UNIVERSITI SAINS MALAYSIA

Second Semester Examination
2009/2010 Academic Session

April/May 2010

MAT 161 – 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 ten** [10] questions.

*[**Arahan:** Jawab **semua 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. In a survey of foreign tourists arriving in Penang, the ages of these tourists are as given in the following frequency distribution table.

Age (X years)	Frequency
16 – 25	11
26 – 35	15
36 – 45	35
46 – 55	40
56 – 65	15
66 – 75	10

(i) Find the mean and standard deviation of the ages of the tourists.
(ii) What percentage of the tourists is between 30 and 50 years old?
(iii) Using Chebyshev's Theorem, find the limits of the interval which contains at least 75% of the ages of foreign tourists arriving in Penang.

[15 marks]

2. The Ministry of Science and Innovation (MOSTI) randomly selects 750 Science and Technology research grants it approved over the last five years. MOSTI then classifies each of these grants according to two criteria: (a) whether it was completed early, on-time or late, and (b) whether or not the research produces a tangible product. The table below provides the outcome of this two-way classification.

Research Output	Completion Time		
	Early	On-Time	Late
No tangible product	266	142	74
Produces a tangible product	54	95	119

(a) If one research grant is selected at random from these 750 research grants, what is the probability that it
  (i) was completed on time?
  (ii) produced a tangible product or it finished late?
  (iii) produced no tangible product given that it finished early?

(b) Are the two events "Research produced a tangible product" and "completed early" independent? Are they mutually exclusive? Explain why or why not?

[10 marks]

1. *Dalam suatu tinjauan ketibaan pelancong asing ke Pulau Pinang, usia (umur) pelancong-pelancong ini adalah seperti yang diberikan dalam taburan kekerapan yang berikut*

Umur (X tahun)	*Kekerapan*
15 – 25	*11*
26 – 35	*15*
36 – 45	*35*
46 – 55	*40*
56 – 65	*15*
66 – 75	*10*

(i) Cari min dan sisihan piawai bagi umur pelancong-pelancong ini.

(ii) Berapakah peratusan pelancong-pelancong yang berumur antara 30 hingga 50 tahun?

(iii) Dengan mengguna Teorem Chebyshev dapatkan had-had selang yang mengandungi sekurang-kurangnya 75% daripada umur pelancong asing yang melawat Pulau Pinang .

[15 markah]

2. *Kementerian Sains dan Inovasi (MOSTI) memilih secara rawak 750 geran penyelidikan yang telah diperakukan sepanjang lima tahun kebelakangan. MOSTI kemudiannya mengelaskan setiap daripada geran-geran penyelidikan ini dengan menggunakan dua kriteria: (a) sama ada penyelidikan disiapkan awal, tepat pada masa, atau lewat, dan (b) sama ada penyelidikan itu menghasilkan suatu produk atau tidak. Jadual di bawah adalah hasil pengelasan dua-hala ini.*

Output Penyelidikan	*Masa Siap Penyelidikan*		
	Awal	*Tepat pada masa*	*Lewat*
Tiada produk	*266*	*142*	*74*
Menghasilkan suatu produk	*54*	*95*	*119*

(a) Sekiranya suatu geran penyelidikan dipilih secara rawak daripada 750 geran penyelidikan di atas, apakah kebarangkalian bahawa penyelidikan ini

(i) siap awal?

(ii) menghasilkan suatu produk?

(iii) tidak menghasilkan sebarang produk jika diketahui ia disiapkan awal?

(b) Adakah perisitiwa "penyelidikan menghasilkan suatu produk" dan "penyelidikan disiapkan awal" tak bersandar? Adakah kedua-duanya saling ekslusif? Terangkan mengapa?

[10 markah]

3. A shooting competition is entered by teams of five people. Each person in a team is given one chance to hit the target. The table below shows the number of points given for the number of hits by a team.

Number of hits	0	1	2	3	4	5
Number of points	0	2	4	8	12	18

For team *M*, the probability that each member of the team independently hits a target is 0.7. Find

(i) the probability of the team hitting the target *r* times, for $r = 0, 1, 2, 3, 4, 5$.
(ii) the team's expected number of points.
(iii) the probability that the team earns more than 10 points.

[20 marks]

4. The cooking section of a popular restaurant takes time to prepare the dishes ordered by its customers. The duration from the time a customer places an order to the time the dishes are served to the customer is normally distributed with a mean of 30 minutes and a standard deviation of 6 minutes.

(i) Calculate the probability that a randomly selected customer will have to wait for less than 24 minutes to get his/her dishes after making the order?
(ii) Find the probability that the mean waiting time for dishes for a sample of 15 customers is within 5 minutes of the population mean.
(iii) Suppose that the management guarantees that a meal voucher of RM10 will be given to customers who had to wait for more than a specific duration. What should this specific duration be if the management wants to ensure that no more than 10% of the customers will be receiving the RM10 meal vouchers?

[10 marks]

5. A machine is programmed to produce 3-cm nails. A sample of 15 nails is randomly selected and the lengths measured. The results (in x cm) are summarized as follows: $\sum x = 45.11$, $\sum x^2 = 135.71$.

(i) Find a 95% confidence interval for μ, the mean length of a nail produced by the machine.
(ii) Determine the minimum number of nails that you need to sample if you would like to be 99% confident that an estimate of the mean length of a nail is within 0.01 cm of the true mean.
(iii) In a sample of 150 nails, 20 nails are rejected because they are either too short or too long. Find a 95% confidence interval for the proportion of nails that are acceptable.

[15 marks]

3. *Suatu peraduan menembak disertai oleh pasukan-pasukan dengan empat orang ahli. Setiap ahli pasukan diberi satu peluang untuk menembak sasaran. Jadual yang berikut menunjukkan bilangan mata yang diberikan untuk bilangan tembakan yang mengenai sasaran bagi satu pasukan.*

Bilangan tembakan yang mengenai sasaran	*0*	*1*	*2*	*3*	*4*	*5*
Biilangan mata	*0*	*2*	*4*	*8*	*12*	*20*

Bagi pasukan M, kebarangkalian tembakan setiap ahlinya, mengenai sasaran secara tak bersandar ialah 0.7. Dapatkan

(i) kebarangkalian tembakan pasukan itu mengenai sasaran sebanyak r kali, bagi r = 0, 1, 2, 3, 4, 5.

(ii) bilangan mata jangkaan pasukan tersebut.

(iii) kebarangkalian bahawa pasukan tersebut mendapat lebih daripada 10 mata.

[20 markah]

4. *Bahagian memasak di sebuah restoran yang popular mengambil masa untuk menyediakan hidangan yang diminta pelanggan. Tempoh masa dari saat pesanan dibuat pelanggan hinggalah hidangan disaji untuk dimakan pelanggan, adalah tertabur secara Normal dengan min 30 minit dan sisihan piawai 6 minit.*

(i) Hitung kebarangkalian bahawa seorang pelanggan yang dipilih secara rawak perlu menunggu kurang daripada 24 minit untuk mendapatkan hidangan yang dipesan?

(ii) Cari kebarangkalian bahawa min masa tunggu untuk hidangan bagi sampel rawak 15 orang pelanggan adalah berbeza tidak melebihi 5 minit daripada min populasi.

(iii) Andaikan bahawa pihak pengurusan menjamin baucer makanan RM10 akan diberikan kepada semua pelanggan yang perlu menunggu lebih daripada suat tempoh tertentu. Apakah tempoh ini, jika pihak pengurusan ingin memastikan bahawa tidak lebih daripada 10% pelanggan-pelanggannya akan mendapat baucer makanan RM10 ini?

[10 markah]

5. *Sebuah mesin diprogramkan untuk menghasilkan batang-batang paku-3-cm. Suatu sampel 15 batang paku dipilih secara rawak dan panjangnya diukur. Hasilnya (dalam x cm) diringkaskan seperti yang berikut:* $\sum x = 45.11$, $\sum x^2 = 135.71$

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

(ii) Tentukan bilangan paku yang minimum yang perlu disampelkan jika anda ingin 99% yakin bahawa anggaran min panjang sebatang paku adalah dalam sekitar 0.01 cm daripada min sebenar.

(iii) Dalam suatu sampel 150 batang paku, 20 batang ditolak kerana terlalu panjang atau terlalu pendek. Dapatkan suatu selang keyakinan 95% bagi kadaran bilangan paku yang diterima.

[15 markah]

…6/-

6. It is reported that the participations of first year undergraduate students in three courses in one university can be studied in their ratios. The participations of students in the course of Mathematics, Statistics and Operations Research take the ratios of 5:3:2. A total of 100 students are surveyed and the survey reveals the following result:

Mathematics	Statistics	Operation Research
56	32	12

(i) If the null hypothesis (5:3:2) is true, what is the expected number of students that participate in the Mathematics course?

(ii) Complete the hypothesis test using the 0.01 significance level.

[15 marks]

7. A producer of canned fruit uses a new branded machine to fill cans. The amount of fruit dispensed into the cans forms a normal distribution with a mean 500 grams and a standard deviation 7 grams. A sample of 10 cans is taken randomly per hour. The hour's production is accepted if the sample mean is between 495.5 and 504.5; otherwise, it is rejected and the machine is recalibrated before continuing.

(i) Find the probability of the type I error by rejecting the previous hour's production when the mean can weight is 500 grams. Give your answer in an approximate value.

(ii) Find the probability of the type II error by accepting the previous hour's production when the mean can weight is actually 498 grams.

[15 marks]

8. Supermarket A requires that all eggs from egg suppliers must meet the store's specifications of a mean diameter of 4.0cm and a standard deviation of no more than 0.2cm. Supermarket A sends a buyer to a potential egg supplier and selects a random sample of 50 eggs from the supplier's farm. The diameter of each egg is measured and the mean is reported to be 3.95 with a standard deviation of 0.24.

(i) Show if the sample sufficiently meets the specification with regard to the mean diameter. Use a significance level of 0.05.

(ii) Is there any sufficient evidence that the eggs do not meet the specification with regard to the standard deviation? Conduct a hypothesis test at the 0.05 significance level.

(iii) Compare the results of (i) and (ii). Should supermarket A takes the eggs from the eggs supplier?

[20 marks]

6. *Terdapat laporan yang menyatakan bahawa penyertaan pelajar ijazah tahun pertama dalam tiga kursus di sebuah universiti boleh dikaji dari nisbah mereka. Penyertaan pelajar dalam kursus Matematik, Statistik dan Penyelidikan Operasi adalah dalam nisbah 5:3:2. Seramai 100 mahasiswa ditinjau dan hasil tinjauan adalah seperti yang berikut.*

Matematik	*Statistik*	*Penyelidikan Operasi*
56	*32*	*12*

(i) Jika hipotesis nol (5:3:2) adalah benar, berapakah bilangan pelajar yang dijangka menyertai kursus Matematik?

(ii) Lengkapkan pengujian hipotesis dengan menggunakan aras keertian 0.01.

[15 markah]

7. *Seorang pengeluar buah-buahan bertin menggunakan sebuah mesin berjenama baru untuk pengisian. Amaun buah-buahan yang diisi ke dalam tin membentuk satu taburan normal dengan min 500 gram dan sisihan piawai 7 gram. Satu sampel 10 tin diambil secara rawak setiap jam. Pengeluaran dalam tempoh tersebut diterima jika min sampel adalah di antara 495.5 dan 504.5; jika sebaliknya, ia ditolak dan mesin itu diselaraskan semula sebelum pengeluaran diteruskan.*

(i) Cari kebarangkalian ralat jenis I dengan menolak pengeluaran pada jam sebelumnya apabila min berat tin adalah 500 gram. Beri jawapan anda dalam nilai yang terdekat.

(ii) Cari kebarangkalian ralat jenis II dengan menerima pengeluaran pada jam sebelumnya apabila min berat tin adalah sebenarnya 498 gram.

[15 markah]

8. *Pasaraya A mensyaratkan bahawa semua telur dari pembekal telur harus memenuhi spesifikasi dengan min diameter 4.0 cm dan sisihan piawai tidak lebih dari 0.2 cm. Pasaraya A menghantar seorang pembeli kepada seorang bakal pembekal dan memilih suatu sampel rawak 50 biji telur daripada ladangnya. Diameter telur diukur dan min dilaporkan 3.95 dengan sisihan piawai 0.24.*

(i) Tunjukkan sama ada sampel tersebut memenuhi spesifikasi dengan secukupnya dari segi min diameter. Gunakan aras keertian 0.05.

(ii) Adakah terdapat bukti yang menunjukkan bahawa telur tidak memenuhi spesifikasi dari segi sisishan piawai? Lakukan ujian hipotesis pada aras keertian 0.05.

(iii) Bandingkan keputusan (i) dan (ii). Haruskah pasaraya A mengambil telur daripada pembekal telur tersebut?

[20 markah]

9. The table below shows the average daily temperature (in Degree Celsius) of 10 days in September for two cities, Hannover and Hamburg. The samples are assumed to be normally distributed.

Hannover	14.0	12.5	11.5	12.2	12.4	12.3	11.9	12.1	12.8	13.2
Hamburg	13.2	12.2	12.8	12.9	13.3	12.9	12.2	11.6	12.8	14.2

Hannover : $\bar{x} = 12.49 \quad s^2 = 0.4988 \quad s = 0.7062$
Hamburg : $\bar{x} = 12.81 \quad s^2 = 0.5054 \quad s = 0.7109$

(i) What parametric test should you apply in order to test if the average temperature in Hamburg is higher than the temperature in Hannover. Why do you choose this test?
(ii) Conduct the parametric test you mention in (i). Use a significance level of 0.05.
(iii) What nonparametric test can be used if we are unsure about the distribution of the samples?
(iv) What does the nonparametric test in part (iii) show? Use the 0.05 level of significance.
(v) Do both tests report the same result?

[20 marks]

10. It is claimed that the passing rate of a mathematic course for the first year undergraduate students in university B is higher than the passing rate of university A. In order to investigate how true this statement holds, two independent samples are taken randomly. The results are summarized in the following table.

Universities	n	No. of students who fail
A	220	23
B	182	12

(i) Conduct a hypothesis test to determine if there is a difference in the proportion of the two independent samples. Use the significance level α =0.05.
(ii) Suggest another possible test to determine the difference in the proportion of two samples. Is this test equivalent to the test conducted in (i)? Explain.

[10 marks]

9. *Jadual di bawah menunjukkan purata suhu harian (dalam darjah celsius) selama 10 hari dalam bulan September untuk dua bandar, Hannover dan Hamburg. Sampel tersebut diandaikan tertabur secara normal.*

Hannover	*14.0*	*12.5*	*11.5*	*12.2*	*12.4*	*12.3*	*11.9*	*12.1*	*12.8*	*13.2*
Hamburg	*13.2*	*12.2*	*12.8*	*12.9*	*13.3*	*12.9*	*12.2*	*11.6*	*12.8*	*14.2*

Hannover : $\bar{x} = 12.49 \quad s^2 = 0.4988 \quad s = 0.7062$

Hamburg : $\bar{x} = 12.81 \quad s^2 = 0.5054 \quad s = 0.7109$

(i) Apakah ujian berparametrik yang patut anda guna untuk menguji sama ada purata suhu di Hamburg lebih tinggi daripada purata suhu di Hannover? Mengapakah anda memilih ujian ini?

(ii) Jalankan ujian berparametrik yang anda sebut dalam (i). Gunakan aras keertian 0.05.

(iii) Apakah ujian tak berparametrik yang boleh digunakan jika kita tidak pasti tentang taburan sampel?

(iv) Apakah keputusan yang ditunjukkan oleh ujian tak berparametrik yang disebut dalam bahagian (iii)? Gunakan aras keertian 0.05.

(v) Adakah kedua-dua ujian melaporkan keputusan yang sama?

[20 markah]

10. *Terdapat suatu dakwaan bahawa kadar lulus suatu kursus matematik untuk pelajar ijazah tahun pertama di Universiti A lebih tinggi daripada kadar lulus di Universiti B. Untuk menyiasat kebenaran dakwaan ini, dua sampel tak bersandar diambil secara rawak. Hasilnya diringkaskan dalam jadual berikut.*

Universiti	*n*	*Bil. pelajar gagal*
A	*220*	*23*
B	*182*	*12*

(i) Lakukan ujian hipotesis untuk menentukan sama ada terdapat perbezaan dalam kadar lulus antara dua sampel itu. Gunakan aras keertian $\alpha = 0.05$.

(ii) Cadangkan ujian lain yang mungkin untuk menentukan perbezaan dalam perkadaran antara dua sampel. Adakah ujian ini menyamai ujian yang dijalankan di bahagian (i)? Jelaskan jawapan anda.

[10 markah]

APPENDIX/FORMULA

$\bar{x} = \frac{\sum xf}{\sum f}$	$s_p^2 = \frac{(n_x - 1)s_2^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$
$s^2 = \frac{\sum (x^2 f) - \frac{(\sum xf)^2}{\sum f}}{\sum f - 1}$	$p'_p = \frac{x_1 + x_2}{n_1 + n_2}$
Confidence Intervals:	$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$	$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$	$(\bar{x}_1 - \bar{x}_2) \pm t_{df,\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$
$p' \pm z_{\alpha/2}\sqrt{\frac{p'(1-p')}{n}}$	$p'_1 - p'_2 \pm z_{\alpha/2}\sqrt{\frac{p'_1(1-p'_1)}{n_1} + \frac{p'_2(1-p'_2)}{n_2}}$
$\frac{(n-1)s^2}{\chi^2_{\alpha/2}}$ *to* $\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$	
Test Statistics:	
$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$	$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n_d}}}$
$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$	$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$

…11/-

Test Statistics:

$$z=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

$$t=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

$$t=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$z=\frac{(p_1'-p_2')-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}}$$

$$z=\frac{(p_1'-p_2')-(p_1-p_2)}{\sqrt{p_p'\ (1-p_p')\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

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

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

Nonparametric Statistics:

Wilcoxon Signed-rank: $W=\sum R^+$, $W=\sum R^-$

$$Z=\frac{T-\mu_W}{\sigma_W}\ ,\quad \mu_W=\frac{n(n+1)}{4},\ \sigma_W=\sqrt{\frac{n(n+1)(2n+1)}{24}}$$

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

$$Z=\frac{T-\mu_T}{\sigma_T}\ ,$$

$$\mu_T=\frac{n_1(n_1+n_2+1)}{2},\ \sigma_T=\sqrt{\frac{n_1n_2(n+1)(n_1+n_2+1)}{12}}$$

…12/-

Mann-Whitney U Test:

Mann-Whitney U test-statistic:

$$U_a = n_a \cdot n_b \cdot \frac{n_b \quad n_b + 1}{2} - R_b$$

$$U_b = n_a \cdot n_b \cdot \frac{n_a \quad n_a + 1}{2} - R_a$$

U^* is the test statistic , i.e. the smaller of U_a and U_b

Normal approximation:

Mean: $\mu_u = \frac{n_a \cdot n_b}{2}$

Standard deviation: $\sigma_u = \sqrt{\frac{n_a \cdot n_b \quad n_a + n_b + 1}{12}}$

Test-statistic: $z^* = \frac{U^* - \mu_u}{\sigma_u}$

- ooo O ooo -