UNIVERSITI SAINS MALAYSIA

Second Semester Examination
Academic Session 2006/2007

April 2007

MAT 161 – Elementary Statistics
[Statistik Permulaan]

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 **all four** [4] questions.

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

.../2-

1. (a) The following table shows the frequency distribution of the number of hours each of the 200 students randomly selected from a college studies per week.

Hours studied, x	Frequency, f
0 – 3	12
4 – 7	29
8 – 11	48
12 – 15	62
16 – 19	34
20 – 23	15

$\sum xf = 2388, \quad \sum x^2 f = 33882$

(i) Calculate estimates for the mean and standard deviation of the number of hours a college student studies per week.

(ii) Using Chebyshev's theorem, find the minimum number of students who studies between 5.88 to 18.00 hours per week.

(iii) What percentage of the students study for more than 18 hours per week?

(iv) Calculate the value of the 70^{th} percentile. Give a brief interpretation of this percentile in the context of the given problem.

(Give your answers in two decimal places.)

(b) A student is to answer seven out of ten questions in an examination. How many choices does he have if

(i) he must answer the first three questions?

(ii) he must answer at least three of the first five questions?

(c) Three cards are identical in form but coloured differently. The first card has both sides coloured red, the second card has both sides coloured black and the third card has one side coloured red and the other coloured black. The three cards are mixed up in a box and one card is randomly selected and placed on a table. If the upper side of the selected card is coloured red, what is the probability that its other side is coloured black?

(d) Suppose that A, B and C are three mutually independent events. Show that

$$P(\overline{A} \cap \overline{B} \cap \overline{C}) = P(\overline{A})P(\overline{B})P(\overline{C}).$$

[100 marks]

1. (a) *Jadual yang berikut menunjukkan taburan kekerapan bagi bilangan jam yang digunakan setiap 200 pelajar yang dipilih secara rawak daripada sebuah kolej untuk belajar setiap minggu.*

Bilangan jam belajar, x	*Kekerapan, f*
0 – 3	*12*
4 – 7	*29*
8 – 11	*48*
12 – 15	*62*
16 – 19	*34*
20 – 23	*15*

$\sum xf = 2388, \quad \sum x^2 f = 33882$

(i) *Hitung anggaran min dan sisihan piawai bagi bilangan jam belajar setiap minggu bagi seorang pelajar kolej.*

(ii) *Dengan menggunakan teorem Chebyshev, dapatkan bilangan pelajar yang minimum yang belajar antara 5.88 hingga 18.00 jam setiap minggu.*

(iii) *Berapakah peratusan pelajar yang belajar lebih daripada 18 jam setiap minggu?*

(iv) *Hitung nilai persentil ke-70. Berikan tafsiran persentil ini dalam konteks masalah yang diberikan.*

(Berikan jawapan anda dalam dua tempat perpuluhan.)

(b) *Seorang pelajar dikehendaki menjawab tujuh daripada sepuluh soalan dalam suatu peperiksaan. Berapakah pilihan yang ia ada jika*

(i) *ia mesti menjawab tiga soalan yang pertama?*

(ii) *ia mesti menjawab sekurang-kurangnya tiga daripada lima soalan yang pertama?*

(c) *Tiga keping kad mempunyai bentuk yang serupa tetapi diwarnai berbeza-beza. Kad pertama mempunyai kedua-dua sisi berwarna merah, kad kedua mempunyai kedua-dua sisi berwarna hitam dan kad ketiga mempunyai satu sisi hitam dan satu sisi lagi merah. Ketiga-tiga kad tersebut dicampur-adukkan di dalam sebuah kotak dan sekeping kad dikeluarkan dan diletakkan di atas meja. Jika sisi atas kad yang terpilih tersebut berwarna merah, apakah kebarangkalian bahawa sisinya yang satu lagi berwarna hitam?*

(d) *Andaikan A, B dan C ialah tiga peristiwa yang saling tak bersandar. Tunjukkan bahawa*

$$P(\overline{A} \cap \overline{B} \cap \overline{C}) = P(\overline{A})P(\overline{B})P(\overline{C}).$$

[100 markah]

.../4-

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

$$f(x)=\begin{cases} x & 0\le x\le 1 \\ k-x & 1\le x\le 2 \\ 0 & otherwise \end{cases}$$

(i) Determine the value of k.

(ii) Calculate $P\left(X>1 \middle| X<\frac{3}{2}\right)$

(iii) Find the value of Q_1.

(b) Two companies, A and B, drill wells in a rural area. Company A charges a flat fee of RM3500 to drill a well regardless of its depth. Company B charges RM1000 plus RM40 per meter to drill a well. The depths of wells in this area have a normal distribution with a mean of 76 meters and a standard deviation of 12 meters.

(i) What is the probability that Company B would charge more than Company A to drill a well?
(ii) Find the mean amount charged by Company B to drill a well.
(iii) Which company would you choose to drill a well? Give your reason(s).
(iv) If six wells are drilled in the area, what is the probability that the mean depth of these wells will be between 70 and 84 meters?

(c) Serious delays on a certain railway line occur at random, at an average of one per week.

(i) Find the probability that at least four serious delays occur in a particular month. Assume there are four weeks in a month.
(ii) Assume that a year consists of thirteen 4-week periods. Find the probability that in a particular year, more than ten of these 4-week periods would have at least four serious delays.
(iii) Approximate the probability that at least 50 serious delays will occur in a year.

[100 marks]

2. (a) *Fungsi taburan kebarangkalian suatu pembolehubah rawak selanjar X diberikan oleh:*

$$f(x)=\begin{cases} x & 0\le x\le 1 \\ k-x & 1\le x\le 2 \\ 0 & \text{di tempat lain} \end{cases}$$

(i) *Tentukan nilai k.*

(ii) *Hitung* $P\left(X>1 \,\middle|\, X<\frac{3}{2}\right)$

(iii) *Dapatkan nilai* Q_1.

(b) *Dua buah syarikat, A dan B, menggali perigi di sebuah kawasan luar bandar. Syarikat A mengenakan bayaran sama rata sebanyak RM3500 tanpa mengambil kira kedalaman perigi yang digali. Syarikat B mengenakan bayaran RM1000 campur RM40 bagi setiap meter kedalaman perigi yang digali. Kedalaman perigi di kawasan tersebut mempunyai taburan normal dengan min 76 meter dan sisihan piawai 12 meter.*

(i) *Berapakah kebarangkalian bahawa Syarikat B akan mengenakan bayaran yang lebih tinggi daripada Syarikat A untuk menggali sebuah perigi?*

(ii) *Dapatkan min amaun yang dikenakan oleh Syarikat B untuk menggali sebuah perigia.*

(iii) *Syarikat manakah yang anda akan pilih untuk menggali sebuah perigi? Berikan alasan anda.*

(iv) *Jika enam buah perigi digali di kawasan tersebut, berapakah kebarangkalian bahawa min kedalaman perigi-perigi tersebut adalah di antara 70 meter dan 84 meter?*

(c) *Kelewatan serius pada sebuah laluan keretapi berlaku secara rawak pada purata satu setiap minggu.*

(i) *Dapatkan kebarangkalian bahawa sekurang-kurangnya empat kelewatan serius berlaku dalam suatu bulan tertentu. Andaikan terdapat empat minggu dalam satu bulan.*

(ii) *Andaikan bahawa satu tahun mengandungi tiga belas tempoh 4-minggu. Dapatkan kebarangkalian bahawa dalam satu tahun tertentu, lebih daripada sepuluh tempoh-tempoh 4-minggu tersebut akan mengandungi sekurang-kurangnya empat kelewatan serius.*

(iii) *Anggarkan kebarangkalian bahawa sekurang-kurangnya 50 kelewatan serius akan berlaku dalam satu tahun.*

[100 markah]

.../6-

3. (a) A machine is programmed to produce 3-cm nails. It is known that the standard deviation of lengths of nails produced by the machine is 0.04 cm. A sample of 10 nails is obtained and the lengths measured. The results (in x centimeters) are as follows:

2.95, 2.98, 3.01, 3.05, 2.95, 2.98, 3.05, 3.06, 3.00, 3.12.

$\sum x = 30.15 \quad \sum x^2 = 90.9289$

(i) Find a 95% confidence interval for μ, the mean length of a nail produced by the machine.

(ii) Determine the minimum sample size if you would like to be 99% confident that your estimate of μ is within 0.001 inch of its true value.

(iii) Over a period of several months, it is found that in a sample of 1500 nails, 105 are unacceptable because they are either too short or too long. Calculate a 95% confidence interval for the proportion of nails that are acceptable.

(b) Air Alpha claims that only 15% of its flight arrive more than 10 minutes late. Let p be the proportion of all Air Alpha's flights that arrive more than 10 minutes late. Consider the hypothesis test

$H_0 : p \leq 0.15$ versus $H_1 : p > 0.15$.

Suppose a random sample of 50 flights by Air Alpha is taken and the significance level is set at 10%.

(i) Determine whether the null hypothesis should be accepted when X, the number of Air Alpha's flights that arrive more than 10 minutes late in the sample, is 12.

(ii) Calculate the Type II error and power of the test when $p = 0.18$.

(c) Attitude toward mathematics was measured for two different groups of students. The attitude scores range from 0 to 80 with the higher scores indicating a more positive attitude. Results obtained from a sample of Elementary Education majors and another sample of majors from several other areas are shown below:

Group (major)	***n***	**mean**	***Standard deviation***
Elementary Education (X)	75	42.7	15.5
Non-Elementary Education (Y)	110	49.3	17.0

(i) At $\alpha = 1\%$, test whether there is a difference in the variance of the attitude scores of the two groups of students.

(ii) Construct a 95% confidence interval for $\mu_X - \mu_Y$.

(iii) Using the p-value method, test at the 5% significance level, that Non-Elementary Education majors have a more positive attitude toward mathematics than the Elementary Education majors.

[100 marks]

.../7-

3. (a) *Sebuah mesin diprogram untuk menghasilkan paku-paku 3-cm. Diketahui bahawa sisihan piawai panjang paku-paku yang dihasilkan oleh mesin tersebut ialah 0.04cm. Suatu sampel 10 batang paku diambil dan panjangnya diukur. Hasilnya (dalam x centimeter) ialah seperti yang berikut:*

2.95, 2.98, 3.01, 3.05, 2.95, 2.98, 3.05, 3.06, 3.00, 3.12.

$\sum x = 30.15 \quad \sum x^2 = 90.9289$

(i) *Dapatkan suatu selang keyakinan 95% bagi μ, min panjang sebatang paku yang dihasilkan oleh mesin tersebut.*

(ii) *Tentukan saiz sampel yang minimum jika anda ingin 99% yakin bahawa anggaran anda bagi μ adalah dalam sekitar 0.001 inci daripada nilainya yang sebenar.*

(iii) *Dalam suatu tempoh beberapa bulan, didapati bahawa dalam suatu sampel 1500 batang paku, 105 batang tidak dapat diterima kerana terlalu pendek atau terlalu panjang. Hitung suatu selang keyalinan 95% bagi kadaran paku yang dapat diterima.*

(b) *Air Alpha mendakwa hanya 15% daripada penerbangannya tiba lewat lebih daripada 10 minit. Andaikan p ialah kadaran semua penerbangan Air Alpha yang tiba lewat lebih daripada 10 minit. Pertimbangkan ujian hipotesis:*

$H_0 : p \leq 0.15$ *berlawanan* $H_1 : p > 0.15$.

Andaikan bahawa suatu sampel rawak 50 penerbangan Air Alpha diambil dan aras keertian ujian ditetapkan pada 10%.

(i) *Ttentukan sama ada hipotesis nol patut diterima jika X, bilangan penerbangan Air Alpha yang tiba lewat lebih daripada 10 minit dalam sampel, ialah 12.*

(ii) *Hitung ralat jenis II dan kuasa ujian apabila $p = 0.18$.*

(c) *Sikap terhadap matematik diukur untuk dua kumpulan pelajar yang berbeza. Skor sikap ini adalah daripada 0 hingga 80 dengan skor tinggi menandakan sikap yang lebih positif. Hasil yang diperoleh daripada suatu sampel pelajar berpengkhususan Pendidikan Asas dan satu lagi daripada sampel pelajar beberapa pengkhususan lain ditunjukkan di bawah:*

Kumpulan (pengkhususan)	***n***	***min***	***sisihan piawai***
Pendidikan Asas (X)	*75*	*42.7*	*15.5*
Bukan Pendidikan Asas (Y)	*110*	*49.3*	*17.0*

(i) *Pada $\alpha = 1\%$, uji sama ada terdapat perbezaan dalam varians skor sikap kedua-dua kumpulam pelajar.*

(ii) *Bina suatu selang keyakinan 95% bagi $\mu_X - \mu_Y$.*

(iii) *Menggunakan kaedah nilai-p, uji sama ada pelajar berpengkhususan bukan Pendidikan Asas mempunyai sikap yang lebih positif terhadap matematik daripada pelajar yang berpengkhususan Pendidikan Asas.*

[100 markah]

.../8-

4. (a) Many water treatment plants supplement the natural fluoride concentration in order to reach a target concentration of fluoride in drinking water. Certain levels are thought to enhance dental health, but very high concentrations can be dangerous. Suppose that a treatment plant targets 0.75 mg/L for their water. The plant tests 25 samples each day to determine whether the median level differs from the target.

(i) Suppose that one day's samples result in 18 values that exceed 0.75 mg/L. Conduct a hypothesis test with $\alpha = 0.10$.

(ii) Explain the implication of a Type I error in the context of this problem.

(iii) When it was suggested to the plant's supervisor that a t-test should be used to conduct the daily test, she replied that the probability distribution of the fluoride concentration was "heavily skewed to the right". Show graphically what she meant by this and explain why her reply is a reason to prefer the sign test to the t-test.

(b) A pharmaceutical company has synthesized two new compounds to be used in two different sleeping drugs. The data in the following table shows the additional hours of sleep gained by 10 patients through the use of the two drugs. Do the data present sufficient evidence to indicate that the probability distributions of additional hours of sleep differ for the two drugs? Test using the Wilcoxon signed-rank test at the 5% significance level.

Patient	A	B	C	D	E	F	G	H	I	J
Drug A	.4	−.7	−.4	−1.4	−1.6	2.9	4.0	.1	3.1	1.9
Drug B	.7	−1.6	−.2	−1.4	−.2	3.4	3.7	.8	.0	2.0

(c) A social science researcher hypothesized that the monthly salaries of workers in a certain sector is normally distributed with mean RM1200 and standard deviation RM200. A random sample of 1000 workers from the sector was obtained and the following incomplete table was prepared. Complete the table and test whether there is evidence that the salary distribution is non-normal at the 5% level of significance.

Monthly Salary	Observed Frequency	Probability
Under RM800	26	.023
RM800 – RM 1000	146	.136
RM1000 – RM1200	361	
RM1200 – RM1400	311	.341
RM1400 – RM1600	143	.136
RM1600 or above		

[100 marks]

4. (a) *Kebanyakan pusat rawatan air menambah kepekatan fluorida asli untuk mencapai suatu kepekatan flourida sasaran dalam air minuman. Sesetengah aras dipercayai menambah kesihatan gigi tetapi kepekatan yang terlalu tinggi boleh membahayakan. Andaikan suatu pusat rawatan mensasar 0.75 mg/L bagi airnya. Pusat tersebut menguji 25 sampel setiap hari untuk menentukan sama ada aras median berbeza daripada aras sasaran.*

   (i) *Andaikan sampel-sampel dalam satu hari menghasilkan 18 nilai yang melebihi 0.75 mg/L. Jalankan suatu ujian hipotesis dengan* $\alpha = 0.10$.

   (ii) *Terangkan implikasi ralat Jenis I dalam konteks masalah ini.*

   (iii) *Apabila dicadangkan kepada penyelia pusat rawatan tersebut bahawa ujian-t patut digunakan untuk ujian hariannya, ia menjawab bahawa taburan kebarangkalian bagi kepekatan flourida "sangat terpencong ke kanan". Tunjukkan secara bergraf maksud penyelia tersebut dan terangkan kenapa jawapan ini merupakan sebab ujian tanda lebih dicenderungi daripada ujian-t.*

   (b) *Sebuah syarikat farmaseutikal telah mensintesis dua kompoun baru untuk digunakan dalam dua ubat tidur yang berbeza. Data dalam jadual yang berikut menunjukkan bilangan jam tambahan dalam tidur yang diperoleh oleh 10 pesakit melalui penggunaan kedua-dua ubat tersebut. Adakah data ini memberikan bukti cukup untuk menandakan bahawa taburan kebarangkalian bagi bilangan jam tambahan dalam tidur berbeza bagi kedua-dua ubat? Uji dengan menggunakan ujian pangkat bertanda Wilcoxon pada aras keertian 5%.*

Patient	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Drug A	.4	−.7	−.4	−1.4	−1.6	2.9	4.0	.1	3.1	1.9
Drug B	.7	−1.6	−.2	−1.4	−.2	3.4	3.7	.8	.0	2.0

   (c) *Seorang penyelidik sains sosial menghipotesiskan bahawa pendapatan bulanan pekerja-pekerja dalam suatu sektor tertabur secara normal dengan min RM1200 dan sisihan piawai RM200. Suatu sampel rawak 1000 pekerja daripada sektor tersebut diperoleh dan jadual tak lengkap yang berikut disediakan. Lengkapkan jadual ini dan uji sama ada terdapat bukti bahawa taburan pendapatan tersebut adalah tak normal pada aras keertian 5%.*

Pendapatan bulanan	*Kekerapan diamati*	*Kebarangkalian*
kurang drpd. RM800	*26*	*.023*
RM800 – RM1000	*146*	*.136*
RM1000 – RM1200	*361*	
RM1200 – RM1400	*311*	*.341*
RM1400 – RM1600	*143*	*.136*
RM1600 atau lebih		

[100 markah]

.../10-

FORMULA

$$\bar{x} = \frac{\sum xf}{\sum f}$$

$$s^2 = \frac{\sum (x^2 f) - \frac{(\sum xf)^2}{\sum f}}{\sum f - 1}$$

$$S_p{}^2 = \frac{(n_x - 1)s_x{}^2 + (n_y - 1)s_y{}^2}{n_x + n_y - 2}$$

$$\bar{p} = \frac{X + Y}{n_x + n_y}$$

Confidence Intervals:

$$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$(\bar{X} - \bar{Y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_x{}^2}{n_x} + \frac{\sigma_y{}^2}{n_y}}$$

$$\bar{X} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$$

$$(\bar{X} - \bar{Y}) \pm t_{\alpha/2}\sqrt{S_p{}^2\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}$$

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$(\hat{p}_x - \hat{p}_y) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}$$

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \quad to \quad \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

Test Statistics:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

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

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

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

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

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

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

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

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

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

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

.../11-