First Semester Examination
2017/2018 Academic Session

January 2018

MAA161 - Statistics for Science Students
[Statistik untuk Pelajar Sains]

Duration : 3 hours
(Masa : 3 jam)

Please check that this examination paper consists of **THIRTEEN (13)** pages of printed material before you begin the examination.

*[Sila pastikan bahawa kertas peperiksaan ini mengandungi **TIGA BELAS (13)** muka surat yang bercetak sebelum anda memulakan peperiksaan ini].*

Instructions : Answer **all seven (7)** questions.

*[**Arahan :** Jawab **semua tujuh (7)** soalan.]*

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

[Sekiranya terdapat sebarang percanggahan pada soalan peperiksaan, versi Bahasa Inggeris hendaklah digunapakai].

Question 1

You are given 12 observations (x) as follows:

3	2	47	118	204	97
68	86	62	57	98	99

(a) $\sum x = 941;\ \sum x^2 = 105689$. Find the standard deviation of these data.

(b) Do you think that the sample standard deviation is a good measurement of variation? Why?

(c) Using Chebyshev's theorem, find at least what percentage of the observations fall in the intervals $\bar{x} \pm 2.5s$

(d) Draw a dotplot using the sample data. Comment on the distribution of the data.

[12 marks]

Soalan 1

Anda diberi 12 cerapan (x) seperti berikut:

3	2	47	118	204	97
68	86	62	57	98	99

(a) $\sum x = 941;\ \sum x^2 = 105689$. *Cari sisihan piawai untuk data tersebut.*

(b) Adakah anda berpendapat sisihan piawai sampel merupakan pengukur yang baik terhadap variasi? Kenapa?

(c) Dengan menggunakan teorem Chebyshev, cari sekurang-kurangnya berapakah peratusan dari cerapan yang berada dalam selang $\bar{x} \pm 2.5s$ *?*

(d) Lukis satu dotplot menggunakan data sampel. Berikan komen terhadap taburan data.

[12 markah]

Question 2

Bob drives to his university. The probability he will face road accident under different weathers are summarized as below:

	Good weather, G	Rain, R	Snow, S	Total
Accident, A	0.1	0.1	0.2	0.4
No accident, N	0.4	0.1	0.1	0.6
Total	0.5	0.2	0.3	1.0

(a) What is the probability that Bob will face an accident (A)?

(b) What is the probability of snow (S)?

(c) What is the probability of an accident given snow?

(d) Are the events accident (A) and snow (S) independent?

[12 marks]

Soalan 2

Bob memandu ke universitinya. Kebarangkalian dia akan menghadapi kemalangan jalanraya pada cuaca berlainan adalah dirumuskan seperti berikut:

	Cuaca baik, G	*Hujan, R*	*Salji, S*	*Jumlah*
Kemalangan, A	*0.1*	*0.1*	*0.2*	*0.4*
Tiada kemalangan, N	*0.4*	*0.1*	*0.1*	*0.6*
Jumlah	*0.5*	*0.2*	*0.3*	*1.0*

(a) Apakah kebarangkalian bahawa Bob akan menghadapi kemalangan (A)?

(b) Apakah kebarangkalian turun salji (S)?

(c) Apakah kebarangkalian berlaku kemalangan jika hari bersalji?

(d) Adakah peristiwa (A) dan (S) bersandar?

[12 markah]

Question 3

The distribution of scores for persons over 16 years of age on a common IQ test is approximately normal with mean 100 and standard deviation 15.

(a) What is the probability that a randomly chosen adult has an IQ score over 105?

(b) What are the mean and standard deviation of the average IQ score on this test for 60 people?

(c) What is the probability that the average IQ on this test of 60 people is 105 or higher?

(d) Would your answer in (a) and (c) be affected if the distribution of IQ scores on this test in the adult population were distinctly non-normal? Explain your answer.

[12 marks]

Soalan 3

Taburan skor untuk mereka yang berusia 16 tahun ke atas dalam ujian IQ biasa adalah berhampiran normal dengan min 100 dan sisihan piawai 15.

(a) Apakah kebarangkalian bahawa seorang dewasa yang dipilih secara rawak mempunyai skor IQ melebihi 105?

(b) Apakah min dan sisihan piawai bagi purata skor IQ ujian ini bagi 60 orang?

(c) Apakah kebarangkalian bahawa purata IQ untuk 60 orang ialah 105 atau lebih tinggi?

(d) Adakah jawapan anda dalam (a) dan (c) terpengaruh jika taburan skor IQ ujian ini untuk populasi dewasa sangat tidak normal? Jelaskan jawapan anda.

[12 markah]

Question 4

(a) Write down two conditions for $X \sim Bin(n, p)$ to be approximated by a normal distribution $Y \sim N(\mu, \sigma^2)$.

(b) A factory manufactures 2000 DVDs everyday. It is known that 3% of DVDs are faulty.

(i) Using a binomial approximation, find the probability that at least 40 faulty DVDs are produced in one day.

(ii) The quality control system in the factory identifies and destroys every faulty DVD at the end of the manufacturing process. It costs \$0.70 to manufacture a DVD and the factory sells non-faulty DVDs for \$11. Find the expected profit made by the factory per day.

[12 marks]

Soalan 4

(a) Tuliskan dua keadaan untuk $X \sim Bin(n, p)$ *untuk diberi penghampiran oleh taburan normal* $Y \sim N(\mu, \sigma^2)$.

(b) Sebuah kilang menghasilkan 2000 DVD setiap hari. Adalah diketahui bahawa 3% DVD mempunyai masalah kerosakan.

(i) Dengan menggunakan penghampiran binomial, cari kebarangkalian bahawa sekurang-kurangnya 40 DVD yang dihasilkan dalam satu hari menghadapi kerosakan.

(ii) Sistem kawalan kualiti dalam kilang mengenalpasti dan memusnahkan setiap DVD yang rosak pada akhir proses pengeluaran. Ia berkos \$0.70 untuk mengeluarkan satu DVD dan kilang menjual DVD yang tiada kerosakan dengan \$11. Cari jangkaan keuntungan yang dibuat oleh kilang untuk satu hari.

[12 markah]

Question 5

A multiple-choice test question has four possible responses. The question is designed to be very difficult, yet with only one correct answer. It first occurs on an exam taken by 400 students. The designers test whether more people answer the question correctly than would be expected just due to coincidence.

(a) Construct the hypotheses for the test.

(b) Perform the hypothesis test using α =0.05. What is your conclusion?

(c) What is the estimate of the sample size if we limit the margin of error to be within 0.02 based on the population proportion for the 95% confidence interval?

[12 marks]

Soalan 5

Satu soalan ujian jenis pelbagai pilihan mempunyai empat kemungkinan jawapan. Soalan tersebut direka supaya sangat susah dengan hanya satu jawapan yang betul. Ujian ini pertama kalinya diambil oleh 400 pelajar. Pereka soalan ingin menguji sama ada lebih ramai pelajar menjawab dengan tepat berbanding seperti yang dijangkakan disebabkan secara kebetulan.

(a) Bina hipotesis untuk ujian ini.

(b) Jalankan ujian hipotesis menggunakan α =0.05. Apakah kesimpulan anda?

(c) Apakah anggaran saiz sampel jika kita hadkan margin ralat kepada dalam linkungan 0.02 berasaskan kadar populasi untuk 95% selang keyakinan?

[12 markah]

Question 6

(a) Two students investigate whether people are more likely to be helpful in a busy street or quiet street. One of them drops a bag of apples at randomly selected times during the day and the other watches to see whether anyone helps by picking them up and returning them to their owner. The results are as follows:

	Busy street	Quiet street
Helped	14	20
Not helped	10	6

Use χ^2 test at the 5% significance level to see whether there appears to be any different between the response of people in a busy street to in a quiet street.

(b) The random variable X has probability distribution

x	1	3	5	7	9
$P(X=x)$	0.2	P	0.2	Q	0.15

(i) Given that $E(X) = 4.5$, write down two equations involving P and Q.

(ii) Find the value of P and Q.

(iii) Find $P(4 < X \leq 7)$

[20 marks]

Soalan 6

(a) *Dua pelajar menyiasat sama ada orang ramai adalah lebih cenderung untuk membantu dalam kesibukan jalan atau kesunyian jalan. Salah seorang pelajar mencicirkan sebuah beg yang mengandungi epal secara rawak pada masa yang dipilih dalam satu hari dan seorang lagi melihat sama ada sesiapa membantu dengan mengambil epal dan memulangkan kepada pemiliknya. Keputusan adalah seperti berikut:*

	Kesibukan jalan	*Kesunyian jalan*
Membantu	*14*	*20*
Tidak membantu	*10*	*6*

Gunakan ujian χ^2 *pada aras keertian 5% untuk menguji sama ada terdapat perbezaan antara reaksi orang ramai dalam kesibukan jalan berbanding dalam kesunyian jalan.*

(b) *Pembolehubah rawak X mempunyai taburan kebarangkalian*

x	1	3	5	7	9
$P(X=x)$	0.2	P	0.2	Q	0.15

(i) *Diberi bahawa E(X)=4.5, tuliskan dua persamaan melibatkan P dan Q.*

(ii) *Cari nilai P dan Q.*

(iii) *Cari* $P(4 < X \leq 7)$

[20 markah]

Question 7

Ten children one each selected from 10 sets of identical twins, were trained by a certain method A and the remaining 10 were trained by method B. At the end of the year, the following IQ scores were obtained:

Pair	1	2	3	4	5	6	7	8	9	10
A	31	25	38	33	42	40	44	26	43	35
B	44	30	34	47	35	32	35	47	48	34

(a) What is nonparametric test? Give two advantages of nonparametric test.

(b) Perform a Mann Whitney rank sum test on the following paired data. Is there sufficient evidence to indicate a difference in the average IQ of the groups given α =0.10?

(c) Perform a paired *t*-test on the same dataset given $\alpha = 0.10$. Do you obtain the same results as in (b)? Which test is preferred? Why?

[20 marks]

Soalan 7

Sepuluh orang kanak-kanak dengan setiap satu dipilih dari 10 pasangan kembar yang sama, di mana mereka dilatih dengan kaedah tertentu A dan selebihnya 10 dilatih dengan kaedah B. Pada akhir tahun, skor IQ berikut diperoleh:

Pasangan	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
A	*31*	*25*	*38*	*33*	*42*	*40*	*44*	*26*	*43*	*35*
B	*44*	*30*	*34*	*47*	*35*	*32*	*35*	*47*	*48*	*34*

(a) Apakah yang dimaksudkan ujian tak berparametrik? Berikan dua kebaikan ujian tak berparametrik.

(b) Jalankan ujian jumlah pangkat Mann Whitney terhadap pasangan data tersebut. Adakah terdapat bukti yang mencukupi untuk menentukan perbezaan dalam purata IQ kumpulan tersebut diberi α *=0.10?*

(c) Jalankan ujian pasangan t terhadap data yang sama diberi $\alpha = 0.10$ *. Adakah anda mendapat keputusan yang sama dalam (b)? Ujian manakah yang akan diutamakan? Kenapa?*

[20 markah]

APPENDIX

Standard Normal Probabilities

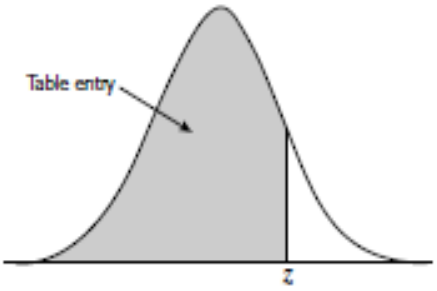
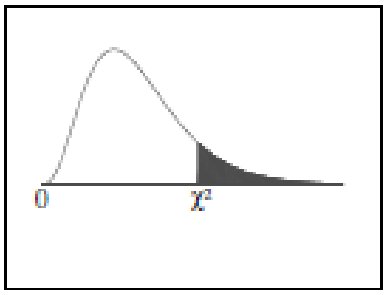


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Chi-Square Distribution Table



The shaded area is equal to α for $\chi^2 = \chi^2_\alpha$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.960}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	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	11.070	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	28.300
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	14.041	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	43.195	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	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	96.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

FORMULA

Descriptive measures:

$$\bar{x} = \frac{\Sigma x_i}{n} \qquad \mu = \frac{\Sigma x_i}{N}$$

$$s = \sqrt{\frac{\Sigma x_i^2 - \left(\Sigma x_i\right)^2 / n}{n-1}} \qquad \sigma = \sqrt{\frac{\Sigma x_i^2}{N} - \mu^2}$$

Probability:

$$P\left(A \cup B\right) = P\left(A\right) + P\left(B\right) - P\left(A \cap B\right)$$

$$P_r = \frac{n!}{(n-r)!} \qquad P\left(A|B\right) = \frac{P\left(A \cap B\right)}{P\left(B\right)}$$

$${}_nC_r = \frac{n!}{\left(n-r\right)!r!}$$

Independence: $P\left(A \cap B\right) = P\left(A\right) \cdot P\left(B\right)$

Random variables:

Discrete:

$$P(X = x) = \binom{n}{x} p^x q^{n-x}$$

$$P\left(X = x\right) = \frac{{}^kC_x \, {}^{(N-k)}C_{n-x}}{{}^NC_n}$$

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad x = 0, 1, 2, \ldots$$

$$\mu = \sum xP(x)$$

$$E(x) = \sum xP(x)$$

Continuous:

$$\mu = E(X) = \int_{-\infty}^{\infty} x \, f(x) \, dx$$

$$Var(X) = \sigma_X^2 = E(X^2) - \mu^2$$
$$= \int_{-\infty}^{\infty} x^2 \, f(x) \, dx - \mu^2$$

$$z = \frac{x - \mu}{\sigma} \qquad \mu = np \text{ and } \sigma = \sqrt{npq}$$

Nonparametric test:

Mann Whitney test:

$$U_1 = n_1 n_2 + \frac{n_2\left(n_1 + 1\right)}{2} - R_2$$

$$U_2 = n_1 n_2 + \frac{n_1\left(n_2 + 1\right)}{2} - R_1$$

$$z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2\left(n_1 + n_2 + 1\right)}{12}}}$$

Chi-Square tests

$$\chi^2 = \sum\left[\frac{(O-E)^2}{E}\right]$$

Goodness of fit:

$E = np$

$df = k - 1$

Independence/ homogeneity:

$df = (r-1)(c-1)$

$$E = \frac{R \cdot C}{n}$$

Test Statistics:

1. $Z = \dfrac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$ or $Z = \dfrac{\bar{X} - \mu}{\dfrac{s}{\sqrt{n}}}$

2. $T = \dfrac{\bar{X} - \mu}{\dfrac{s}{\sqrt{n}}}$

3. $T = \dfrac{\bar{d} - \mu_d}{\dfrac{s_d}{\sqrt{n_d}}}$

4. $Z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$

5. $Z = \dfrac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{\sigma_x^2}{n_x} + \dfrac{\sigma_y^2}{n_y}}}$

6. $T = \dfrac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{S_p^2\left(\dfrac{1}{n_x} + \dfrac{1}{n_y}\right)}}$

7. $T = \dfrac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}}}$

with $df = \dfrac{\left(\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}\right)^2}{\dfrac{\left(\dfrac{s_x^2}{n_x}\right)^2}{n_x - 1} + \dfrac{\left(\dfrac{s_y^2}{n_y}\right)^2}{n_y - 1}}$

8. $Z = \dfrac{(\hat{p}_x - \hat{p}_y) - (p_x - p_y)}{\sqrt{\dfrac{p_x(1-p_x)}{n_x} + \dfrac{p_y(1-p_y)}{n_y}}}$

9. $Z = \dfrac{(\hat{p}_x - \hat{p}_y) - (p_x - p_y)}{\sqrt{\bar{p}(1-\bar{p})\left(\dfrac{1}{n_x} + \dfrac{1}{n_y}\right)}}$

10. $\chi^2 = \dfrac{(n-1)s^2}{\sigma^2}$

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

12. $\chi^2 = \sum \dfrac{(O-E)^2}{E}, \quad E = np$

- ooo00ooo -