
UNIVERSITI SAINS MALAYSIA

First Semester Examination
2011/2012 Academic Session

January 2012

MAA 161 –Statistics for Science Students
[Statistik untuk Pelajar Sains]

Duration : 3 hours
[Masa : 3 jam]

Please check that this examination paper consists of TWELVE pages of printed material before you begin the examination.

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

Instructions: Answer **all eight** [8] questions.

Arahan: Jawab **semua lapan** [8] soalan.

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

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

1. Over a period of three years, a bank keeps a weekly record of the number of cheques with errors that are presented for payment. The results for 100 accounting weeks are shown as follows

Number of cheques with errors (x)	Number of weeks (f)
0	3
1	11
2	25
3	21
4	13
5	14
6	8
7	2
8	2
9	1

- (a) Construct a suitable frequency table consisting of additional columns – boundaries, $\sum xf$ and $\sum x^2 f$.
- (b) Form a histogram of the above table.
- (c) Also, construct a boxplot of these data.
- (d) Some textbooks measure the skewness of a distribution by $\frac{3 \text{ mean-median}}{\text{standard deviation}}$, and others measure it by $\frac{\text{mean-mode}}{\text{standard deviation}}$. Calculate and compare the values of these two measures of skewness for the above data.

[10 marks]

1. *Selama tempoh tiga tahun, sebuah bank telah menyimpan rekod mingguan bilangan cek dengan kesalahan yang digunakan untuk pembayaran. Keputusan untuk 100 minggu perakaunan ditunjukkan seperti berikut*

<i>Bilangan cek yang mengandungi kesalahan (x)</i>	<i>Bilangan minggu (f)</i>
0	3
1	11
2	25
3	21
4	13
5	14
6	8
7	2
8	2
9	1

- (a) Bina jadual kekerapan yang sesuai mengandungi lajur-lajur tambahan, iaitu selang kelas, $\sum xf$ dan $\sum x^2 f$.
- (b) Bina histogram untuk jadual di atas.
- (c) Juga, bina boxplot untuk data-data tersebut.
- (d) Seseengah buku teks mengukur kepencongkan sesuatu taburan sebagai $\frac{3 \text{ min-median}}{\text{sisihan piawai}}$, dan selainnya sebagai $\frac{\text{min-mod}}{\text{sisihan piawai}}$. Kira dan bandingkan nilai-nilai daripada dua ukuran kepencongkan data di atas.

[10 markah]

2. In a class consisting of 10 boys and 20 girls, two students are selected at random (without replacement). What is the probability of selecting

- (a) Two boys, i.e. a boy and a boy?
- (b) Two girls?
- (c) Either two boys or two girls?

Now, if three students are randomly selected without replacement, what is the probability of selecting

- (d) Three boys?
- (e) Three girls?
- (f) Either three boys or three girls?
- (g) Neither three boys nor three girls?
- (h) Conditional probability of three boys given that at least one boy has been chosen at random.

[15 marks]

2. Dalam sebuah kelas yang mengandungi 10 pelajar lelaki dan 20 pelajar perempuan, dua pelajar telah dipilih secara rawak (tanpa dikembalikan). Apakah kebarangkalian untuk memilih

- (a) Dua lelaki, iaitu lelaki dan lelaki?
- (b) Dua perempuan?
- (c) Sama ada dua lelaki atau dua perempuan?

Sekarang, jika tiga pelajar dipilih secara rawak (tanpa pengembalian), apakah kebarangkalian

- (d) Tiga pelajar lelaki?
- (e) Tiga pelajar perempuan?
- (f) Sama ada tiga lelaki atau tiga perempuan?
- (g) Bukan tiga lelaki mahupun tiga perempuan?
- (h) Kebarangkalian bersyarat tiga lelaki diberi bahawa sekurang-kurangnya seorang lelaki telah dipilih secara rawak.

[15 markah]

3. Proton Holding Bhd is conducting a study to evaluate the performance of her two types of vehicles in consuming the petrol over a long journey. The data quoted are in unit of litre consumption per 100 km (*litre per 100 km*). The saver of the vehicle in fuel consumption is indicated by the less value of fuel consumption. Ten observations have been quoted respectively from difference types of specification of the cars – namely 1.5L and 2.0L. The results are presented as follows:-

1.5L	8.208	9.103	10.87	7.884	9.035	10.253	8.187	8.523	8.595	8.575
<i>x</i>										
2.0L	11.3	9.013	9.999	10.799	9.221	11.227	8.679	10.025	10.548	9.532
<i>y</i>										

Also given, $\sum x = 89.233$, $\sum y = 100.343$, $\sum x^2 = 804.3786$ and $\sum y^2 = 1014.54$

- At the significance level of 5%, test whether the variance of the first type (1.5L) is greater than the second one (2.0L).
- With the assumption of common variance, that is 0.9, and at the significance level of 5%, test whether the average fuel consumption for 1.5L vehicles are saver in fuel consumption.
- Assumed with common unknown variance, and using the significance level of 5%, test whether the average fuel consumption for 1.5L vehicles are saver in fuel consumption.

[15 marks]

3. *Proton Holding Bhd sedang menjalankan satu kajian untuk menilai prestasi dua jenis kereta keluarannya terhadap penggunaan petrol bagi perjalanan jauh. Data-data yang dicerap adalah dalam unit penggunaan liter per 100 km (iaitu, litre per 100 km). Penggunaan petrol yang jimat oleh kenderaan ditunjukkan oleh nilai penggunaan petrol yang kurang. Sepuluh cerapan telah dikutip masing-masing daripada dua jenis kereta tersebut – iaitu 1.5L dan 2.0L. Keputusan dibentangkan seperti berikut:-*

1.5L	8.208	9.103	10.87	7.884	9.035	10.253	8.187	8.523	8.595	8.575
<i>x</i>										
2.0L	11.3	9.013	9.999	10.799	9.221	11.227	8.679	10.025	10.548	9.532
<i>y</i>										

Juga diberi, $\sum x = 89.233$, $\sum y = 100.343$, $\sum x^2 = 804.3786$ dan $\sum y^2 = 1014.54$

- Pada aras keyakinan 5%, uji sama ada varians bagi kereta jenis pertama (1.5L) adalah lebih besar daripada yang kedua (2.0L).*
- Dengan andaian varians sepunya, iaitu 0.9, dan aras keyakinan 5%, uji sama ada purata penggunaan petrol bagi kereta 1.5L adalah lebih jimat dari segi penggunaan petrol.*
- Andaikan varians sepunya yang tidak diketahui, dan dengan menggunakan aras keyakinan 5%, uji sama ada purata penggunaan petrol bagi kenderaan 1.5L adalah lebih jimat dalam penggunaan petrol.*

[15 markah]

...5/-

4. In a reality program show, it has been studied that Mawi's winning votes are according to Poisson distribution with 5 votes per minute. Meanwhile, Siti's winning votes are following the Poisson distribution with 4.5 votes per minute.
- (a) Calculate the probability that Mawi wins 5 votes or more within three minutes in time.
 - (b) Within 1 hour program, calculate the probability that Siti wins the votes more 300 votes.[Hint: use normal approximation to solve the question].
 - (c) After two hours reality program run, using the normal approximation and assuming that Mawi's votes and Siti's votes are mutually independent, find the probability that total Siti's votes is higher than Mawi's.

[10 marks]

4. Dalam suatu program realiti, telah dikaji bahawa Mawi memenangi undian berdasarkan taburan Poisson dengan 5 undian per minit. Manakala Siti meraih undian berdasarkan taburan Poisson dengan 4.5 undian per minit.

- (a) Kira kebarangkalian bahawa Mawi meraih 5 atau lebih undian dalam tempoh tiga minit.
- (b) Dalam tempoh 1 jam, kira kebarangkalian Siti meraih undian lebih daripada 300 undian.[Petunjuk: gunakan penghampiran normal untuk menyelesaikan soalan ini].
- (c) Selepas dua jam program tersebut berlangsung, dengan menggunakan penghampiran normal serta andaian bahawa undian Mawi dan Siti adalah tidak bersandar, cari kebarangkalian keseluruhan undian Siti adalah lebih tinggi daripada keseluruhan undian Mawi.

[10 markah]

5. A new driver is facing a risk of mild accident on the road, but the events of the accidents do not cause him to death. It is also assumed that the mild accident causing in claiming of insurance, happens once per year.

- (a) In three-year periods, state the set of possible accidents happening to him.
- (b) Let X be a random variable of the number of mild accident within five years period, what is the mean and the variance if the probability of the accident is homogenously 0.15?
- (c) What is the probability that he will claim the insurance more than twice due to the accident cause for the next five years?

[10 marks]

5. Seorang pemandu baru mengalami risiko kemalangan kecil di jalan raya, tetapi kemalangan tersebut tidak mengakibatkan kematiannya. Juga diandaikan bahawa kemalangan kecil ini yang mengakibatkan tuntutan insurans, berlaku sekali sahaja setiap tahun.

- (a) Dalam tempoh tiga tahun, nyatakan set-set kemalangan yang mungkin berlaku terhadapnya.
- (b) Biar X sebagai pembolehubah rawak bilangan kemalangan kecil dalam tempoh lima tahun, apakah min (jangkaan) dan varians jika kebarangkalian kemalangan tersebut adalah secara homogennya 0.15?
- (c) Apakah kebarangkalian bahawa dia akan membuat tuntutan insurans lebih daripada dua kali disebabkan kemalangan untuk lima tahun akan datang ini?

[10 markah]

...6/-

6. A test for a disease correctly diagnoses a diseased person is having the disease with probability 0.85. The test incorrectly diagnoses someone without the disease as having the disease with a probability of 0.10. It is studied in a certain population, 5% of these have the disease. For a certain period of years, 500 patients are being tested on that disease test. The results are shown as follows.

Tests results	Disease and correctly diagnosed	Disease and wrongly diagnosed	Non disease and correctly diagnosed	Non disease and wrongly diagnosed
N(test)	30	3	426	41

Total = 500 patients

- Find probabilities for all outcomes (i.e. Disease and correctly diagnosed, Disease and wrongly diagnosed, Non disease and correctly diagnosed, and Non disease and wrongly diagnosed).
- For a sample containing 100 patients, what is the number of people who are expectedly, not having that disease, but are diagnosed as positively diseased? (Hint: the positive test indicates that the patient is tested to have the particular disease.)
- Perform the chi-squared test to verify the correctness of the theory by considering 500 patients and use a significance level of 0.01.

[15 marks]

6. *Suatu ujian terhadap kaedah ujikaji yang teruji benar ke atas seseorang yang mempunyai suatu penyakit, mempunyai kebarangkalian 0.85. Kajian terhadap ujikaji yang teruji salah bagi orang yang tidak mempunyai penyakit tersebut pula ialah 0.10. Telah dikaji bahawa dalam sesuatu populasi, 5% daripadanya mempunyai penyakit tersebut. Bagi suatu tempoh beberapa tahun, 500 pesakit telah diuji dengan ujian penyakit tersebut. Keputusan ditunjukkan seperti berikut*

<i>Keputusan Ujian</i>	<i>Berpenyakit dan teruji benar</i>	<i>Berpenyakit tetapi teruji salah</i>	<i>Tidak berpenyakit dan teruji benar</i>	<i>Tidak berpenyakit tetapi teruji salah</i>
<i>N(ujian)</i>	<i>30</i>	<i>3</i>	<i>426</i>	<i>41</i>

Jumlah keseluruhan= 500 pesakit

- Cari kebarangkalian untuk kesemua keputusan (iaitu Berpenyakit dan teruji benar, Berpenyakit tetapi teruji salah, Tidak berpenyakit dan teruji benar, and Tidak berpenyakit tetapi teruji salah).*
- Untuk suatu sampel 100 pesakit, apakah bilangan pesakit yang dijangka, tidak berpenyakit sedangkan dia diuji secara positif berpenyakit? (Petunjuk: uji positif menandakan bahawa seseorang itu didapati mempunyai penyakit tersebut)*
- Laksanakan ujian khi kuasa dua untuk menentusahkan teori di atas dengan mempertimbangkan 500 pesakit dan gunakan aras keertian 0.01.*

[15 markah]

7. An organizer is reconsidering in organizing a bowling charity tournament for the next year. They are deciding whether to separate the group of tournament based on the following groups – primary school, secondary school, female adult and male adult – or mix them up as an open tournament. Based on the previous tournament, they have gathered the number of pins dropped in one game as shown in the table below:-

	Primary school	Secondary school	Female adult	Male Adult
Number of dropped pins below than 50	37	52	33	26
Number of dropped pins 50 and above	13	23	27	54

The tournament will be conducted separately based on the stated groups if they found out that the groups and the pins dropped are dependent.

- State the necessary hypothesis
- What is the suitable test of this study?
- Perform the test and conclude the hypothesis above. Should the organizer run the tournament separately by groups? Use a significance level of 0.01.

[10 marks]

7. *Suatu organisasi sedang mempertimbangkan semula untuk melaksanakan pertandingan boling amal pada tahun hadapan. Mereka sedang mempertimbangkan sama ada membahagikan kumpulan-kumpulan bagi pertandingan ini berdasarkan kumpulan-kumpulan berikut – sekolah rendah, sekolah menengah, perempuan dewasa dan lelaki dewasa – atau gabungkan kesemua kumpulan sebagai pertandingan terbuka. Berdasarkan pertandingan sebelumnya, mereka telah memperolehi bilangan pin yang dijatuhkan dalam satu permainan, seperti ditunjukkan di jadual di bawah:-*

	<i>Sekolah rendah</i>	<i>Sekolah menengah</i>	<i>Perempuan dewasa</i>	<i>Lelaki dewasa</i>
<i>Bilangan pin jatuh di bawah 50</i>	37	52	33	26
<i>Bilangan pin jatuh lebih daripada 50</i>	13	23	27	54

Pertandingan ini akan dijalankan secara kumpulan yang berasingan jika didapati kumpulan dan jatuhan pin adalah bersandar.

- Nyatakan hipotesis yang sesuai*
- Apakah ujian yang sesuai dengan kajian ini?*
- Laksanakan ujian tersebut dan simpulkan hipotesis di atas. Adakah patut organisasi itu menjalankan pertandingan tersebut secara berasingan mengikut kumpulan? Gunakan aras keyakinan 0.01.*

[10 markah]

8. A team of economists were studying the impact of the economic downturn for a specific period of years. They initially anticipated the variables of the return on equity (*ROE*) for the randomly selected public listed companies' annual reports. 10 companies were selected and their *ROE* for the respective years (before crisis and after crisis) were shown as follows:-

Companies	Return on Equities (ROE)	
	Before crisis	After crisis
C1	0.596	0.766
C2	-0.302	-0.497
C3	0.479	0.166
C4	0.255	0.479
C5	-0.125	0.001
C6	0.529	-0.136
C7	0.922	-0.198
C8	-0.473	0.509
C9	0.68	0.025
C10	0.971	0.845

- What parametric test should be used to test the impact of the crisis towards the companies' *ROE*, assuming that the *ROE* are normally distributed? Why do you choose this test?
- Conduct the parametric test mentioned in (a). Use a significance level of 0.01.
- What nonparametric test can be used if you are unsure about the distribution of the samples? Why do you choose this test?
- Conduct the nonparametric test mentioned in (c). Use a significance level of 0.01.

[15 marks]

8. *Sekumpulan pakar ekonomi telah mengkaji kesan ekonomi merudum untuk suatu tempoh tahun tertentu. Pada permulaannya, mereka menerima pakai pembolehubah pulangan daripada ekuiti (ROE) untuk laporan kewangan tahunan syarikat tersenarai awam yang dipilih secara rawak. Sepuluh syarikat telah dipilih dan ROE mereka bagi tahun masing-masing (sebelum dan selepas krisis) telah ditunjukkan seperti berikut:-*

Syarikat	Pulangan daripada Ekuiti (ROE)	
	Sebelum krisis	Selepas krisis
C1	0.596	0.766
C2	-0.302	-0.497
C3	0.479	0.166
C4	0.255	0.479
C5	-0.125	0.001
C6	0.529	-0.136
C7	0.922	-0.198
C8	-0.473	0.509
C9	0.68	0.025
C10	0.971	0.845

- (a) *Ujian berparameter apakah yang patut digunakan untuk menguji kesan krisis ekonomi tersebut terhadap ROE syarikat ini, dengan andaian bahawa ROE adalah tertabur secara normal? Kenapa anda memilih ujian ini?*
- (b) *Jalankan ujian berparameter dalam bahagian (a). Gunakan aras keertian 0.01.*
- (c) *Apakah ujian tidak berparameter yang boleh digunakan jika anda tidak pasti akan taburan sampel tersebut? Kenapa anda memilih ujian ini?*
- (d) *Jalankan ujian tidak berparameter seperti dinyatakan pada bahagian (c). Gunakan aras keertian 0.01.*

[15 markah]

FORMULAE for MAA 161/4

$\bar{x} = \frac{\sum xf}{\sum f} = \frac{\sum x}{n}$ $s^2 = \frac{\sum(x^2 f) - \frac{\sum xf^2}{\sum f}}{\sum f - 1}$ $= \frac{\sum x^2 - \frac{\sum x^2}{n}}{n - 1}$	$S_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$ $\bar{p} = \frac{X + Y}{n_x + n_y}$
<p>For $g \leq p < n+1 \leq g+1$, where $0 \leq p \leq 1$</p> $\frac{\hat{\pi}_p - x_g}{x_{g+1} - x_g} = p < n+1 - g$	
<p>Confidence Intervals:</p> <ol style="list-style-type: none"> $\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ or $\bar{X} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$ $\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$ $\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $(\bar{X} - \bar{Y}) \pm z_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$ 	<ol style="list-style-type: none"> $(\bar{X} - \bar{Y}) \pm t_{\alpha/2} \sqrt{S_p^2 \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}$ $(\hat{p}_x - \hat{p}_y) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}$ $\frac{(n-1)s^2}{\chi_{\alpha/2}^2}$ to $\frac{(n-1)s^2}{\chi_{1-\alpha/2}^2}$

<p>Test Statistics:</p> <p>1. $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ or $Z = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$</p> <p>2. $T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$</p> <p>3. $T = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n_d}}}$</p> <p>4. $Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$</p>	<p>5. $Z = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$</p> <p>6. $T = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{S_p^2 \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}}$</p> <p>7. $T = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$ with $df = \frac{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y} \right)^2}{\frac{\left(\frac{s_x^2}{n_x} \right)^2}{n_x - 1} + \frac{\left(\frac{s_y^2}{n_y} \right)^2}{n_y - 1}}$</p>
<p>8. $Z = \frac{(\hat{p}_x - \hat{p}_y) - (p_x - p_y)}{\sqrt{\frac{p_x(1-p_x)}{n_x} + \frac{p_y(1-p_y)}{n_y}}}$</p> <p>9. $Z = \frac{(\hat{p}_x - \hat{p}_y) - (p_x - p_y)}{\sqrt{\bar{p}(1-\bar{p}) \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}}$</p>	<p>10. $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$</p> <p>11. $F = \frac{s_1^2}{s_2^2}$</p> <p>12. $\chi^2 = \sum \frac{(O - E)^2}{E}, \quad E = np$</p>

Nonparametric Statistics:

1. Sign Test:

Small sample: $X =$ Number of (+) signs [or (-) signs]

Large sample: $Z = \frac{2X - n}{\sqrt{n}}$

2. Wilcoxon Signed-rank:

Small sample: $W = \min(\sum(+), \sum(-))$

Large sample: $Z = \frac{W - \mu_W}{\sigma_W}$, $\mu_W = \frac{n(n+1)}{4}$, $\sigma_W = \sqrt{\frac{n(n+1)(2n+1)}{24}}$

3. Wilcoxon Rank Sum Test:

Small sample: $U = R - \frac{n(n+1)}{2}$