UNIVERSITI SAINS MALAYSIA

Final Examination
2015/2016 Academic Session

May/June 2016

JIB 323 – Biostatistics
[Biostatistik]

Duration : 3 hours
[Masa : 3 jam]

Please ensure that this examination paper contains **TWELVE** printed pages before you begin the examination.

Ensure Appendix (formula and distribution tables) are enclosed with the question paper.

Answer **FIVE** questions. You may answer **either** in Bahasa Malaysia or English.

All answers must be written in the answer booklet provided.

Each question is worth 20 marks and the mark for each sub question is given at the end of that sub question.

In the event of any discrepancies in the exam questions, the English version shall be used.

*Sila pastikan bahawa kertas peperiksaan ini mengandungi **DUA BELAS** muka surat yang bercetak sebelum anda memulakan peperiksaan ini.*

Sila pastikan Appendiks (formula dan jadual taburan) disertakan bersama kertas soalan.

*Jawab **LIMA** soalan. Anda dibenarkan menjawab soalan **sama ada** dalam Bahasa Malaysia atau Bahasa Inggeris.*

Setiap jawapan mesti dijawab di dalam buku jawapan yang disediakan.

Setiap soalan bernilai 20 markah dan markah subsoalan diperlihatkan di penghujung subsoalan itu.

Sekiranya terdapat sebarang percanggahan pada soalan peperiksaan, versi Bahasa Inggeris hendaklah diguna pakai.

Answer FIVE questions
Jawab LIMA soalan

1. In a study of the relationship between heights and trunk diameters of trees, botany students collected sample data. Listed below are the tree circumference (in centimeters).

Table 1/*Jadual 1*

Tree circumference (centimeters)/*Lilitan pokok (sentimeter)*									
54.9	57.9	54.9	73.1	155.4	94.5	167.6	155.4	253.0	94.5

Using the circumference listed in the Table 1,

(a) Calculate the mean, median and mode

(b) Determine the range of data

(c) Calculate the standard deviation

(d) Calculate the variance

(e) Calculate the coefficient of variation

(f) Are the given values in centimeters of the population of trees are discrete or continuous data? Why?

(g) What is the level of measurement of these values (Nominal, ordinal, interval or ratio)? Why?

Sekumpulan pelajar botani telah mengambil sampel data untuk mengkaji hubungan antara tinggi dan diameter pokok-pokok. Jadual di bawah menyenaraikan lilitan pokok (dalam sentimeter).

Menggunakan lilitan yang disenaraikan dalam Jadual 1,

(a) Kira min, median dan mod

(b) Tentukan julat data tersebut

(c) Kira sisihan piawai

(d) Kira varians

(e) Kira pekali perubahan data tersebut

(f) *Adakan nilai yang diberi dalam sentimeter untuk populasi pokok merupakan data diskret atau berterusan? Mengapa?*

(g) *Apakah tahap pengukuran nilai-nilai tersebut (nominal, ordinal, selang atau nisbah)? Mengapa?*

(20 marks/*markah*)

2. A researcher in the United States has developed a theoretical model for predicting eye colour. After examining random sample of parents, she predicts the eye colour of the first child. Table 2 below list the eye colour of offsprings. Based on her theory, she predicted that 87% of the offsprings would inherit brown eyes, 8% would inherit blue eyes, and 5% would inherit green eyes.

Table 2/*Jadual 2*

	Brown eyes/ *mata coklat*	**Blue eyes/ *mata biru***	**Green eyes/ *mata hijau***
Frequency/ Frekuensi	132	17	0

Use a 0.05 significance level to test the claim that the actual frequencies corresponds to her predicted distribution.

Seorang penyelidik dari Amerika syarikat telah membangunkan satu model teori untuk meramalkan warna mata. Selepas pemeriksaan sampel rawak ibu bapa, beliau meramalkan warna mata anak pertama pasangan yang disampel. Jadual 2 menyenaraikan warna mata anak pertama pasangan. Berdasarkan teori, beliau meramalkan bahawa 87% daripada anak-anak akan mewarisi mata coklat, 8% akan mewarisi mata biru dan 5% akan mewarisi mata hijau.

Gunakan aras keertian 0.05 untuk menguji dakwaan bahawa frekuensi sebenar sepadan dengan ramalan beliau.

(20 marks/*markah*)

3. A biologist assumes that there is a linear relationship between the amount of fertilizer supplied to long bean plants and the subsequent yield of long beans obtained. Eight long bean plants, of the same variety, were selected at random and treated weekly, with a solution in which x gram of fertilizer was dissolved in a fixed quantity of water. The yield, y kilograms of long beans was recorded in Table 3.

Table 3: Effect of different amounts of fertilizer on the yield of long bean./

Jadual 3: Kesan jumlah baja yang berbeza terhadap hasil kacang panjang.

Plant/ *Tanaman*	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
x	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5
y	3.9	4.4	5.8	6.6	7.0	7.1	7.3	7.7

(a) Which is the independent variable (IV) and dependent variable (DV)?

(b) Determine the slope (b_1) of the regression line.

(c) Determine the y-axis (b_0) intercept of the regression line.

(d) Write the regression equation.

(e) What is the estimated yield of plant treated weekly with 3.2 gram of fertilizer?

Seorang ahli biologi mengandaikan terdapat pertalian linear antara jumlah baja yang digunakan dan hasil tanaman kacang panjang. Sebanyak lapan tumbuhan kacang panjang daripada varieti yang sama telah dipilih secara rawak dan dirawat setiap minggu dengan x gram baja yang dilarutkan dalam kuantiti air yang ditetapkan. Hasilnya, y kilogram kacang panjang telah direkodkan dalam Jadual 3.

(a) *Nyatakan yang mana satu pembolehubah tak bersandar (IV) dan bersandar (DV)?*

(b) Tentukan kecerunan (b_1) bagi garisan regresi tersebut.

…5/-

(c) *Tentukan kepintasan paksi-y (b_0) bagi garisan regresi tersebut.*

(d) *Tuliskan persamaan regresi.*

(e) *Apakah anggaran hasil tumbuhan yang dirawat baja sebanyak 3.2 gram setiap minggu?*

(20 marks/*markah*)

4. Write a detail essay on various sampling methods with an appropriate example for each method.

Tuliskan esei terperinci mengenai pelbagai kaedah persampelan dengan menggunakan satu contoh yang bersesuaian untuk setiap kaedah.

(20 marks/*markah*)

5. Table 4 shows data for systolic and diastolic blood pressure from Ministry of Health Malaysia of eight randomly selected males.

Table 4 : Systolic and diastolic / *Jadual 4: Tekanan darah sistolik dan diastolik*

Systolic/ Sistolik	125	107	126	110	110	107	113	126
Diastolic/ Diastolik	78	54	81	68	66	83	71	72

Using the data in Table 4, use 0.05 significance level.

(a) Calculate the correlation coefficient.

(b) Calculate the coefficient of determination.

(c) What is the strength of the correlation?

(d) Test H_o: $\rho = 0$, vs H_a: $\rho \neq 0$.

Jadual 4 menunjukkan data untuk tekanan darah sistolik dan diastolik dari Kementerian Kesihatan Malaysia yang diambil daripada lapan lelaki yang dipilih secara rawak.

Dengan menggunakan data di Jadual 4, gunakan aras keertian 0.05.

(a) *Kira pekali korelasi*

(b) *Kira pekali penentuan.*

(c) *Apakah kekuatan pertalian korelasi?*

(d) *Uji* H_o: $\rho = 0$, *terhadap* H_a: $\rho \neq 0$.

(20 marks/*markah*)

6. The following sepal length (mm) were obtained from 4 each of Setosa, Versicolor and Virginica plants. Use a 0.05 significance level to test the claim that the three different varieties come from populations having the same mean sepal length.

Table 5 : Sepal length (mm) / *Jadual 5 : Panjang sepal (mm)*

Setosa	Versicolor	Virginica
5.1	7.0	6.3
4.9	6.4	5.8
4.7	6.9	7.1
4.6	5.5	6.3

Panjang sepal (mm) berikut diperolehi daripada 4 pokok Setosa, Versicolor dan Virginica masing-masing. Gunakan aras keertian 0.05 untuk menguji dakwaan bahawa varieti tersebut berasal daripada populasi yang mempunyai min panjang sepal yang sama.

(20 marks/*markah*)

APPENDIX – FORMULA AND TABLES

Ch. 2: Descriptive Statistics

$\bar{x} = \frac{\Sigma x}{n}$ Mean

$\bar{x} = \frac{\Sigma f \cdot x}{\Sigma f}$ Mean (frequency table)

$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}$ Standard deviation

$s = \sqrt{\frac{n(\Sigma x^2) - (\Sigma x)^2}{n(n - 1)}}$ Standard deviation (shortcut)

$s = \sqrt{\frac{n[\Sigma(f \cdot x^2)] - [\Sigma(f \cdot x)]^2}{n(n - 1)}}$ Standard deviation (frequency table)

variance $= s^2$

$CV = \frac{s}{\bar{x}} \times 100$, coefficient of variation

Ch. 9: Linear Correlation/Regression

Correlation $r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$

$b_1 = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$

$b_0 = \bar{y} - b_1\bar{x}$ or $b_0 = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$

$\hat{y} = b_0 + b_1 x$ Estimated eq. of regression line

$r^2 = \frac{\text{explained variation}}{\text{total variation}}$

$s_e = \sqrt{\frac{\Sigma(y - \hat{y})^2}{n - 2}}$ or $\sqrt{\frac{\Sigma y^2 - b_0\Sigma y - b_1\Sigma xy}{n - 2}}$

$\hat{y} - E < y < \hat{y} + E$

where $E = t_{\alpha/2} s_e \sqrt{1 + \frac{1}{n} + \frac{n(x_0 - \bar{x})^2}{n(\Sigma x^2) - (\Sigma x)^2}}$

Ch. 10: Multinomial and Contingency Tables

$$\chi^2 = \Sigma\frac{(O - E)^2}{E} \quad \text{Multinomial (df} = k - 1)$$

$$\chi^2 = \Sigma\frac{(O - E)^2}{E} \quad \text{Contingency table } [\text{df} = (r - 1)(c - 1)]$$

$$\text{where } E = \frac{(\text{row total})(\text{column total})}{(\text{grand total})}$$

Ch. 11: One-Way Analysis of a Variance

$$F = \frac{ns_{\bar{x}}^2}{s_p^2} \quad k \text{ samples each of size } n \; (\text{num. df} = k - 1;\ \text{den. df} = k(n - 1))$$

$$F = \frac{\text{MS(treatment)}}{\text{MS(error)}} \quad \leftarrow \text{df} = k - 1$$

$$\text{MS(treatment)} = \frac{\text{SS(treatment)}}{k - 1}$$

$$\text{MS(error)} = \frac{\text{SS(error)}}{N - k} \qquad \text{MS(total)} = \frac{\text{SS(total)}}{N - 1}$$

$$\text{SS(treatment)} = n_1(\bar{x}_1 - \bar{\bar{x}})^2 + \cdots + n_k(\bar{x}_k - \bar{\bar{x}})^2$$

$$\text{SS(error)} = (n_1 - 1)s_1^2 + \cdots + (n_k - 1)s_k^2$$

$$\text{SS(total)} = \Sigma(x - \bar{\bar{x}})^2$$

$$\text{SS(total)} = \text{SS(treatment)} + \text{SS(error)}$$

Note: MS (treatment) also known as MS(between)
SS (treatment) also known as SS(between groups)/SS(between samples)/SS(between)
SS (error) also known as SS(within group)/SS(within samples)/SS(within)

Test statistic for one way ANOVA

F = <u>variance between samples</u>
variance within samples

TABLE A-3 *t* Distribution: Critical *t* Values

	Area in One Tail				
	0.005	0.01	0.025	0.05	0.10
	Area in Two Tails				
Degrees of Freedom	0.01	0.02	0.05	0.10	0.20
1	63.657	31.821	12.706	6.314	3.078
2	9.925	6.965	4.303	2.920	1.886
3	5.841	4.541	3.182	2.353	1.638
4	4.604	3.747	2.776	2.132	1.533
5	4.032	3.365	2.571	2.015	1.476
6	3.707	3.143	2.447	1.943	1.440
7	3.499	2.998	2.365	1.895	1.415
8	3.355	2.896	2.306	1.860	1.397
9	3.250	2.821	2.262	1.833	1.383
10	3.169	2.764	2.228	1.812	1.372
11	3.106	2.718	2.201	1.796	1.363
12	3.055	2.681	2.179	1.782	1.356
13	3.012	2.650	2.160	1.771	1.350
14	2.977	2.624	2.145	1.761	1.345
15	2.947	2.602	2.131	1.753	1.341
16	2.921	2.583	2.120	1.746	1.337
17	2.898	2.567	2.110	1.740	1.333
18	2.878	2.552	2.101	1.734	1.330
19	2.861	2.539	2.093	1.729	1.328
20	2.845	2.528	2.086	1.725	1.325
21	2.831	2.518	2.080	1.721	1.323
22	2.819	2.508	2.074	1.717	1.321
23	2.807	2.500	2.069	1.714	1.319
24	2.797	2.492	2.064	1.711	1.318
25	2.787	2.485	2.060	1.708	1.316
26	2.779	2.479	2.056	1.706	1.315
27	2.771	2.473	2.052	1.703	1.314
28	2.763	2.467	2.048	1.701	1.313
29	2.756	2.462	2.045	1.699	1.311
30	2.750	2.457	2.042	1.697	1.310
31	2.744	2.453	2.040	1.696	1.309
32	2.738	2.449	2.037	1.694	1.309
34	2.728	2.441	2.032	1.691	1.307
36	2.719	2.434	2.028	1.688	1.306
38	2.712	2.429	2.024	1.686	1.304
40	2.704	2.423	2.021	1.684	1.303
45	2.690	2.412	2.014	1.679	1.301
50	2.678	2.403	2.009	1.676	1.299

Formulas and Tables

for *Elementary Statistics, Ninth Edition*, by Mario F. Triola

TABLE A-4 Chi-Square (χ^2) Distribution

Degrees of Freedom	Area to the Right of the Critical Value									
	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672

F Values for $\alpha = 0.05$

d_2 \ d_1	1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.3	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27

TABLE A-6 Critical Values of the Pearson Correlation Coefficient r

n	$\alpha = .05$	$\alpha = .01$
4	.950	.999
5	.878	.959
6	.811	.917
7	.754	.875
8	.707	.834
9	.666	.798
10	.632	.765
11	.602	.735
12	.576	.708
13	.553	.684
14	.532	.661
15	.514	.641
16	.497	.623
17	.482	.606
18	.468	.590
19	.456	.575
20	.444	.561
25	.396	.505
30	.361	.463
35	.335	.430
40	.312	.402
45	.294	.378
50	.279	.361
60	.254	.330
70	.236	.305
80	.220	.286
90	.207	.269
100	.196	.256

- oooOooo -