
UNIVERSITI SAINS MALAYSIA

First Semester Examination
2012/2013 Academic Session

January 2013

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 nine** [9] questions.

[Arahan: Jawab **semua sembilan** [9] 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. (a) The following table gives the total number of DVDs sold at retail stores between 2003 and 2008.

Year	U.S. Retail Sales of DVDs (millions of DVDs)
2003	11.3
2004	15.1
2005	16.0
2006	16.3
2007	15.8
2008	15.2

State and describe the meanings of a variable, a measurement and a data set with reference to this table.

- (b) The following frequency distribution table gives the age distribution of drivers who were at fault in auto accidents that occurred during 1-week period in a city.

Age (years)	<i>f</i>
<i>18 to less than 20</i>	7
<i>20 to less than 25</i>	12
<i>25 to less than 30</i>	18
<i>30 to less than 40</i>	14
<i>40 to less than 50</i>	15
<i>50 to less than 60</i>	16
<i>60 and over</i>	35

- (i) Construct a relative frequency histogram for this table.
- (ii) In what way(s) is this histogram misleading?
- (iii) How can you change the frequency distribution so that the resulting histogram gives a clearer picture?
- (iv) Note that there were 50% more accidents in the *25 to less than 30* age group than in the *20 to less than 25* age group. Does this suggest that the older group of drivers in the city is more accident-prone than the younger group? What other explanation might account for the difference in accident rates?

[15 marks]

1. (a) Jadual berikut memberikan jumlah bilangan DVD yang dijual di kedai-kedai runcit antara 2003 dengan 2008.

Tahun	Jualan Runcit DVD di U.S. (juta DVD)
2003	11.3
2004	15.1
2005	16.0
2006	16.3
2007	15.8
2008	15.2

Dengan merujuk kepada jadual ini, nyatakan dan huraikan makna pembolehubah, ukuran dan set data.

- (b) Jadual taburan kekerapan berikut memberikan taburan umur pemandu yang bersalah dalam kemalangan yang berlaku dalam tempoh 1 minggu di bandar.

Umur (tahun)	f
18 hingga kurang daripada 20	7
20 hingga kurang daripada 25	12
25 hingga kurang daripada 30	18
30 hingga kurang daripada 40	14
40 hingga kurang daripada 50	15
50 hingga kurang daripada 60	16
60 dan ke atas	35

- (i) Bina satu histogram kekerapan relatif bagi jadual ini.
- (ii) Apakah yang mengelirukan dalam histogram ini?
- (iii) Apakah yang anda boleh ubah dalam taburan kekerapan supaya histogram yang dihasilkan memberi gambaran yang lebih jelas?
- (iv) Perhatikan bahawa kemalangan dalam kumpulan umur 25 hingga kurang daripada 30 melebihi 50% daripada kemalangan dalam kumpulan umur 20 hingga kurang daripada 25. Adakah ini mencadangkan bahawa kumpulan pemandu yang lebih tua di bandar adalah lebih terdedah kepada kemalangan daripada kumpulan yang lebih muda? Apa penjelasan lain yang mungkin boleh diambil kira dalam perbezaan kadar kemalangan?

[15 markah]

2. Actuaries at an insurance company must determine a premium for a new type of insurance. A random sample of 40 potential purchasers of this type of insurance were found to have suffered the following values of losses (in dollars) during the past year. These losses would have been covered by the insurance if it were available.

100	32	0	0	470	50	0	14589	212	93
0	0	1127	421	0	87	135	420	0	250
12	0	309	0	177	295	501	0	143	0
167	398	54	0	141	0	3709	122	0	0

- (a) Find the mean, median and mode of these 40 losses.
- (b) Which of the mean, median or mode is largest?
- (c) Construct a box-and-whisker plot for these data and describe the skewness if any.
- (d) Which measure of central tendency should the actuaries use to determine the premium for this insurance? Explain.

[12 marks]

2. Para akturi di sebuah syarikat insurans mesti menentukan premium untuk insurans jenis baru. Satu sampel rawak 40 pembeli berpotensi untuk insurans jenis ini telah didapati mengalami nilai-nilai kerugian (dalam dolar) berikut sepanjang tahun lalu. Kerugian ini akan dilindungi oleh insurans jika ia telah disediakan.

100	32	0	0	470	50	0	14589	212	93
0	0	1127	421	0	87	135	420	0	250
12	0	309	0	177	295	501	0	143	0
167	398	54	0	141	0	3709	122	0	0

- (a) Cari min, median dan mod untuk 40 kerugian ini.
- (a) Min, median dan mod manakah yang memberi nilai terbesar?
- (c) Bina satu plot “kotak dan misai” untuk data ini dan huraikan kepencongan jika ada.
- (d) Manakah ukuran kecenderungan memusat yang harus digunakan oleh para aktuari untuk menentukan premium insurans ini? Jelaskan.

[12 markah]

3. Scott offers you the following game: You will roll two fair dice. If the sum of the two numbers obtained is 2, 3, 4, 9, 10, 11 or 12, Scott will pay you \$20. However, if the sum of the two numbers is 5, 6, 7, or 8, you will pay Scott \$20. Scott points out that you have seven winning numbers and only four losing numbers. Is this game fair to you? Should you accept this offer? Support your conclusion with appropriate calculations

[12 marks]

3. *Scott menawarkan anda permainan berikut: Anda akan melempar dua dadu adil. Jika jumlah dua nombor yang diperolehi ialah 2, 3, 4, 9, 10, 11 atau 12, Scott akan membayar anda \$20. Walau bagaimanapun, jika jumlah dua nombor adalah 5, 6, 7, atau 8, anda akan membayar Scott \$20. Scott menunjukkan bahawa anda mempunyai tujuh nombor yang menang dan hanya empat nombor yang kalah. Adakah permainan ini adil kepada anda? Haruskah anda menerima tawaran ini? Sokong kesimpulan anda dengan pengiraan yang sesuai.*

[12 markah]

4. An old car is not parked inside the garage at night. On the morning after a rainy night, the probability that the car engine will not start is $\frac{1}{3}$. On the morning after a rainless night, the probability that the car engine will not start is $\frac{1}{25}$. The event of the engine will start each morning does not depend on the event of the engine starting on other days.
- (a) For 6 consecutive nights, the rain falls. Find the probability that the car engine will not start on at least 2 of the 6 mornings.
- (b) During a rainy season, there were 32 rainy nights. Determine the probability that the car engine will not start on less than 16 of the 32 mornings.
- (c) During a dry season, there were 100 rainless nights. Find the probability that the car engine will not start on 5 or more of the 100 mornings.

[13 marks]

4. *Sebuah kereta lama tidak ditempatkan di dalam garaj pada waktu malam. Pada pagi berikutan suatu malam berhujan, kebarangkalian enjin kereta ini tidak boleh dihidupkan ialah $\frac{1}{3}$. Pada pagi berikutan suatu malam tidak berhujan pula, kebarangkalian enjin kereta ini tidak boleh dihidupkan ialah $\frac{1}{25}$. Peristiwa enjin kereta ini boleh dihidupkan setiap pagi adalah tidak bersandar pada peristiwa enjin kereta itu dihidupkan pada pagi-pagi yang lepas.*
- (a) *Pada malam hujan turun berturut-turut. Tentukan kebarangkalian bahawa enjin kereta ini tidak boleh dihidupkan pada sekurang-kurangnya 2 daripada 6 pagi ini.*
- (b) *Pada suatu musim hujan, terdapat 32 malam berhujan. Tentukan kebarangkalian bahawa enjin kereta ini tidak boleh dihidupkan pada kurang daripada 16 di antara 32 pagi ini.*
- (c) *Pada suatu musim kemarau, terdapat 100 malam tidak berhujan. Tentukan kebarangkalian bahawa enjin kereta ini tidak boleh dihidupkan pada 5 atau lebih daripada 100 pagi ini.*

[13 markah]

5. A mail-order company promises its customers that the products ordered will be mailed within 72 hours after an order is placed. The quality control department at the company checks from time to time to see if this promise is fulfilled. Recently the quality control department took a sample of 50 orders and found that 35 of them were mailed within 72 hours of the placement of the orders.
- (a) Construct a 98% confidence interval for the percentage of all orders that are mailed within 72 hours of their placement.
- (b) Suppose the confidence interval obtained in part a is too wide. How can the width of this interval be reduced? Discuss all possible alternatives. Which alternative is the best?

[8 marks]

5. *Sebuah syarikat pesanan mel menjanjikan pelanggan mereka bahawa produk yang dipesan akan dihantar dalam masa 72 jam selepas pesanan telah dibuat. Jabatan kawalan kualiti di syarikat itu memeriksa dari masa ke semasa untuk melihat sama ada janji ini dipenuhi. Kebelakangan ini jabatan kawalan kualiti mengambil sampel sebanyak 50 pesanan dan mendapati bahawa 35 daripada mereka telah dihantar dalam tempoh 72 jam selepas pesanan telah dibuat.*
- (a) *Binakan selang keyakinan 98% untuk peratusan semua pesanan yang diposkan dalam tempoh 72 jam selepas pesanan telah dibuat..*
- (b) *Katakan selang keyakinan yang diperolehi dalam bahagian a adalah terlalu luas. Bagaimana lebar selang ini boleh dikurangkan? Bincangkan semua alternatif yang mungkin. Alternatif manakah yang terbaik?*

[8 markah]

6. The print on the package of 100-watt General Electric (GE) soft-white light bulbs states that these light bulbs have an average life of 750 hours. Assume that the standard deviation of the lengths of lives of these light bulbs is 50 hours. A sceptical consumer does not think these light bulbs last as long as the manufacturer claims, and she decides to test 64 randomly selected light bulbs. She has set up the decision rule that if the average life of these 64 light bulbs is less than 735 hours, then she will conclude that GE has printed too high an average length of life on packages and will write them a letter to that effect. Approximately what significance level is the consumer using? Approximately what significance level is she using if she decides that GE has printed too high an average length of life on the package if the average life of the 64 light bulbs is less than 700 hours? Interpret the values you get.

[10 marks]

6. *Cetakan pada pakej 100-watt elektrik mentol lampu putih menyatakan bahawa mentol ini mempunyai hayat purata sebanyak 750 jam. Anggapkan bahawa sisihan piawai untuk tempoh hayat mentol ini ialah 50 jam. Seorang pengguna tidak percaya mentol lampu ini boleh tahan selama tempoh yang diberi oleh pengilang, dan dia memutuskan untuk menguji 64 mentol lampu yang dipilih secara rawak. Beliau telah membuat keputusan bahawa jika hayat purata 64 mentol lampu adalah kurang daripada 735 jam, maka dia akan menyimpulkan bahawa GE telah mencetak tempoh hayat terlalu tinggi pada pakej dan akan menulis surat kepada mereka tentang keputusan yang didapati. Anggarkan, aras keertian yang digunakan? Anggarkan aras keertian yang dia gunakan jika dia memutuskan bahawa GE telah mencetak tempoh hayat terlalu tinggi di pakej jika hayat purata 64 mentol lampu adalah kurang daripada 700 jam? Tafsirkan nilai yang anda dapat.*

[10 markah]

7. A company psychologist wanted to test if company executives have job-related stress scores higher than those of university professors. He took a sample of 40 executives and 50 professors and tested them for job-related stress. The sample of 40 executives gave a mean stress scores of 7.6. The sample of 50 professors produced a mean stress score of 5.4. Assume that the standard deviations of the two populations are 0.8 and 1.3 respectively.
- (a) Construct a 99% confidence interval for the difference between the mean stress score of all executives and all professors.
- (b) Test at the 2.5% significance level whether the mean stress scores of all executives is higher than that of all professors.

[12 marks]

7. *Seorang ahli psikologi syarikat ingin menguji bahawa eksekutif syarikat mempunyai skor tekanan kerja lebih tinggi daripada skor tekanan profesor universiti. Beliau mengambil sampel sebanyak 40 eksekutif dan 50 profesor dan menguji tekanan kerja mereka. Sampel dengan 40 eksekutif memberi min skor tekanan kerja sebanyak 7.6. Sampel dengan 50 orang profesor menghasilkan min skor tekanan sebanyak 5.4. Andaikan bahawa sisihan piawai bagi dua populasi ini adalah 8 dan 1.3 masing-masing.*
- (a) *Binakan 99% selang keyakinan bagi perbezaan antara skor min tekanan semua eksekutif dan semua professor.*
- (b) *Uji pada aras keertian 2.5% sama ada min skor tekanan bagi semua eksekutif adalah lebih tinggi daripada semua profesor.*

[12 markah]

8. Each of five boxes contains a large (but unknown) number of red and green marbles. You have been asked to find if the proportions of red and green marbles are the same for each of the five boxes. You sample 50 times, with replacement, from each of the five boxes and observe 20, 14, 23, 30, and 18 red marbles, respectively. Can you conclude that the five boxes have the same proportions of red and green marbles? Use a 0.05 level of significance.

[10 marks]

8. *Setiap lima kotak mengandungi sebilangan besar (tetapi tidak diketahui) guli merah dan hijau. Anda telah diminta untuk mencari jika perkadaran guli merah dan hijau adalah sama bagi setiap lima kotak. Anda membuat 50 kali sampel, dengan penggantian, daripada setiap lima kotak dan mendapati 20, 14, 23, 30, dan 18 biji guli merah, masing-masing. Bolehkah anda buat kesimpulan bahawa lima kotak mempunyai kadaran yang sama bagi guli merah dan hijau? Gunakan aras keertian 0.05.*

[10 markah]

9. A private agency claims that the short course it offers significantly increases the writing speed of secretaries. The following table gives the writing speeds of eight secretaries before and after they attended this course.

Before	84	75	88	91	65	71	90	75
After	97	72	93	110	78	69	115	75

Using the 2.5% significance level, can you conclude that attending this course increases the writing speed of secretaries?

[8 marks]

9. *Sebuah agensi swasta mendakwa bahawa kursus pendek yang ditawarkan boleh meningkatkan kelajuan penulisan setiausaha. Jadual berikut memberikan kelajuan bertulis lapan setiausaha sebelum dan selepas mereka menghadiri kursus ini.*

Sebelum	84	75	88	91	65	71	90	75
Selepas	97	72	93	110	78	69	115	75

Dengan menggunakan aras keertian 2.5%, bolehkah anda menyimpulkan bahawa menghadiri kursus ini akan meningkatkan kelajuan menulis setiausaha?

[8 markah]

APPENDIX

$\bar{x} = \frac{\sum xf}{\sum f} = \frac{\sum x}{n}$ $s^2 = \frac{\sum(x^2 f) - \frac{\sum xf^2}{\sum f}}{\sum f - 1}$ $= \frac{\sum x^2 - \frac{\sum x^2}{n}}{n - 1}$	$S_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$ $\bar{p} = \frac{X + Y}{n_x + n_y}$
<p>Confidence Intervals:</p> <ol style="list-style-type: none"> 1. $\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ or $\bar{X} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$ 2. $\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$ 3. $\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$ 4. $(\bar{X} - \bar{Y}) \pm z_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$ 	<ol style="list-style-type: none"> 5. $(\bar{X} - \bar{Y}) \pm t_{\alpha/2} \sqrt{S_p^2 \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}$ 6. $(\hat{p}_x - \hat{p}_y) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_x(1 - \hat{p}_x)}{n_x} + \frac{\hat{p}_y(1 - \hat{p}_y)}{n_y}}$ 7. $\frac{(n-1)s^2}{\chi^2_{\alpha/2}}$ to $\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$

Test Statistics:

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

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

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

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

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

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

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

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

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

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

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

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

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

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