
UNIVERSITI SAINS MALAYSIA

Second Semester Examination
Academic Session 2005/2006

April/May 2006

MAT 161E – Elementary Statistics
[Statistik Permulaan]

Duration : 3 hours
[Masa : 3 jam]

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

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

Instructions: Answer **all four** [4] questions.

Arahan: Jawab **semua empat** [4] soalan].

1. (a) The following stem-and-leaf diagram gives the distances (in thousands of miles) driven during the past year by a sample of drivers in a city .

0	3 6 9
1	2 8 5 1 0 5
2	5 1 6
3	8
4	1
5	
6	2

Key: 1 | 2 = 12 thousand miles

- (i) Compute the sample mean, median, and mode for the data on distances driven.
- (ii) Compute the standard deviation and the interquartile range for these data.
- (iii) Make a box-and-whisker plot for the data and describe the shape of the distribution.
- (iv) Between the standard deviation and the interquartile range, which is preferable as a measure of the variation of the above data? Why?
- (b) A thief has stolen an automatic teller machine (ATM) card. The card has a four-digit personal identification number (PIN). The thief knows that the first two digits are 2 and 6, but he does not know the last two digits. Thus, the PIN could be any number from 2600 to 2699.
- (i) How many possible numbers can the thief attempt for the last two digits?
- (ii) The automatic teller machine will not allow more than three unsuccessful attempts to enter the PIN. After the third wrong PIN, the machine keeps the card and allows no further attempts. Find the probability that the thief finds the correct PIN within the three attempts? (Assume that the thief will not try the wrong PIN twice.)
- (c) A random sample of 80 lawyers was selected, and they were asked if they are in favour of or against the death penalty. The following table shows the distribution of their responses by age.

		<i>Response</i>		
		Favours (<i>F</i>)	Opposes (<i>O</i>)	Neutral (<i>N</i>)
<i>Age</i> (years)	<i>< 35</i> (<i>Y</i>)	5	15	3
	<i>35 - 45</i> (<i>M</i>)	10	15	5
	<i>> 45</i> (<i>E</i>)	15	10	2

.../3-

- (i) If a lawyer is randomly selected from this sample, what is the probability that he/she opposes *or* is neutral about the death penalty?
- (ii) Two lawyers are selected at random from those in favour the death penalty. Calculate the probability that one is less than 35 years old and the other is more than 35 years old.

Given that 90% of the lawyers in favour of the death penalty are males, as do 70% of those who oppose and 30% of those who are neutral.

- (iii) What is the probability that a lawyer selected at random is a male?
- (iv) Suppose that a lawyer selected at random is a male. What is the probability that he is in favour of the death penalty?

[100 marks]

1. (a) *Gambarajah tangkai dan daun yang berikut menunjukkan jarak (dalam ribuan batu) yang dipandu oleh suatu sampel pemandu di sebuah bandaraya.*

0	3 6 9
1	2 8 5 1 0 5
2	5 1 6
3	8
4	1
5	
6	2

Petunjuk: $1/2 = 12$ ribu batu

- (i) *Hitung min sampel, median sampel dan mod sampel bagi data mengenai jarak yang dipandu.*
- (ii) *Hitung sisihan piawai dan julat antara kuartil bagi data di atas*
- (iii) *Buat suatu plot kotak bagi data di atas dan perihalkan bentuk taburannya.*
- (iv) *Antara sisihan piawai dan julat antara kuartil, sukatan apakah yang lebih dicenderung sebagai sukatan serakan bagi data di atas? Kenapa?*
- (b) *Seorang pencuri telah mencuri sekeping kad mesin pelayan automatik (ATM). Kad tersebut mempunyai nombor pengenalan peribadi (PIN) yang terdiri daripada empat digit. Pencuri tersebut tahu bahawa dua digit pertama ialah 2 dan 6, tetapi ia tidak tahu dua digit yang terakhir. Oleh itu, PIN kad tersebut mungkin sebarang nombor daripada 2600 sehingga 2699.*
- (i) *Berapakah nombor yang mungkin dicuba oleh pencuri tersebut untuk dua digit yang terakhir?*

.../4-

[100 markah]

- (iii) Apakah keberangkalian bahawa seorang peguam yang dipilih secara rawak adalah seorang lelaki?
 (iv) Andaikan bahawa seorang peguam yang dipilih secara rawak adalah seorang lelaki. Apakah keberangkalian bahawa ia bersetuju dengan hukuman mati?

- (i) Jika seorang peguam dipilih secara rawak daripada sampel ini, apakah keberangkalian bahawa ia tidak setuju dengan hukuman mati atau neutral mengenai hukuman mati?
 (ii) Dua orang peguam dipilih secara rawak daripada kalangan peguam yang bersetuju dengan hukuman mati. Hitung keberangkalian bahawa seorang berrumur kurang daripada 35 tahun dan seorang lagi berrumur lebih daripada 35 tahun.

Maklum balas	Umur (tahun)	
	< 35 (Y)	35 - 45 (M)
Setuju (F)	5	10
Tidak Setuju (O)	15	15
Neutral (N)	3	5
	15	10
	2	2

- (c) Suatu sampel rawak 80 orang peguam dipilih dan mereka ditanya sama ada mereka bersetuju atau tidak bersetuju dengan hukuman mati. Jadual yang berikut menunjukkan taburan maklum balas mereka berdasarkan umur.

- (ii) Mesin pelayan automatik tidak membenarkan lebih daripada tiga percubaan yang gagal dalam memasuki PIN. Selepas PIN ketiga yang salah, mesin tersebut akan menyimpan kadnya dan tidak membenarkan sebarang percubaan lagi. Dapatkan keberangkalian bahawa pencuri tersebut mendapat PIN yang betul dalam sekitar tiga percubaan. (Andaikan bahawa ia tidak akan mencuba PIN yang salah dua kali.)

2. (a) The discrete random variable X has a probability function given by

$$P(X = x) = \begin{cases} \left(\frac{1}{2}\right)^x & 1 \leq x \leq 5 \\ C & x = 6 \\ 0 & \text{otherwise} \end{cases}$$

- (i) Determine the value of C , a constant.
 (ii) Find $P(X \leq 5 \mid X > 4)$.
 (iii) Find $E(X)$ and show that $\text{Var}(X) = 1.6553$
 (iv) Approximate the probability that a sample of 100 independent observations of X has a total sum of less than 200.
- (b) A history teacher gives a test consisting of 50 multiple-choice questions. Each question has four choices of which only one is correct. The scoring includes a penalty for guessing; each correct answer is worth 1 point, each wrong answer costs 1/2 point and an unanswered question gets 0 point. For example, if a student answers 35 questions correctly, 8 questions incorrectly, and does not answer 7 questions, the total score for the student will be $35(1) + 8(-1/2) + 7(0) = 31$.
- (i) Suppose a student has no idea what the correct answer to a question is. By guessing the answer, does he increase or decrease his expected score? Explain.

Suppose a student answers 38 questions correctly and guesses on the other 12 questions by randomly choosing one of the four answers for each.

- (ii) What is the probability that the student will score more than 40 points?
 (iii) What is his expected score?
- (c) The number of cars arriving at a petrol station in a period of t minutes may be assumed to have a Poisson distribution with mean $\frac{7t}{20}$.
- (i) What is the probability that fewer than 6 cars will arrive in a 10 minute period?
 (ii) Suppose a car has just arrived at the petrol station. What is the probability that it will be at least 5 minutes before the next car arrives?
 (iii) Approximate the probability that more than 24 cars will arrive in an hour.

[100 marks]

.../6-

2. (a) *Pembolehubah rawak diskrit X mempunyai fungsi kebarangkalian seperti yang berikut:*

$$P(X = x) = \begin{cases} \left(\frac{1}{2}\right)^x & 1 \leq x \leq 5 \\ C & x = 6 \\ 0 & \text{di tempat lain} \end{cases}$$

- (i) *Tentukan nilai C , suatu pemalar.*
 (ii) *Dapatkan $P(X \leq 5 \mid X > 4)$.*
 (iii) *Dapatkan $E(X)$ dan tunjukkan bahawa $\text{Var}(X) = 1.6553$*
 (iv) *Dapatkan kebarangkalian hampiran bahawa suatu sampel rawak 100 cerapan tak bersandar X mempunyai jumlah kurang daripada 200.*
- (b) *Seorang guru sejarah memberi suatu ujian yang terdiri daripada 50 soalan pilihan berganda. Setiap soalan mempunyai empat pilihan jawapan yang mana hanya satu sahaja adalah betul. Pemarkahan ujian termasuk suatu penalti untuk penekaan; setiap jawapan betul bernilai 1 markah, setiap jawapan salah ditolak $\frac{1}{2}$ markah dan soalan yang tidak dijawab mendapat 0 markah. Contohnya, jika seorang pelajar menjawab 35 soalan dengan betul, 8 soalan dijawab salah dan 7 soalan tidak dijawab, maka jumlah markah yang akan diperolehnyalah $35(1) + 8(-1/2) + 7(0) = 31$.*

- (i) *Andaikan seorang pelajar tidak tahu langsung jawapan yang betul bagi suatu soalan. Dengan meneka jawabannya, adakah ia meningkatkan atau menurunkan markah yang dijangka? Jelaskan jawapan anda.*

Andaikan seorang pelajar menjawab 38 soalan dengan betul dan meneka pada 12 soalan yang lain dengan memilih secara rawak satu daripada empat pilihan jawapan bagi setiap satu.

- (ii) *Apakah kebarangkalian bahawa pelajar tersebut akan mendapat lebih daripada 40 markah?*
 (iii) *Berapakah markah yang dijangka diperolehnyalah?*
- (c) *Bilangan kereta yang tiba di sebuah stesen minyak dalam suatu tempoh t minit, dapat diandaikan tertabur secara Poisson dengan min $\frac{7t}{20}$.*
- (i) *Apakah kebarangkalian bahawa kurang daripada 6 buah kereta akan tiba dalam suatu tempoh 10 minit?*

- (ii) *Andaikan sebuah kereta baru sahaja tiba di stesen minyak tersebut. Apakah kebarangkalian bahawa sekurang-kurangnya 5 minit akan berlalu sebelum kereta seterusnya tiba?*
- (iii) *Dapatkan kebarangkalian hampiran bahawa lebih daripada 24 buah kereta akan tiba dalam satu jam.*

[100 markah]

3. (a) A bag contains a very large number of marbles, identical except for their colour. Of these, an unknown proportion p are blue. Suppose you are required to test the null hypothesis $H_0 : p = 0.2$ against the alternative $H_1 : p < 0.2$. To perform the test, you are to take a random sample of 100 marbles and note X , the number of blue marbles in the sample.

- (i) If the significance level is 10%, determine whether the null hypothesis should be accepted when $X = 15$.
- (ii) If the significance level is to be as close as possible to 5%, find the rejection region in the form of $0 \leq X \leq a$, where a is an integer.
- (iii) Calculate the power of the test when the rejection region is $0 \leq X \leq 14$, and $p = 0.1$

- (b) A drug manufacturer produces 250-milligram capsules of a new antibiotic. A random sample is selected, and the amount of antibiotic in each capsule is determined. The results are as follows (in milligrams):

252, 246, 242, 250, 255, 258, 250, 252, 250, 258.

$$\sum x = 2,513 \quad \sum x^2 = 631,741$$

- (i) Find a 95% confidence interval for μ , the mean amount of antibiotic per capsule.
- (ii) Determine the minimum number of capsules that you need to sample if you would like to be 99% confident that an estimate of the mean amount of antibiotic per capsule is within 1 milligram of the true mean.
- (iii) Over a period of months, the manufacturer found that of a sample of 150 capsules, 18 capsules have unacceptable amounts of the antibiotic. Calculate an approximate 95% confidence interval for the proportion of capsules having an acceptable amount of the antibiotic.
- (c) Attitude toward mathematics was measured for two different groups. The attitude scores range from 0 to 80 with the higher scores indicating a more positive attitude. One group consisted of Elementary Education majors, and the other group consisted of majors from several other areas. The data are shown below:

.../8-

Group (major)	n	\bar{x}	s
Elementary Education (1)	75	42.7	15.5
Non-Elementary Education (2)	110	49.3	17.0

- (i) Assuming a common population variance, obtain a pooled estimate of this variance.
- (ii) Construct a 95% confidence interval for $\mu_1 - \mu_2$, the difference between the mean attitude scores of Elementary Education majors and Non-Elementary Education majors.
- (iii) Using the p -value method, test that Non-Elementary Education majors have a more positive attitude toward mathematics than the Elementary Education majors at the 5% significance level.

[100 marks]

3. (a) Sebuah beg mengandungi suatu bilangan besar guli yang serupa tetapi berbeza warna. Daripada bilangan ini, suatu kadaran p yang tidak diketahui adalah guli warna biru. Andaikan anda perlu menguji hipotesis nol $H_0 : p = 0.2$ berlawanan hipotesis alternatif $H_1 : p < 0.2$. Bagi menjalankan ujian, anda perlu mengambil suatu sampel rawak 100 biji guli dan mencatat nilai X , iaitu bilangan guli biru dalam sampel tersebut.

- (i) Jika aras keertian ujian ialah 10%, tentukan sama ada hipotesis nol patut diterima apabila $X = 15$.
- (ii) Jika aras keertian ujian dikehendaki sedekat mungkin dengan 5%, dapatkan kawasan penolakan dalam bentuk $0 \leq X \leq a$, a ialah suatu integer.
- (iii) Hitung kuasa ujian apabila kawasan penolakan ialah $0 \leq X \leq 14$, dan $p = 0.1$

(b) Sebuah syarikat pengeluar ubat menghasilkan sejenis antibiotik baru dalam bentuk kapsul-kapsul 250 milligram. Suatu sampel rawak dipilih dan amaun antibiotik dalam setiap kapsul ditentukan. Hasilnya ialah seperti yang berikut (dalam milligram):

252, 246, 242, 250, 255, 258, 250, 252, 250, 258.

$$\sum x = 2,513 \quad \sum x^2 = 631,741$$

- (i) Dapatkan suatu selang keyakinan 95% bagi μ , iaitu min amaun antibiotik di dalam setiap kapsul yang dihasilkan.

.../9-

- (ii) Tentukan bilangan kapsul yang minimum yang perlu disampelkan jika anda ingin 99% yakin bahawa suatu anggaran bagi min amaun antibiotik di dalam setiap kapsul adalah dalam sekitar 1 milligram daripada min yang sebenar.
- (iii) Dalam suatu tempoh beberapa bulan, pengeluar ubat tersebut mendapati bahawa daripada suatu sampel 150 biji kapsul, 18 biji mempunyai amaun antibiotik yang tidak dapat diterima. Dapatkan suatu selang keyakinan 95% hampiran bagi kadaran kapsul yang mempunyai amaun antibiotik yang dapat diterima.
- (c) Sikap terhadap Matematik diukur bagi dua kumpulan pelajar yang berbeza. Julat bagi skor sikap adalah daripada 0 sehingga 80 dengan skor tinggi menandakan sikap yang lebih positif. Satu kumpulan terdiri daripada pelajar dalam pengkhususan Pendidikan Asas dan kumpulan yang satu lagi terdiri daripada pelajar dalam pengkhususan beberapa disiplin lain. Data yang diperoleh ialah seperti yang berikut:

Kumpulan (pengkhususan)	n	\bar{x}	s
Pendidikan Asas (1)	75	42.7	15.5
Bukan Pendidikan Asas (2)	110	49.3	17.0

- (i) Dengan andaian bahawa varians kedua-dua populasi adalah seponya, dapatkan anggaran tergembeleng bagi varians tersebut.
- (ii) Bina suatu selang keyakinan 95% bagi $\mu_1 - \mu_2$, iaitu perbezaan antara min bagi skor sikap pelajar pengkhususan Pendidikan Asas dan min bagi skor pelajar bukan pengkhususan Pendidikan Asas.
- (iii) Dengan menggunakan kaedah nilai- p , uji bahawa pelajar bukan pengkhususan Pendidikan Asas mempunyai sikap yang lebih positif terhadap Matematik daripada pelajar pengkhususan Pendidikan Asas pada aras keertian 5%.

[100 markah]

4. (a) (i) Explain why nonempty, mutually exclusive events A and B must be dependent.
- (ii) Show that if A and B are independent events, then A and \bar{B} are also independent.
- (b) In a 2003 survey conducted in a city, respondents answered the question "What is the combined income of your household?" The responses appear in the table below

Household Income	% of Households
Under RM20,000	25
RM20,000 to less than RM40,000	43
RM40,000 to less than RM60,000	22
RM60,000 and over	10

Assume that the above results hold true for the entire population in the city. In a recent poll of 200 households in the same city, 39 had a total income of under RM20,000, 83 had an income between RM20,000 and RM40,000, 45 had an income between RM40,000 to less than RM60,000 and the rest had an income over RM60,000. Test at the 5% level of significance whether the current percentages of the city's households in each income category are the same now as they were in 2003.

- (c) The following table lists the ages (in years) and the prices (in thousands of RM) for a sample of cars of the same model.

Age (x years)	8	11	4	2	6	5	13
Price (RM y 000)	40	30	95	130	80	90	15

$$\sum x = 49 \quad \sum y = 480 \quad \sum x^2 = 435 \quad \sum y^2 = 43,150 \quad \sum xy = 2,415$$

- Find the least square regression line.
- Explain the meaning of the y -intercept and the slope of the regression line.
- Calculate r and r^2 and explain their meaning.
- Test at $\alpha = .01$ whether ρ is negative.

[100 marks]

- 4 (a) (i) Terangkan kenapa peristiwa A dan peristiwa B yang tak kosong dan saling eksklusif semestinya saling bersandaran.
- (ii) Tunjukkan bahawa jika A dan B adalah peristiwa-peristiwa saling bersandaran, maka A dan \bar{B} juga adalah saling bersandaran.
- (b) Dalam suatu tinjauan pada tahun 2003 yang dijalankan di sebuah bandaraya, responden menjawab soalan "Berapakah jumlah pendapatan keluarga anda?". Jawapan yang diberi ditunjukkan dalam jadual yang berikut:

Pendapatan keluarga	% keluarga
Kurang daripada RM20,000	25
RM20,000 hingga kurang daripada RM40,000	43
RM40,000 hingga kurang daripada RM60,000	22
RM60,000 dan lebih	10

.../11-

Andaikan hasil di atas adalah benar bagi seluruh populasi di bandaraya tersebut. Dalam suatu tinjauan terkini terhadap 200 buah keluarga di bandaraya yang sama, 39 buah keluarga mempunyai jumlah pendapatan kurang daripada RM20,000, 83 buah keluarga mempunyai jumlah pendapatan antara RM20,000 dan RM40,000, 45 buah keluarga mempunyai jumlah pendapatan antara RM40,000 dan RM60,000 dan yang selebihnya mempunyai jumlah pendapatan lebih daripada RM60,000. Uji pada aras keertian 5% sama ada peratusan semasa keluarga di bandaraya tersebut dalam setiap kategori pendapatan adalah sama seperti dalam tahun 2003.

- (c) Jadual yang berikut menyenaraikan usia (dalam tahun) dan harga (dalam ribuan RM) bagi suatu sampel kereta yang sama model.

Usia (x tahun)	8	11	4	2	6	5	13
Harga (RM y 000)	40	30	95	130	80	90	15

$$\sum x = 49 \quad \sum y = 480 \quad \sum x^2 = 435 \quad \sum y^2 = 43,150 \quad \sum xy = 2,415$$

- (i) Dapatkan garis regresi kuasa dua terkecil.
- (ii) Terangkan makna pintasan- y dan kecerunan garis regresi yang diperoleh.
- (iii) Hitung nilai r dan nilai r^2 dan terangkan makna nilai masing-masing.
- (iv) Uji pada $\alpha = .01$ sama ada ρ adalah negatif.

[100 markah]

FORMULA

$\bar{x} = \frac{\sum x}{n}$ $s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}$	$S_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$ $\bar{p} = \frac{X + Y}{n_x + n_y}$
--	---

Confidence Interval

$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ $\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$ $\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$(\bar{X} - \bar{Y}) \pm z_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$ $(\bar{X} - \bar{Y}) \pm t_{\alpha/2} \sqrt{S_p^2 \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}$ $(\hat{p}_x - \hat{p}_y) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}$
---	--

Test Statistics

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

Regression and Correlation

$$S_{XY} = \sum xy - \frac{(\sum x)(\sum y)}{n}$$

$$s_e = \sqrt{\frac{S_{YY} - bS_{XY}}{n-2}} \quad ; \quad s_b = \frac{s_e}{\sqrt{S_{XX}}} \quad ; \quad r = \frac{S_{XY}}{\sqrt{S_{XX} \cdot S_{YY}}}$$

- 000 O 000 -