
UNIVERSITI SAINS MALAYSIA

Second Semester Examination
2015/2016 Academic Session

June 2016

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

Duration : 3 hours
[Masa : 3 jam]

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

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

Instructions: Answer **TEN** (10) questions.

Arahan: Jawab **SEPULUH** (10) 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. Biologists from USM are investigating fish in the Lake Harapan. They trap the fish in nets, measure them, then release them again. For each fish, they record its species (snapper, pomfret, threadfin), its weight in grams and its length in millimeters.
- (a) State the type of each three variables.
- (b) Explain briefly, using a sketch, how to display these three variables on a single diagram.
- [6 marks]

1. *Ahli biologi dari USM sedang menyiasat ikan di Tasik Harapan. Mereka memerangkap ikan tersebut dalam jaring, mengukur, kemudian melepaskan ikan tersebut. Bagi setiap ikan, mereka merakam spesies (jenahak, bawal, senangin), berat dalam gram dan panjangnya dalam milimeter.*
- (a) *Nyatakan jenis ketiga-tiga pembolehubah.*
- (b) *Terangkan secara ringkas, dengan menggunakan lakaran, bagaimana untuk memaparkan ketiga-tiga pembolehubah pada gambar rajah.*
- [6 markah]

2. In a study to compare two different methods of teaching reading, a class of 30 children was divided into two groups at random. Each group was taught reading using one method. After six months the two groups were given the same test. One pupil in the first group was ill and missed the test. The scores are given below.

Method 1	Method 2
80 85 79 81 89	79 62 75 68 62
73 50 71 76 87	73 76 81 73 61
72 85 86 85	72 68 75 63 62

Draw a box and whisker plots for the two groups using the same scale. Comment on the differences between the two groups.

[10 marks]

2. Dalam satu kajian untuk membandingkan dua kaedah pengajaran pembacaan yang berbeza, kelas yang mempunyai 30 kanak-kanak telah dibahagikan kepada dua kumpulan secara rawak. Setiap kumpulan telah diajar membaca menggunakan satu kaedah. Selepas enam bulan kedua-dua kumpulan diberi ujian yang sama. Seorang murid dalam kumpulan pertama sakit dan terlepas ujian. Skor yang diperoleh diberikan di bawah.

Kaedah 1	Kaedah 2
80 85 79 81 89	79 62 75 68 62
73 50 71 76 87	73 76 81 73 61
72 85 86 85	72 68 75 63 62

Lukis plot kotak dan misai bagi kedua-dua kumpulan dengan menggunakan skala yang sama. Komen perbezaan antara kedua-dua kumpulan.

[10 markah]

3. It is estimated that heights of adult males are normally distributed with a mean of 70.0 inches and a standard deviation of 3.5. In one state, the law requires a person to be 68.0 inches or taller to become a firefighter.
- (a) What percentage of adult males will meet this height requirement for becoming a firefighter?
- (b) A person who was denied to become a firefighter learned that his height was at the 25th percentile. What is his actual height?

[6 marks]

3. Dianggarkan ketinggian lelaki dewasa adalah bertaburan normal dengan min 70.0 inci dan sisihan piawai 3.5. Dalam satu negara, undang-undang bagi seseorang untuk menjadi bomba memerlukan ketinggian 68.0 inci atau lebih.
- (a) Berapa peratus daripada lelaki dewasa akan memenuhi keperluan ketinggian untuk menjadi bomba?
- (b) Seseorang yang telah dinafikan untuk menjadi bomba tahu bahawa ketinggian beliau adalah pada peratusan ke-25. Apakah ketinggian sebenar beliau?

[6 markah]

4. Among all users of a gym facility, 70% of them use the treadmill regularly, 45% use the swimming pool regularly, and 35% use both the treadmill and swimming pool regularly. Suppose that you just met a person who uses the gym facility.
- What is the probability that the person whom you met regularly uses the treadmill, swimming pool, or both?
 - What is the probability that the person whom you met is a regular user of the swimming pool, given that he or she uses the treadmill regularly?
 - Are the two events, “the person you met regularly uses the treadmill” and “the person you met regularly uses the swimming pool” independent? Justify your answer.

[8 marks]

4. *Antara semua pengguna kemudahan gimnasium, 70% daripada mereka menggunakan mesin larian setempat dengan kerap, 45% menggunakan kolam renang secara tetap, dan 35% menggunakan kedua-dua mesin larian setempat dan kolam renang secara kerap. Katakan anda bertemu seseorang yang menggunakan kemudahan gimnasium.*
- Apakah kebarangkalian bahawa orang yang anda bertemu dengan kerap menggunakan mesin larian setempat, kolam renang, atau kedua-duanya?*
 - Apakah kebarangkalian bahawa orang yang anda bertemu adalah pengguna kerap kolam renang, diberikan dia menggunakan mesin larian setempat dengan kerap?*
 - Adakah kedua-dua peristiwa, "orang yang anda bertemu dengan kerap menggunakan mesin larian setempat " dan "orang yang anda bertemu dengan kerap menggunakan kolam renang" tak bersandar? Jelaskan jawapan anda.*

[8 markah]

5. (a) Haneen is taking a multiple choice exam. There are 5 questions, and each question has three possible choices. Haneen did not study for the exam and so she randomly guesses her answer to each question. Compute the probability that Haneen passes the exam. (Note: Haneen must answer at least 3 questions correctly in order to pass.)
- (b) Suppose, instead, that there are 72 questions in the exam. (Again, each question has three possible choices.) Use the normal approximation to the binomial to compute the probability that Haneen passes. (Note: Haneen must answer at least 36 questions correctly in order to pass.)

[8 marks]

...5/-

5. (a) *Haneen mengambil peperiksaan aneka pilihan. Terdapat 5 soalan, dan setiap soalan mempunyai tiga pilihan yang mungkin. Haneen tidak belajar untuk peperiksaan dan dia secara rawak meneka jawapan bagi setiap soalan. Kirakan kebarangkalian bahawa Haneen lulus peperiksaan. (Perhatian: Haneen perlu menjawab sekurang-kurangnya 3 soalan dengan betul untuk lulus.)*
- (b) *Katakan, sebaliknya, bahawa terdapat 72 soalan peperiksaan. (Sekali lagi, setiap soalan mempunyai tiga pilihan yang mungkin). Gunakan penghampiran normal kepada binomial untuk mengira kebarangkalian bahawa Haneen lulus. (Perhatian: Haneen perlu menjawab sekurang-kurangnya 36 soalan dengan betul untuk lulus).*

[8 markah]

6. Twist-off caps are very common for bottled drinks these days. However, many consumers are concerned about the cleanliness of the contents of the bottles (bottles can easily be opened before a customer purchases the drinks). To address this issue, a major soft drink company intends to conduct a survey, and if they determine that more than 75% of all consumers have the concern, they will switch back to using regular caps.
- (a) If the company wishes their estimate of the proportion of the customers with the concern to be within 0.06 of the true value with 95% certainty, how many customers should they survey? Do not use any information from (b) below to answer this question.
- (b) Suppose that the survey was conducted with 250 customers, and that 195 of them said that they were concerned about twist-off caps. Construct a 95% confidence interval for the true proportion.

[12 marks]

6. *Keadaan penutup botol tertanggal sangat biasa untuk minuman botol hari ini. Walau bagaimanapun, ramai pengguna bimbang tentang kebersihan kandungan botol (botol boleh dibuka dengan mudah sebelum pelanggan membeli minuman). Untuk menangani isu ini, syarikat utama minuman ringan utama bercadang untuk menjalankan kaji selidik, dan jika mereka mendapati bahawa lebih daripada 75% daripada semua pengguna mempunyai kebimbangan, mereka akan beralih kembali menggunakan penutup biasa.*
- (a) *Jika syarikat ingin anggaran perkadaran pelanggan dengan kebimbangan yang berada dalam 0.06 daripada nilai sebenar dengan 95% kepastian, berapakah bilangan pelanggan yang mereka perlu tinjau? Jangan gunakan sebarang maklumat dari (b) di bawah untuk menjawab soalan ini.*
- (b) *Katakan bahawa kaji selidik itu dijalankan dengan 250 pelanggan, dan 195 daripada mereka berkata bahawa mereka bimbang tentang penutup tertanggal. Binakan 95% selang keyakinan bagi perkadaran sebenar.*

[12 markah]

...6/-

7. A firm makes sprinkler systems for the prevention of fires in buildings. It is important that each sprinkler is activated quickly when a fire starts. The engineer in charge assumes that their activation times (in seconds) are approximately normally distributed, with unknown mean μ and unknown variance σ^2 . He has designed the sprinklers to have $\mu = 25$. If $\mu > 25$ then he needs to change the design. He takes a random sample of 50 sprinklers, starts a fire near each one, and records their activation times x_i in seconds. He finds that

$$\sum_{i=1}^{50} x_i = 1396 \text{ and } \sum_{i=1}^{50} x_i^2 = 45165$$

- State the engineer's null and alternative hypotheses.
- Find the sample mean and the sample standard deviation of the data.
- Carry out the appropriate hypothesis test at the 10% significance level, and report the conclusion (use the p -value method).

[14 marks]

7. Sebuah syarikat membuat sistem semburan untuk mencegah kebakaran dalam bangunan. Ia adalah penting bahawa setiap semburan diaktifkan dengan cepat apabila kebakaran bermula. Jurutera yang bertanggungjawab menganggap bahawa masa pengaktifan (dalam saat) adalah menghampiri taburan normal, dengan min tidak diketahui μ dan varians tidak diketahui σ^2 . Beliau telah merekabentuk penyembur untuk mempunyai $\mu = 25$. Jika $\mu > 25$ maka dia perlu menukar rekabentuk. Beliau mengambil sampel rawak 50 unit penyembur, memulakan kebakaran berhampiran antara satu sama lain, dan merekodkan masa pengaktifan mereka x_i dalam saat. Beliau mendapati bahawa

$$\sum_{i=1}^{50} x_i = 1396 \text{ and } \sum_{i=1}^{50} x_i^2 = 45165$$

- Nyatakan hipotesis nol dan alternatif jurutera.
- Cari min sampel dan sisihan piawaian sampel data.
- Jalankan ujian hipotesis yang sesuai pada aras keertian 10%, dan laporkan kesimpulan (gunakan kaedah nilai- p).

[14 markah]

8. A study used X-ray computed tomography (CT) to collect data on brain volumes for a group of patients with obsessive-compulsive disorders and control group of healthy persons. Sample results (in mL) are given below for total brain volumes.

Obsessive-compulsive disorders: $n = 10, \bar{x} = 1390.03, s = 156.84$

Control group: $n = 10, \bar{x} = 1268.41, s = 137.97$

- (a) Test the claim that the population of total brain volumes for obsessive-compulsive disorders patients and the control group have different amount of variation. Use 0.05 significance level.
- (b) Based on (a), use a 0.05 significance level to test the claim that there is no difference between the mean for obsessive-compulsive disorders patients and the mean for healthy persons.

[14 marks]

8. *Satu kajian menggunakan sinaran-X tomografi berkomputer (CT) untuk mengumpul data isipadu otak untuk sekumpulan pesakit yang mengalami gangguan obsesif-kompulsif dan kumpulan kawalan orang sihat. Keputusan sampel (dalam mL) diberikan di bawah bagi isipadu keseluruhan otak.*

Gangguan obsesif-kompulsif: $n = 10, \bar{x} = 1390.03, s = 156.84$

Kumpulan Kawalan: $n = 10, \bar{x} = 1268.41, s = 137.97$

- (a) *Uji dakwaan bahawa populasi isipadu keseluruhan otak untuk pesakit gangguan obsesif-kompulsif dan kumpulan kawalan mempunyai jumlah variasi yang berbeza. Gunakan aras keertian 0.05.*
- (b) *Berdasarkan (a), gunakan aras keertian 0.05 untuk menguji dakwaan bahawa tidak ada perbezaan min antara pesakit gangguan obsesif-kompulsif dan orang yang sihat.*

[14 markah]

9. In a study of store checkout scanning systems, samples of purchases were used to compare the scanned prices to the posted prices. The accompanying table summarizes results for sample of 819 items. When stores use scanners to check out items, are the error rates the same for regular-priced items as they are advertised-special items? How might the behavior of consumers change if they believe that disproportionately more overcharges occur with advertised-special item? Use $\alpha = 0.05$.

	Regular-Priced Items	Advertised-special Items
Undercharge	20	7
Overcharge	15	29
Correct price	384	364

[10 marks]

9. *Dalam kajian ke atas sistem pengimbasan keluar kedai, sampel pembelian telah digunakan untuk membandingkan harga yang diimbas dengan harga yang dipamerkan. Jadual yang disertakan meringkaskan keputusan untuk sampel 819 item. Apabila kedai menggunakan pengimbas untuk memeriksa barang-barang, adakah kadar ralat yang sama untuk barangan harga biasa berbanding barangan diiklankan-khas? Bagaimana tingkah laku pengguna mungkin berubah jika mereka percaya bahawa lebihan harga berlaku dengan item yang diiklankan-istimewa? Guna $\alpha = 0.05$.*

	<i>Barangan harga biasa</i>	<i>Barangan iklan istimewa</i>
<i>Kurang harga</i>	20	7
<i>Lebih harga</i>	15	29
<i>Harga betul</i>	384	364

[10 markah]

10. In a study of techniques used to measure lung volumes, physiological data were collected for 10 subjects. The values given in the table are in liters and represent the measured functional residual capacities of the 10 subjects both in a sitting position and in a supine (lying) position. At the 0.05 significance level, use the appropriate nonparametric method to test the claim that there is no significant difference between the measurements taken in the two positions.

Sitting	2.96	4.65	3.27	2.50	2.59	5.97	1.74	3.51	4.37	4.02
Supine	1.97	3.05	2.29	1.68	1.58	4.43	1.53	2.81	2.70	2.70

[12 marks]

10. Dalam kajian ke atas teknik yang digunakan untuk mengukur isipadu paru-paru, data fisiologi telah dikumpulkan untuk 10 subjek. Nilai yang diberi dalam jadual adalah dalam liter dan ukuran kapasiti fungsi reja 10 subjek dalam posisi duduk dan berbaring. Pada aras keertian 0.05, gunakan kaedah tak berparameter yang bersesuaian untuk menguji dakwaan bahawa tidak terdapat perbezaan yang bererti di antara ukuran diambil pada kedua-dua kedudukan.

Duduk	2.96	4.65	3.27	2.50	2.59	5.97	1.74	3.51	4.37	4.02
Baring	1.97	3.05	2.29	1.68	1.58	4.43	1.53	2.81	2.70	2.70

[12 markah]

APPENDIX/ LAMPIRAN

$\bar{x} = \frac{\sum xf}{\sum f} = \frac{\sum x}{n}$ $s^2 = \frac{\sum(x^2 f) - \frac{(\sum xf)^2}{\sum f}}{\sum f - 1}$ $= \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n - 1}$	$S_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$ $\bar{p} = \frac{X + Y}{n_x + n_y}$
<p>Confidence Intervals:</p> <p>1. $\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ or $\bar{X} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$</p> <p>2. $\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$</p> <p>3. $\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$</p> <p>4. $(\bar{X} - \bar{Y}) \pm z_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$</p>	<p>5. $(\bar{X} - \bar{Y}) \pm t_{\alpha/2} \sqrt{S_p^2 \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}$</p> <p>6. $(\hat{p}_x - \hat{p}_y) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}$</p> <p>7. $\frac{(n-1)s^2}{\chi_{\alpha/2}^2}$ to $\frac{(n-1)s^2}{\chi_{1-\alpha/2}^2}$</p>

Test Statistics:

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

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

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

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

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

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

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

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

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

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

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

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

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

$$13. \chi^2 = \frac{(n-1)s^2}{\sigma^2}$$