UNIVERSITI SAINS MALAYSIA

Peperiksaan Kursus Semasa Cuti Panjang
Sidang Akademik 2005/2006

Jun 2006

MAT 161E – Elementary Statistics
[Statistik Permulaan]

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 four** [4] questions.

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

1. (a) The speeds of 55 randomly selected cars that passed through a road in a town were monitored by radar and recorded as in the following table:

Speed of car, X (miles/hour)	Frequency, f_i
12 - 18	1
18 - 24	14
24 - 30	22
30 - 36	8
36 - 42	5
42 - 48	3
48 - 54	2

(i) Calculate the mean speed and its standard deviation.

(ii) Cars that exceed a specified speed limit will be remanded by the police. If the speed limit is fixed at 40 miles/hour, what percentage of the cars will be remanded by the police?

(iii) Use Chebyshev's Theorem to obtain an interval for the speeds of at least 70% of cars that passed through the road.

[Give yours answers to two decimal places.]

(b) A box contains six white marbles, three red marbles and a blue marble. Three marbles were randomly taken out of the box, without replacement. Let X represents the number of red marbles that were taken out.

(i) Write the probability distribution of X.

(ii) Find the probability that one white, one red and one blue marble were selected.

(iii) Find the probability that at least one red marble was taken out.

(c) 500 adults are classified based on their smoking habits and frequency of visits to the doctor in a year. The results are shown below:

	Frequency of visit to doctor	
	More than 5 times	At most 5 times
Heavy smoker	120	60
Light or nonsmoker	75	245

An individual is randomly selected from the group of 500 people.

(i) Given that the individual is a heavy smoker, what is the probability that he/she visits the doctor at most 5 times a year?

(ii) What is the probability that the selected individual is a heavy smoker or visits the doctor more than 5 times a year?

[100 marks]

1. (a) *Kelajuan 55 buah kereta yang dipilih secara rawak yang melalui sebatang jalan dalam sebuah bandar dipantau dengan alat radar dan direkodkan seperti dalam jadual yang berikut:*

Kelajuan kereta, X (km/jam)	*Kekerapan,* f_i
12 - 18	*1*
18 - 24	*14*
24 - 30	*22*
30 - 36	*8*
36 - 42	*5*
42 - 48	*3*
48 - 54	*2*

(i) *Hitung purata kelajuan kereta yang melalui jalan tersebut dan sisihan piawainya.*

(ii) *Kereta yang melebihi had kelajuan yang ditetapkan akan ditahan oleh polis. Jika had kelajuan ditetapkan pada 40 km/jam, berapakah peratusan kereta yang ditahan oleh polis?*

(iii) *Gunakan Teorem Chebyshev untuk mendapatkan suatu selang kelajuan kereta bagi sekurang-kurangnya 70% daripada kereta-kereta yang melalui jalan tersebut*

[Beri jawapan anda dalam dua tempat perpuluhan]

(b) *Sebuah kotak mengandugi enam biji guli putih, tiga biji guli merah dan sebiji guli biru. Tiga biji guli dikeluarkan secara rawak daripada kotak tersebut, tanpa pengembalian. Andaikan X mewakili bilangan guli merah yang dikeluarkan.*

(i) *Tuliskan taburan kebarangkalian bagi X.*

(ii) *Dapatkan kebarangkalian bahawa sebiji guli putih, sebiji guli merah dan sebiji guli biru terpilih.*

(iii) *Dapatkan kebarangkalian bahawa sekurang-kurangnya sebiji guli merah dikeluarkan.*

(c) 500 orang dewasa dikelaskan berdasarkan kebiasaan merokok dan kekerapan mereka berjumpa dengan doktor dalam setahun. Hasilnya ditunjukkan dalam jadual yang berikut.

	Kekerapan jumpa doktor	
	Lebih daripada 5 kali	*Sebanyak-banyaknya 5 kali*
Kuat merokok	*120*	*60*
Perokok ringan atau bukan perokok	*75*	*245*

Seorang dewasa dipilih secara rawak daripada kumpulan 500 orang ini.

(i) Jika diketahui bahawa orang tersebut kuat merokok, apakah kebarangkalian ia berjumpa dengan doktor sebanyak-banyaknya 5 kali dalam setahun?

(ii) Apakah kebarangkalian bahawa orang yang dipilih tersebut kuat merokok atau berjumpa dengan doktor lebih daripada 5 kali dalam setahun?

[100 markah]

2. (a) The probability density function of a continuous variable X is given by

$$f(x)=\begin{cases} \dfrac{(x+1)}{6}, & 1<x<3 \\ 0, & \textit{otherwise,} \end{cases}$$

(i) Find E(X) and Var(X).

(ii) The median for a continuous probability distribution is a value m that divides the distribution into two equal areas, i.e.

$$\int_{-\infty}^{m} f(x)\,dx = \int_{m}^{\infty} f(x)\,dx = 0.5.$$

Find the median of $f(x)$.

(b) A history teacher gives a test consisting of 20 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 2 points, each wrong answer costs 1 point and an unanswered question gets 0 point. For example, if a student answers 10 questions correctly and 5 questions incorrectly, and does not answer 5 questions, the total score for the student will be 10(2) + 5(-1) + 5(0) = 15.

(i) Suppose a student has no idea what the correct answer to a question is. By guessing the answer, what is his expected score?

Suppose a student answers 10 questions correctly and guesses on the other 10 questions by randomly choosing one of the four answers for each.

(ii) What is the probability that the student will score more than 30 points?
(iii) What is his expected score?

(c) A machine makes metal rods. The diameters of the rods are normally distributed with mean 1.00cm and standard deviation 0.02cm. A rod is considered oversized if its diameter exceeds a specified value. After an inspection, it is found that that 1% of the rods produced by the machine are oversized.

(i) Calculate the minimum diameter of an oversized rod.

Suppose 200 rods are chosen at random. Find the probability that

(ii) at least four of the rods are oversized. (Use a suitable approximation method)
(iii) the sum of the diameters of the rods in the sample is at most 205 cm.

[100 marks]

2. (a) *Taburan kebarangkalian pembolehubah selanjar X diberikan oleh*

$$f(x)=\begin{cases}\dfrac{(x+1)}{6}, & 1<x<3\\ 0, & di\,tempat\ lain,\end{cases}$$

(i) *Dapatkan E(X) dan Var(X).*
(ii) *Median bagi taburan kebarangkalian selanjar ialah suatu nilai m yang membahagikan taburan tersebut kepada dua bahagian yang sama besar iaitu.:*

$$\int_{-\infty}^{m} f(x)\,dx=\int_{m}^{\infty} f(x)\,dx=0.5\,.$$

Dapatkan median f(x).

(b) *Seorang guru sejarah memberi suatu ujian yang terdiri daripada 20 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 2 markah, setiap jawapan salah ditolak 1* *markah dan soalan yang tidak dijawab mendapat 0 markah. Contohnya, jika seorang pelajar menjawab 10 soalan dengan betul, 5 soalan dijawab salah dan 5 soalan tidak dijawab, maka jumlah markah yang akan diperolehnya ialah 10(2) + 5(-1) + 5(0) = 15.*

(i) *Andaikan seorang pelajar tidak tahu langsung jawapan bagi suatu soalan. Dengan meneka jawapannya, berapakah markah yang dijangka diperolehnya?*

Andaikan seorang pelajar menjawab 10 soalan dengan betul dan meneka jawapan bagi setiap 10 soalan yang lain dengan memilih secara rawak satu daripada empat pilihan jawapan.

(ii) *Apakah kebarangkalian bahawa pelajar tersebut akan mendapat lebih daripada 30 markah?*
(iii) *Berapakah markah yang dijangka diperolehnya?*

(c) *Sebuah mesin membuat batang-batang besi. Diameter batang-batang besi tersebut tertabur secara normal dengan min 1.00cm dan sisihan piawai 0.02 cm. Sebatang besi dianggap terlalu besar jika diameternya melebihi nilai yang ditetapkan. Selepas suatu pemeriksaan, didapati bahawa 1% daripada batang-batang besi yang dihasilkan oleh mesin tersebut adalah terlalu besar,*

(i) *Hitung diameter minimum bagi sebatang besi yang terlalu besar*

Andaikan 200 batang besi dipilih secara rawak. Dapatkan kebarangkalian bahawa

(ii) *sekurang-kurangnya empat batang besi adalah terlalu besar. (Guna suatu kaedah penghampiran yang sesuai*
(iii) *jumlah diameter batang besi dalam sampel tersebut adalah sebanyak-banyaknya 205 cm.*

[100 markah]

3. (a) It is known that the variable $X \sim N(\mu,16)$. It is desired that the null hypothesis $\mu = 12$ be tested against the alternative hypothesis $\mu > 12$ with probability of a Type I error of 1%. A random sample of 15 observations of X is taken and the sample mean $\overline{X}$ is taken to be the test statistic.

(i) Find the acceptance and rejection regions in terms of $\overline{X}$.
(ii) For the case of $\mu = 15$, find the probability of a type II error and the power of the test.

(b) The label on a sack of rice sold at a sundry shop qoutes that the weight of its content is 25 kg. To estimate the true mean weight of a 25-kg sack of rice, a random sample taken from a shop yields the following data (in X kg):

21.24, 24.81, 23.62, 26.82, 25.50, 24.88, 25.08

$$\sum x = 171.95, \quad \sum x^2 = 4242.16$$

(i) Find a 95% confidence interval for the true mean weight of a 25-kg sack of rice.

(ii) Over a period of time, it was found that in a random sample of 150 sacks of rice, 18 have weights less than 25 kg. Calculate a 95% confidence interval for the proportion of rice sacks with weights less than 25 kg.

(c) A survey was conducted to compare the mean cost of a meal at fast food restaurants in two different cities. The data obtained is summarized in the following table:

City	*n*	$\overline{x}$	*s*
A	*40*	*RM4.05*	*RM0.55*
B	*35*	*RM4.85*	*RM0.85*

(i) Assuming a common population variance, obtain a pooled estimate of this common variance.

(ii) Construct a 95% confidence interval for $\mu_A - \mu_B$, the difference between the mean cost of a meal at fast food restaurants in cities A and B.

(iii) Use the p-value method to test that mean cost of a meal at fast food restaurants in city A is less than mean cost of a meal at fast food restaurants in city B. Test at the 5% significance level.

[100 marks]

3. *(a) Diketahui bahawa pembolehubah* $X \sim N(\mu, 16)$. *Hipotesis nol* $\mu = 12$ *dikehendaki diuji berlawanan dengan hipotesis alternatif* $\mu > 12$ *dengan kebarangkalian ralat Jenis I sebanyak 1%. Suatu sampel rawak 15 cerapan X diambil dan min sampel* $\overline{X}$ *digunakan sebagai statistik ujian.*

(i) Dapatkan kawasan penerimaan dan kawasan penolakan dalam sebutan $\overline{X}$.

(ii) Bagi kes $\mu = 15$, *dapatkan kebarangkalian ralat Jenis II dan kuasa ujian.*

(b) Label pada sekarung beras yang dijual di sebuah kedai runcit menyatakan bahawa berat kandungannya ialah 25 kg. Untuk menganggar berat sebenar sekarung beras 25-kg, suatu sampel rawak yang diambil di kedai tersebut menghasilkan data yang berikut (dalam X kg):

21.24, 24.81, 23.62, 26.82, 25.50, 24.88, 25.08

$\sum x = 171.95, \quad \sum x^2 = 4242.16$

(i) Dapatkan suatu selang keyakinan 95% bagi min berat sebenar sekarung beras 25-kg.

(ii) Selepas suatu jangka masa, didapati bahawa dalam suatu sampel rawak 150 karung beras, 18 karung mempunyai berat kurang daripada 25 kg. Hitung suatu selang keyakinan 95% bagi kadaran karung beras yang mempunyai berat kurang daripada 25 kg.

(c) Suatu tinjauan dijalankan untuk membandingkan harga purata suatu hidangan di restoran-restoran makanan segera di dua buah bandaraya berbeza. Data yang diperoleh diringkaskan dalam jadual yang berikut:

Bandaraya	n	$\bar{x}$	s
A	*40*	*RM4.05*	*RM0.55*
B	*35*	*RM4.85*	*RM0.85*

(i) Dengan andaian varians sepunya, dapatkan anggaran tergembeleng bagi varians sepunya tersebut.

(ii) Bina suatu selang keyakinan bagi $\mu_A - \mu_B$, iaitu perbezaan antara harga purata suatu hidangan di restoran-restoran makanan segera di bandaraya A dan di bandaraya B.

(iii) Gunakan kaedah nilai-p untuk menguji bahawa harga purata suatu hidangan di restoran-restoran makanan segera di bandaraya A adalah lebih rendah daripada di bandaraya B. Uji pada aras keyakinan 5%.

[100 markah]

4. (a) A survey carried out in 1995 reported that 25% of residents in a certain region opposed the idea of a construction project in their area. To test that the percentage opposing is now less than 25%, a marketing research firm randomly selects 50 residents in the region and determines that 12 are opposed to the idea of the construction.

(i) At the 5% significance level, test whether the sample is sufficient evidence to reject the 1995 report

(ii) If the firm wants to be 90% confident that the sample proportion will be within 0.05 of the true proportion, what sample size should be used?

(b) Suppose that a null hypothesis is that a discrete random variable X has probability distribution given by:

$$P(X = x) = \begin{cases} x/15 & x = 1, 2, 3, 4, 5 \\ 0 & otherwise \end{cases}$$

A random sample of 105 observations of X is summarized in the table below:

Value of X	1	2	3	4	5
Frequency	5	19	18	28	35

Is there sufficient evidence to reject the null hypothesis? Test at the 1% significance level,

(c) Students following a certain course were given a reading competency test (scores range from 0 to 48) and also a math competency test (scores range from 50 to 100). The scores are given in the following table:

Reading (X)	40	36	42	29	44	35	38	42	45	40
Math (Y)	78	80	90	60	95	70	77	83	90	80

$\sum x = 391,\quad \sum x^2 = 15{,}495,\quad \sum y = 803,\quad \sum y^2 = 65{,}427,\quad \sum xy = 31{,}812$

(i) Find an estimate of the equation of the line of best fit for the data.
(ii) Give the interpretation of your estimates of the slope and intercept.
(iii) Calculate the coefficient of linear correlation and coefficient of determination. Interpret the values obtained in the context of the problem.
(iv) Suppose that for each value of X, the measured Y value has a random error which is normally distributed with zero mean and variance 9. What is the probability that, when $X = 30$, the measured value of T exceeds 65?

[100 marks]

4. (a) *Suatu tinjauan yang dijalankan pada tahun 1995 melaporkan bahawa 25% penduduk di sebuah kawasan menentang cadangan suatu projek pembinaan di kawasan mereka. Bagi menguji bahawa peratusan yang menentang sekarang adalah kurang daripada 25%, sebuah firma penyelidikan pemasaran mengambil suatu sampel rawak 50 orang penduduk dan mendapati bahawa 12 orang menentang cadangan pembinaan tersebut.*

(i) *Pada aras keyakinan 5%, uji sama ada sampel yang diperoleh adalah bukti cukup untuk menolak laporan 1995.*

(ii) *Jika firma penyelidikan pemasaran tersebut ingin 90% pasti bahawa kadaran sampelnya adalah dalam sekitar 0.05 daripada kadaran sebenar, berapah saiz sampel yang patut diguna?*

(b) *Andaikan suatu hipotesis nol ialah bahawa pembolehubah rawak diskrit X mempunyai taburan kebarangkalian seperti berikut:*

$$P(X = x) = \begin{cases} \frac{x}{15} & x = 1, 2, 3, 4, 5 \\ 0 & \text{di tempat lain} \end{cases}$$

Suatu sampel rawak 105 cerapan X diringkaskan dalam jadual yang berikut:

Value of X	1	2	3	4	5
Frequency	5	19	18	28	35

Adakah sampel ini bukti cukup untuk menolak hipotesis nol tersebut? Uji pada aras keyakinan 5%.

(c) *Pelajar-pelajar yang mengikuti suatu kursus diberi suatu ujian kecekapan membaca (julat skor daripada 0 hingga 48) dan juga suatu ujian kecekapan matematik (julat skor daripada 50 hingga 100). Skor-skornya diberikan dalam jadual yang berikut:*

Membaca (X)	40	36	42	29	44	35	38	42	45	40
Matematik (Y)	78	80	90	60	95	70	77	83	90	80

$$\sum x = 391, \quad \sum x^2 = 15{,}495, \quad \sum y = 803, \quad \sum y^2 = 65{,}427, \quad \sum xy = 31{,}812$$

(i) *Dapatkan persamaan garis penyuaian terbaik bagi data di atas.*
(ii) *Berikan tafsiran bagi anggaran kecerunan dan pintasan garis yang diperoleh*
(iii) *Hitung pekali korelasi linear dan pekali penentuan. Tafsirkan nilai yang diperoleh dalam konteks masalah yang diberikan.*
(iv) *Andaikan bahawa bagi setiap nilai X, nilai Y yang disukat mempunyai ralat rawak yang tertabur secara normal dengan min sifar dan varians 9. Apakah kebarangkalian bahawa apabila X = 30, nilai Y yang disukat melebihi 65?*

[100 markah]

FORMULA

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

Confidence Interval

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

Test Statistics

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

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

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

- 000 O 000 -