
UNIVERSITI SAINS MALAYSIA

First Semester Examination
2014/2015 Academic Session

December 2014/January 2015

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. The number of road accident over a long festivity weekend in 2013 have been recorded and the results are shown below.

6	15	13	36	25	18	12	18	5	30
24	7	10	22	33	24	19	14	12	9

- (a) Construct a frequency table with the first class of 1-7. Compare the mean calculated from the original data and that calculated from the frequency table.
- (b) Find the 5-number summary and construct a boxplot for the data set. What can you say about the distribution of the number of road accident?

[30 marks]

1. *Bilangan kemalangan jalanraya semasa perayaan hujung minggu yang panjang pada tahun 2013 telah dicatat dan keputusan ditunjukkan seperti di bawah.*

6	15	13	36	25	18	12	18	5	30
24	7	10	22	33	24	19	14	12	9

- (a) *Bina jadual kekerapan dengan kelas pertama adalah 1-7. Bandingkan purata yang dihitung daripada data asal dengan nilai yang dihitung daripada jadual kekerapan.*
- (b) *Cari ringkasan 5-nombor dan bina plot kotak bagi data tersebut. Apa yang boleh anda katakan tentang taburan bagi bilangan kemalangan jalanraya?*

[30 markah]

2. (a) During a piano competition, each contestant has to perform five classical music pieces. She has a choice of five compositions by Beethoven, four compositions by Mozart and six compositions by Chopin. To impress the judges he decides to perform three difficult compositions by Chopin and one each by Beethoven and Mozart. How many different possible combinations of piano compositions can he perform?
- (b) (i) The probability that a family owns a coffee maker is 0.34, that they own a slim-45 inch television is 0.58, and that they own both a coffee maker and a slim-45 inch television is only 0.29. What is the probability that a randomly selected family owns a coffee maker or a slim-45 inch television?
- (ii) M&M and Smarties are two different brands of milk chocolates in a crisp coloured shell. Each piece of confectionery is about the same size and each brand comes in a mixture of colours. A large bowl contains a mixture of the two brands in a ratio of three M&M to two Smarties in just four colours. The proportion of M&M which are red, yellow, orange and green are 0.3, 0.4, 0.1 and 0.2 respectively, while the corresponding proportions of Smarties are 0.25, 0.2, 0.3 and 0.25. If a piece of chocolate is chosen at random from the bowl, what is the probability that it is an M&M brand if it is a green colour.

[30 marks]

2. (a) *Semasa pertandingan piano, setiap peserta perlu mempersembahkan lima buah muzik klasik. Peserta mempunyai pilihan lima ciptaan oleh Beethoven, empat ciptaan oleh Mozart dan enam ciptaan oleh Chopin. Untuk menarik hati para hakim beliau mengambil keputusan untuk mempersembahkan tiga ciptaan yang sukar oleh Chopin dan masing-masing satu ciptaan oleh Beethoven dan Mozart. Berapa banyak kombinasi ciptaan piano yang mungkin dipersembahkan?*
- (b) (i) *Kebarangkalian bahawa sebuah keluarga memiliki sebuah mesin pembancuh kopi adalah 0.34, bahawa mereka memiliki sebuah televisyen nipis 45 inci adalah 0.58, dan bahawa mereka memiliki kedua-dua mesin pembancuh kopi dan televisyen nipis 45 inci adalah hanya 0.29. Apakah kebarangkalian bahawa sebuah keluarga yang dipilih secara rawak memiliki mesin pembancuh kopi atau televisyen nipis 45 inci?*
- (ii) *M&M dan Smarties adalah dua jenama susu coklat yang berlainan dengan kulit berwarna yang rangup. Setiap biji susu coklat mempunyai saiz hamper sama dan setiap jenama datang dengan campuran berwarna-warni. Semangkuk besar mengandungi campuran dua jenama susu coklat dengan nisbah tiga M&M dan dua Smarties, hanya dalam empat warna. Pecahan M&M yang berwarna merah, kuning, oren dan hijau adalah masing-masing 0.3, 0.4, 0.1 dan 0.2, manakala pecahan yang sepadan bagi Smarties adalah 0.25, 0.20, 0.30 dan 0.25. Sekiranya sebiji coklat dipilih secara rawak dari mangkuk tersebut, apakah kebarangkalian bahawa ia adalah berjenama M&M jika ia berwarna hijau?*

[30 marks]

3. (a) The table below describes the probability distribution for the number of rotten mango in a packet of four, two days after it was put on the shelf. Find the mean and standard deviation for the distribution shown.

x	0	1	2	3	4
$P(x)$	0.265	0.412	0.239	0.076	0.008

- (b) (i) In a certain faculty of a university, 32% of the students have not been to foreign countries. If 10 students are selected at random from the faculty, what is the probability that at most four have not travelled overseas?
- (ii) At an archaeological excavation site, workers find an average of 3 prehistoric artifacts per 10 litres of sediment. If a student is sent to dig up 50 litres of sediment, find the probability that he or she will find at least ten artifacts.

[20 marks]

3. (a) *Jadual di bawah menerangkan taburan kebarangkalian bagi bilangan biji mangga yang rosak dalam satu paket mengandungi empat biji mangga, dua hari selepas ia diletak di atas rak. Cari min dan sisihan piawai bagi taburan yang ditunjuk.*

x	0	1	2	3	4
$P(x)$	0.265	0.412	0.239	0.076	0.008

- (b) (i) *Dalam suatu fakulti di sebuah universiti, 32% daripada pelajar belum pernah ke negara asing. Sekiranya 10 orang pelajar dipilih secara rawak daripada fakulti tersebut, apakah kebarangkalian bahawa paling ramai empat orang belum pernah melancong ke luar Negara.*
- (ii) *Di suatu tapak penggalian arkeologi, pekerja mendapat secara purata 3 artifak pra-sejarah bagi setiap 10 liter mendakan. Jika seorang pelajar dihantar untuk menggali 50 liter mendakan, cari kebarangkalian bahawa beliau akan mendapat sekurang-kurangnya 10 artifak.*

[20 markah]

4. (a) A machine is adjusted to fill boxes of cereal with mean 250g and standard deviation 12g. Assume the amount of fill follows a normal distribution.
- (i) What is the probability that a randomly selected cereal box contains more than 256g?
 - (ii) If 40 boxes of cereal are randomly selected, find the probability that the mean weight is less than 245g.
- (b) The Registrar at a local university would like to estimate, within three minutes, the mean travel time to work for all lecturers who live within twenty kilometers of the university. He knows that the standard deviation for all travel times is 12.5 minutes. If he wants to achieve a 95% confidence level on the estimate, how many lecturers should he sample?

[20 marks]

4. (a) *Sebuah mesin diselaraskan untuk mengisi kotak bijirin dengan purata 250g dan sisihan piawai 12g. Andaikan kandungan mempunyai taburan normal.*
- (i) *Apakah kebarangkalian bahawa satu kotak bijirin yang dipilih secara rawak mengandungi lebih daripada 256g?*
 - (ii) *Sekiranya 40 kotak bijirin dipilih secara rawak, cari kebarangkalian bahawa purata berat adalah kurang daripada 245g.*
- (b) *Pendaftar di sebuah universiti tempatan ingin menganggar, dalam lingkungan tiga minit, purata waktu perjalanan ke tempat kerja bagi semua pensyarah yang tinggal dalam lingkungan dua puluh kilometer dari universiti. Dia mengetahui bahawa sisihan piawai untuk semua waktu perjalanan adalah 12.5 minit. Sekiranya dia ingin mencapai 95% aras keyakinan bagi nilai anggaran, berapa ramai pensyarah yang patut diambil sebagai sampel?*

[20 markah]

5. (a) Of 300 randomly interviewed students during the convocation period in August 2013, only 75 said that they have successfully obtained a job and have already started working. Find a 99% confidence interval for the true proportion of all students who have obtained a job and started working prior to convocation.
- (b) The following data represent the number of minutes the Air Asia flights from Kuala Lumpur arriving later than originally scheduled at Penang International Airport. The data are in order of arrival during a particular 5-weekday period.

12	3	9	18	10	20	6	5	15	12
24	9	14	16	18	21	25	16	22	28

- (i) Find the 95% confidence interval for the variance of minutes of late arrival.
- (ii) If 8% of the flights arrive more than 20 minutes late, what is the probability that of 200 flights, more than 15 flights arrive more than 20 minutes late.

[20 marks]

5. (a) *Daripada 300 orang pelajar yang ditemuduga secara rawak semasa tempoh konvokesyen pada Ogos 2013, hanya 75 orang menyatakan bahawa mereka telah berjaya memperoleh pekerjaan dan telah mula bekerja. Cari 99% selang keyakinan bagi peratusan sebenar pelajar yang telah mempunyai pekerjaan dan telah mula bekerja sebelum konvokesyen.*
- (b) *Data berikut mewakili bilangan minit penerbangan Air Asia dari Kuala Lumpur yang tiba lewat daripada yang dijadualkan di Lapangan terbang Antarabangsa Pulau Pinang. Data adalah mengikut turutan ketibaan bagi tempoh 5-hari bekerja yang tertentu.*

12	3	9	18	10	20	6	5	15	12
24	9	14	16	18	21	25	16	22	28

- (i) *Cari 95% selang keyakinan bagi varians bilangan minit ketibaan lewat*
- (ii) *Sekiranya 8% daripada bilangan penerbangan tiba lewat lebih daripada 20 minit, apakah kebarangkalian bahawa daripada 200 penerbangan, lebih daripada 15 penerbangan tiba lewat lebih daripada 20 minit?*

[20 markah]

6. An economist wants to see if the mean income during the first year of work for university graduates who took statistics course exceeds that of those who did not. The mean first-year income of students who have not taken statistics is assumed to be RM26500 and that the standard deviation of the distribution of first-year incomes is assumed to be RM6250. The economist surveys 25 recent graduates who have taken a statistics course during university years and finds that the average first year income is RM28900.

Calculate the observed p-value. What can be said about the amount of evidence to support the claim that the mean income for the first year of work for university graduates who took statistics course exceeds that of those who did not?

[20 marks]

6. *Seorang pakar ekonomi ingin melihat sekiranya purata pendapatan semasa tahun pertama bekerja bagi graduan universiti yang mengambil kursus statistik melebihi pendapatan mereka yang tidak mengambil kursus tersebut. Purata pendapatan tahun pertama bagi para pelajar yang tidak mengambil statistik diandaikan sebanyak RM26500 dan bahawa sisihan piawai bagi taburan untuk pendapatan tahun pertama diandaikan sebanyak RM6250. Pakar ekonomi tersebut menjalankan kaji selidik terhadap 25 graduan yang telah mengambil kursus statistik semasa di universiti dan mendapati bahawa purata pendapatan tahun pertama adalah RM28900.*

Hitung nilai-p yang terhasil. Apakah yang boleh dikatakan tentang kuantiti bukti untuk menyokong dakwaan bahawa purata pendapatan semasa tahun pertama bekerja bagi graduan universiti yang mengambil kursus statistik melebihi nilai bagi mereka yang tidak mengambil kursus tersebut.

[20 markah]

7. The management of a district council is to modify its waste collection policy. It was believed that although terrace housing estates generally dispose less rubbish as compared to those dispose by the high rise housing estates, the amount of rubbish dispose by the terrace housing estates are more varying from day to day. To test the two claims, 16 high rise estates and 14 terrace estates of approximately equal number of housing units were randomly selected and the amount of rubbish (in kg) collected on a given day is recorded below.

	High rise				Terrace			
48	42	50	49	42	41	35	33	
44	59	52	43	38	48	60	45	
52	45	47	53	37	54	40	53	
53	45	51	58	39	61			

- (a) At 5% significance level, is there enough evidence to support the claim that terrace housing estates have greater variation in the amount of daily rubbish being disposed? Use rejection region approach.
- (b) At 5% significance level, is there enough evidence to support the claim that high rise housing estates dispose greater amount of rubbish as compared to the amount dispose by the terrace housing estates? Use p-value approach.

[30 marks]

7. Pihak pengurusan suatu majlis daerah akan mengubahsuai polisi pengutipan sisa mereka. Adalah dipercayai bahawa walaupun kawasan perumahan teres pada amnya kurang membuang sampah berbanding pembuangan oleh kawasan rumah pangsa, kuantiti sampah yang dibuang oleh kawasan perumahan teres adalah lebih berubah-ubah dari hari ke hari. Untuk menguji dua dakwaan tersebut, 16 perumahan pangsa dan 14 perumahan teres, yang mempunyai bilangan unit rumah yang hampir sama, telah dipilih secara rawak dan kuantiti sampah (dalam kg) yang dikutip pada suatu hari tertentu direkodkan seperti di bawah.

	Pangsa				Teres			
48	42	50	49	42	41	35	33	
44	59	52	43	38	48	60	45	
52	45	47	53	37	54	40	53	
53	45	51	58	39	61			

- (a) Pada aras kesignifikanan 5%, adakah terdapat bukti yang cukup untuk menyokong dakwaan bahawa kawasan perumahan teres mempunyai variasi yang lebih besar bagi kuantiti sampah harian yang dibuang? Gunakan kaedah rantau penolakan.
- (b) Pada aras kesignifikanan 5%, adakah terdapat bukti yang cukup untuk menyokong dakwaan bahawa kawasan rumah pangsa membuang kuantiti sampah yang lebih banyak berbanding kuantiti yang dibuang oleh kawasan rumah teres. Gunakan kaedah nilai-p.

[30 markah]

8. (a) The Penang Global Tourism have been given the mandate to look into factors affecting the level of satisfaction among foreign tourists. A survey was conducted on sample of foreign tourists and the distribution on the level of satisfaction on public infrastructure and facilities are given in the table below.

Satisfaction Level of Public Infrastructure and Facilities	Overall Level of Satisfaction		
	Not Satisfied	Neutral	Satisfied
Not satisfied	29	14	8
Neutral	13	29	16
Satisfied	17	21	38

Conduct appropriate test to investigate if the level of satisfaction of foreign tourists on public infrastructure and facilities is related to the level of overall satisfaction.

- (b) A Northern Region manager of McDonalds fast food restaurant strongly feels that the time required for drive-through services at Penang Island's Greenlane branch is shorter than that at Penang's Seberang Jaya branch. The manager randomly observed eight cars using the drive-through services at both branches and records the time for completing the services. The following results were obtained.

Greenlane	115	126	49	56	98	76	78	95
Seberang Jaya	120	67	89	97	124	68	72	144

Using a suitable non-parametric test and at 5% significance level, is there sufficient evidence to support the claim made by the regional manager?

[30 marks]

8. (a) *Pelancongan Global Penang telah diberi mandat untuk melihat faktor-faktor yang mempengaruhi tahap kepuasan di kalangan pelancong asing. Satu kajian telah dijalankan ke atas sampel pelancong asing dan taburan tahap kepuasan terhadap infrastruktur dan kemudahan awam adalah diberikan dalam jadual di bawah.*

<i>Tahap kepuasan terhadap infrastruktur dan kemudahan awam</i>	<i>Tahap kepuasan keseluruhan</i>		
	<i>Tidak memuaskan</i>	<i>Neutral</i>	<i>Memuaskan</i>
<i>Tidak memuaskan</i>	29	14	8
<i>Neutral</i>	13	29	16
<i>Memuaskan</i>	17	21	38

Jalankan ujian yang sesuai untuk menyelidik sekiranya tahap kepuasan pelancong asing terhadap infrastruktur dan kemudahan awam adalah berkaitan dengan tahap kepuasan keseluruhan.

- (b) Seorang pengarah Zon Utara restoran makanan segera McDonalds amat merasakan bahawa masa yang perlu untuk perkhidmatan pandu-lalu di cawangan Greenlane Pulau Pinang adalah lebih pendek daripada masa yang diperlukan di cawangan Seberang Jaya. Pengurus tersebut memilih secara rawak lapan kereta yang menggunakan perkhidmatan pandu-lalu di kedua-dua cawangan dan mencatatkan masa untuk menyempurnakan perkhidmatan tersebut. Keputusan berikut telah diperolehi.

<i>Greenlane</i>	115	126	49	56	98	76	78	95
<i>Seberang Jaya</i>	120	67	89	97	124	68	72	144

Menggunakan ujian tidak-berparameter yang sesuai dan pada aras kesignifikanan 5%, adakah terdapat bukti yang cukup untuk menyokong dakwaan yang dibuat oleh pengurus zon tersebut?

[30 markah]

APPENDIX

Confidence Interval

$\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ $\bar{X} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$ $\bar{d} \pm t_{\frac{\alpha}{2}} \frac{s_d}{\sqrt{n_d}}$ $b \pm t_{\frac{\alpha}{2}} s_b$	$(\bar{X} - \bar{Y}) \pm t_{\alpha/2} S_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$
$\hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $(\hat{p}_x - \hat{p}_y) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}$ $(\bar{X} - \bar{Y}) \pm Z_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$ $(\bar{X} - \bar{Y}) \pm t_{\alpha/2} \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$	$\left(\frac{(n-1)s^2}{\chi_{\frac{\alpha}{2}, n-1}^2}, \frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}, n-1}^2} \right)$ $\left(\frac{s}{Z_{\frac{\alpha}{2}} \sqrt{2n}}, \frac{s}{Z_{\frac{\alpha}{2}} \sqrt{2n}} \right)$ $\left(\frac{s_1^2}{s_2^2} F_{1-\frac{\alpha}{2}, (v_2, v_1)}, \frac{s_1^2}{s_2^2} F_{\frac{\alpha}{2}, (v_2, v_1)} \right)$

Test Statistic

$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ $T = \frac{\bar{X} - \mu}{s / \sqrt{n}}$ $T = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n_d}}$ $T = \frac{b - \beta_1}{s_b}$ $T = r \sqrt{\frac{n-2}{1-r^2}}$ $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$	$Z = \frac{s - \sigma}{\sigma / \sqrt{2n}}$ $Z = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$ $T = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{S_p^2 \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}}$ $S_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$ $F = \frac{s_x^2}{s_y^2}$	$Z = \frac{(\hat{p}_x - \hat{p}_y) - (p_x - p_y)}{\sqrt{\hat{p}(1-\hat{p}) \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}}$ $T = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$ $dk = \frac{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y} \right)^2}{\frac{\left(\frac{s_x^2}{n_x} \right)^2}{n_x - 1} + \frac{\left(\frac{s_y^2}{n_y} \right)^2}{n_y - 1}}$ $\chi^2 = \sum \frac{(O-E)^2}{E}, \quad E = np$
---	--	---

Table A5 Table of Critical *T* Values for Wilcoxon's Signed-Ranks and Matched-Pairs Signed-Ranks Test

<i>n</i>	One-tailed level of significance				One-tailed level of significance				
	.05	.025	.01	.005	.05	.025	.01	.005	
	Two-tailed level of significance				Two-tailed level of significance				
	.10	.05	.02	.01	.10	.05	.02	.01	
5	0	-	-	-	28	130	116	101	91
6	2	0	-	-	29	140	126	110	100
7	3	2	0	-	30	151	137	120	109
8	5	3	1	0	31	163	147	130	118
9	8	5	3	1	32	175	159	140	128
10	10	8	5	3	33	187	170	151	138
11	13	10	7	5	34	200	182	162	148
12	17	13	9	7	35	213	195	173	159
13	21	17	12	9	36	227	208	185	171
14	25	21	15	12	37	241	221	198	182
15	30	25	19	15	38	256	235	211	194
16	35	29	23	19	39	271	249	224	207
17	41	34	27	23	40	286	264	238	220
18	47	40	32	27	41	302	279	252	233
19	53	46	37	32	42	319	294	266	247
20	60	52	43	37	43	336	310	281	261
21	67	58	49	42	44	353	327	296	276
22	75	65	55	48	45	371	343	312	291
23	83	73	62	54	46	389	361	328	307
24	91	81	69	61	47	407	378	345	322
25	100	89	76	68	48	426	396	362	339
26	110	98	84	75	49	446	415	379	355
27	119	107	92	83	50	466	434	397	373