Zea mays*) (A, B, C, D) were grown in the fields to determine the effects of different varieties on the yield (kg/plot). The researcher realised that there was a stream flowing by the side of the study plots and there were big trees lining the end of the plots. Based on the information given, with an aid of a diagram how you will assign the different varieties of corn on the study plots.

(10 marks)

1. [a] Satu fungisid sistemik yang baru telah diuji terhadap penyakit sigatoka pada pohon pisang dewasa. Sembilan pohon pisang dewasa dipilih di Balik Pulau dan fungisid tersebut disembur dua kali dalam selang masa dua minggu. Keparahan penyakit daripada skala 0 (tiada gejala penyakit) hingga 10 (hampir 100% jangkitan) sebelum dan setelah semburan. Berdasarkan data ini, lakukan satu analisis statistik untuk membuktikan sama ada fungisid ini berkesan dalam menyembuhkan penyakit ini (α = 0.01).

(11 markah)

Penyemburan	Skala keparahan penyakit pada 9 pohon pisang dewasa								
	1	2	3	4	5	6	7	8	9
Sebelum	6.5	6.2	5.9	5.6	6.5	6.9	5.8	6.3	6.5
Setelah	6.3	6.1	6.1	5.7	6.4	6.7	5.9	6.1	6.3

[b] Tulis satu nota ringkas tentang setiap satu yang berikut:

[i] Taburan Poisson.

[ii] Taburan Binomial.

[iii] Pensampelan rawak berstratum.

(9 markah)

2. Seorang penyelidik dari Institut Pertanian telah menjalankan projek di sebuah dusun betik eksotika di Pulau Pinang untuk menentukan saiz betik yang boleh dipasarkan. Dia mengambil 99 sampel buah betik yang sedia untuk dipasarkan daripada dusun tersebut dan mendapati min berat dan sisihan piawai masing-masing ialah 314.7 dan 52.8 g. Jika kita mengambil sebiji betik yang sedia untuk dipasarkan secara rawak dari dusun tersebut,

(i) Apakah kebarangkalian untuk berat buah betik tersebut terletak di antara 305-325 g?

(5 markah)

(ii) Hitung selang keyakinan 95% min berat buah betik berkenaan.

(6 markah)

[iii]   Datanya juga menunjukkan bahawa kebarangkalian memperolehi berat buah betik yang melebihi 300 g ialah 53%. Jika anda mengambil 5 biji betik tersebut secara rawak, apakah kebarangkalian 4 biji melebihi berat 300 g setiap satu?

(4 markah)

[iv]   Tulis satu nota ringkas mengenai "Lengkuk Normal Piawai". Apakah perbezaan utama antara Lengkuk Normal Piawai dengan lengkuk normal yang lain?

(5 markah)

3.   [a]   Melalui kajian awal, min kandungan bahan organik di dalam tanah aluvium di Kuala Juru ialah 268 mg/kg dengan sisihan piawai 75 mg/kg. Jika seorang penyelidik lain ingin menghitung anggaran   selang bagi min kandungan bahan organik di kawasan tersebut dengan nilai ketepatan 50 mg/kg, berapa banyakkah sampel yang diperlukan? ($\alpha = 0.05$).

(6 markah)

[b]   Dalam satu kajian, seorang penyelidik ingin tahu sama ada pesakit hipertensi primer (PH) mempunyai aras jumlah kolesterol yang lebih tinggi daripada pesakit normotensi (NT). Untuk kajiannya, dia berjaya mendapat 21 pesakit PH dan 23 pesakit NT. Aras jumlah kolesterol (mg/dl) pesakit-pesakit tersebut adalah seperti berikut. Jalankan satu ujian statistik yang sesuai untuk menunjukkan sama ada pesakit hipertensi primer mempunyai aras jumlah kolesterol yang lebih tinggi daripada pesakit normotensi (NT).

(14 markah)

Aras Jumlah kolesterol (mg/dl)								
Pesakit hipertensi primer (PH)					Pesakit normotensi (NT)			
207	208	226	202		286	196	203	163
172	199	221	218		226	168	189	196
191	185	223	216		187	229	196	189
221	235	181			204	184	142	142
203	214	217			203	186	179	168
241	134	208			206	211	212	

4.   Satu eksperimen telah dijalankan untuk mengkaji kesan masa mengambil darah, faktor A, dan olahan dietilstibesterol (suatu sebatian estrogenik), faktor B, ke atas fosfolipid plasma kambing. Lima ekor kambing diagihkan kepada setiap kumpulan olahan. Kombinasi olahan adalah masa mengambil darah iaitu pagi dan petang dengan atau tanpa olahan dietilstibesterol. Data yang diperolehi adalah seperti berikut:-

Faktor Dietilstilbesterol	Masa Mengambil Darah	
	Pagi	petang
Tanpa olahan	8.53	39.14
	20.53	26.20
	12.53	31.33
	14.00	45.80
	10.80	40.20
Dengan olahan	17.53	32.00
	21.07	23.80
	20.80	28.87
	17.33	25.06
	20.07	29.33

[a]   Senarai semua kombinasi olahan.

(2 markah)

[b]   Jalankan analisis data dan simpulkan keputusan kajian.

(12 markah)

[c]   Dengan menggunakan graf, tunjukkan interaksi gerakan antara dua faktor dalam bentuk graf.

(6 markah)

5. Seorang ahli fisiologi tumbuhan mengkaji kadar fotosintesis sentang (*Azadirachta excelsa*) dan pertaliannya dengan saiz stomata (μm). Gunakan data berikut:

Tumbuhan	Saiz Stomata (μm.)	Kadar fotosintesis
1	31	280
2	34	331
3	35	348
4	38	360
5	41	402
6	48	439
7	52	498
8	58	526

[a]   Namakan pembolehubah bersandar dan tak bersandar.

(2 markah)

[b]   Hitungkan pekali korelasi bagi dua pembolehubah tersebut.

(5 markah)

[c]   Tentukan persamaan regresi yang mengaitkan kadar fotosintesis dan saiz stomata daun.

(10 markah)

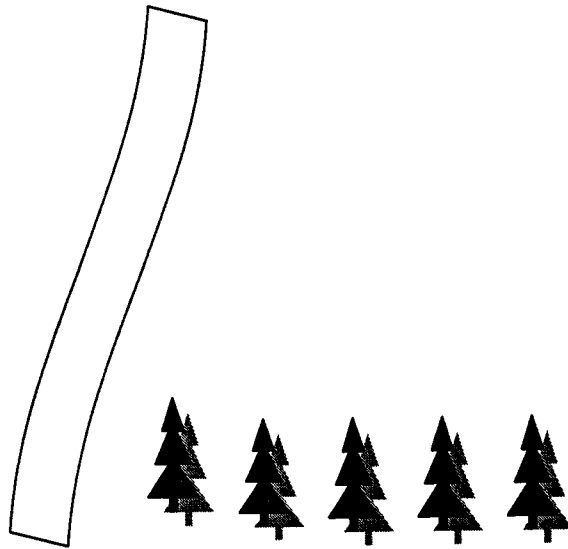[d]   Tentukan kadar fotosintesis apabila saiz stomata ialah 45 μm.

(3 markah)

6. [a] Untuk menentukan corak pertaburan sesuatu spesies cangkerik, sejumlah 200 kuadrat dilontar secara rawak ke padang. Setiap kuadrat diperhati untuk kehadiran cangkerik, dan data berikut diperolehi berdasar kepada bilangan cangkerit yang didapati :

Bilangan Kuadrat	Bilangan cangkerit/kuadrat
98	0
52	1
28	2
22	3

Jalankan analisis data di atas untuk menentukan sama ada cangkerik bertabur secara rawak di seluruh kawasan itu.

(10 markah)

[b]   Empat varieti jagung Zea mays) (A, B, C, D) ditanam untuk menentukan kesan varieti yang berbeza terhadap hasilnya (kg/plot). Ahli penyelidik mendapati terdapat satu sungai kecil di tepi plot kajian dan satu barisan pokok besar di hujung plot kajian. Berdasar maklumat ini, dengan bantuan gambarajah tunjukkan bagaimana anda akan mengagihkan varieti jagung yang berbeza untuk ditanam di plot kajian.

(10 markah)

- oooOooo -

Useful formula

A.
$$r = \dfrac{n\sum_i x_i y_i - \sum_i x_i \sum_i y_i}{\sqrt{\left[n\sum_i x_i^2 - \left(\sum_i x_i\right)^2\right]\left[n\sum_i y_i^2 - \left(\sum_i y_i\right)^2\right]}}$$

B.
$$t = \dfrac{\left(\overline{y_1} - \overline{y_2}\right)}{s\sqrt{\left(\dfrac{1}{n_1}\right) + \left(\dfrac{1}{n_2}\right)}}$$

C.
$$= \dfrac{\overline{y_1} - \overline{y_2}}{s y_1 - y_2}$$

D.
$$t = \dfrac{\left(\overline{y_1} - \overline{y_2}\right)}{\sqrt{\left(\dfrac{s_1^2}{n_1}\right) + \left(\dfrac{s_2^2}{n_2}\right)}}$$

E.
$$z = \dfrac{\overline{y} - 0.5n}{\sqrt{0.25n}}$$

F.
$$t = \dfrac{\overline{d}}{\dfrac{s_d}{\sqrt{n}}} = z$$

G.
$$F = \dfrac{s_1^2}{s_2^2}$$

H.
$$z = \dfrac{\overline{y} - \mu}{s_{\overline{y}}}$$

Critical values for Duncan's multiple range test*

Least significant studentized ranges for testing p successive values out of a linearly ordered arrangement of k sample means from a normal population with v degrees of freedom.

	$\alpha = 0.05$						$\alpha = 0.01$				
v \ p	2	3	4	5	6	v \ p	2	3	4	5	6
1	17.97	17.97	17.97	17.97	17.97	1	90.03	90.03	90.03	90.03	90.03
2	6.085	6.085	6.085	6.085	6.085	2	14.04	14.04	14.04	14.04	14.04
3	4.501	4.516	4.516	4.516	4.516	3	8.261	8.321	8.321	8.321	8.321
4	3.927	4.013	4.033	4.033	4.033	4	6.512	6.677	6.740	6.756	6.756
5	3.635	3.749	3.797	3.814	3.814	5	5.702	5.893	5.989	6.040	6.065
6	3.461	3.587	3.649	3.680	3.694	6	5.243	5.439	5.549	5.614	5.655
7	3.344	3.477	3.548	3.588	3.611	7	4.949	5.145	5.260	5.334	5.383
8	3.261	3.399	3.475	3.521	3.549	8	4.746	4.939	5.057	5.135	5.189
9	3.199	3.339	3.420	3.470	3.502	9	4.596	4.787	4.906	4.986	5.043
10	3.151	3.293	3.376	3.430	3.465	10	4.482	4.671	4.790	4.871	4.931
11	3.113	3.256	3.342	3.397	3.435	11	4.392	4.579	4.697	4.780	4.841
12	3.082	3.225	3.313	3.370	3.410	12	4.320	4.504	4.622	4.706	4.767
13	3.055	3.200	3.289	3.348	3.389	13	4.260	4.442	4.560	4.644	4.706
14	3.033	3.178	3.268	3.329	3.372	14	4.210	4.391	4.508	4.591	4.654
15	3.014	3.160	3.250	3.312	3.356	15	4.168	4.347	4.463	4.547	4.610
16	2.998	3.144	3.235	3.298	3.343	16	4.131	4.309	4.425	4.509	4.572
17	2.984	3.130	3.222	3.285	3.331	17	4.099	4.275	4.391	4.475	4.539
18	2.971	3.118	3.210	3.274	3.321	18	4.071	4.246	4.362	4.445	4.509
19	2.960	3.107	3.199	3.264	3.311	19	4.046	4.220	4.335	4.419	4.483
20	2.950	3.097	3.190	3.255	3.303	20	4.024	4.197	4.312	4.395	4.459
24	2.919	3.066	3.160	3.226	3.276	24	3.956	4.126	4.239	4.322	4.386
30	2.888	3.035	3.131	3.199	3.250	30	3.889	4.056	4.168	4.250	4.314
40	2.858	3.006	3.102	3.171	3.224	40	3.825	3.988	4.098	4.180	4.244
60	2.829	2.976	3.073	3.143	3.198	60	3.762	3.922	4.031	4.111	4.174
120	2.800	2.947	3.045	3.116	3.172	120	3.702	3.858	3.965	4.044	4.107
∞	2.772	2.918	3.017	3.089	3.146	∞	3.643	3.796	3.900	3.978	4.040

*Reproduced with kind permission from H. Leon Harter and N. Balakrishnan, 1998. *Tables for the Use of Range and Studentized Range in Tests of Hypotheses,* CRC Press, New York, 558–561.

. Tables of Distributions and Critical Values

Critical values for the Spearman rank correlation coefficient r_s*

2-tail	0.10	0.05	0.02	0.01
1-tail	0.05	0.025	0.01	0.005
n: 4	1.000			
5	0.900	1.000	1.000	
6	0.829	0.886	0.943	1.000
7	0.714	0.786	0.893	0.929
8	0.643	0.738	0.833	0.881
9	0.600	0.700	0.783	0.833
10	0.564	0.648	0.745	0.794
11	0.536	0.618	0.709	0.755
12	0.503	0.587	0.678	0.727
13	0.484	0.560	0.648	0.703
14	0.464	0.538	0.626	0.679
15	0.446	0.521	0.604	0.654
16	0.429	0.503	0.582	0.635
17	0.414	0.485	0.566	0.615
18	0.401	0.472	0.550	0.600
19	0.391	0.460	0.535	0.584
20	0.380	0.447	0.520	0.570
21	0.370	0.435	0.508	0.556
22	0.361	0.425	0.496	0.544
23	0.353	0.415	0.486	0.532
24	0.344	0.406	0.476	0.521
25	0.337	0.398	0.466	0.511
26	0.331	0.390	0.457	0.501
27	0.324	0.382	0.448	0.491
28	0.317	0.375	0.440	0.483
29	0.312	0.368	0.433	0.475
30	0.306	0.362	0.425	0.467