UNIVERSITI SAINS MALAYSIA

First Semester Examination
2011/2012 Academic Session

January 2012

MAT 161 – Elementary Statistics
[Statistik Permulaan]

Duration : 3 hours
[Masa : 3 jam]

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

[Sila pastikan bahawa kertas peperiksaan ini mengandungi SEPULUH 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. The following table shows the daily allowance for all 60 employees of Company A.

Daily Allowance (RM)	No. of employees
50-59	6
60-69	10
70-79	16
80-89	14
90-99	10
100-109	4

(i) Calculate the mean and standard deviation of the distribution of the daily allowance

(ii) Estimate the percentage of workers who get an allowance of less than RM90 a day.

(iii) Use Chebyshev's theorem to obtain an interval of the daily allowance for at least 75% of the employees of the company.

[10 marks]

1. Jadual berikut menunjukkan elaun harian bagi kesemua 60 orang pekerja di Syarikat A.

Elaun harian (RM)	*Bil. Pekerja*
50-59	*6*
60-69	*10*
70-79	*16*
80-89	*14*
90-99	*10*
100-109	*4*

(i) Hitungkan min dan sisihan piawai bagi taburan elaun harian

(ii) Anggarkan peratusan pekerja yang mendapat elaun kurang daripada RM90 sehari.

(iii) Gunakan Teorem Chebyshev untuk mendapatkan suatu selang elaun harian bagi sekurang-kurangnya 75% daripada pekerja-pekerja syarikat tersebut.

[10 markah]

2. A problem in organ transplant is the body's rejection of the transplanted tissue. If the antigens attached to the tissue cells of the donor and receiver match, the body will accept the transplanted tissue. While the antigens in identical twins always match, the probability of a match in other siblings is 0.25 and that of a match in two people from the population at large is 0.001. Suppose you need a kidney, and you have two brothers and a sister.

(i) If one of your three siblings offers a kidney, what is the probability that the antigens will match?

(ii) If all three siblings offer a kidney, what is the probability that all three antigens will match?

(iii) If all three siblings offer a kidney, what is the probability that none of the antigens will match?

(iv) Repeat parts (ii) and (iii), this time assuming that the three donors were obtained from the population at large.

[10 marks]

2. *Satu masalah dalam pemindahan organ adalah penolakan badan terhadap tisu yang dipindahkan. Jika antigen yang ada pada sel-sel tisu penderma dan penerima adalah sepadan, badan akan menerima tisu yang dipindahkan. Sementara antigen dalam kembar seiras sentiasa sepadan, kebarangkalian kesepadanan dalam adik-beradik lain adalah 0.25 dan kesepadanan dalam dua orang dari populasi keseluruhannya ialah 0.001. Katakan anda memerlukan buah pinggang, dan anda mempunyai dua abang dan seorang kakak.*

(i) Jika salah seorang daripada tiga adik-beradik anda menawarkan buah pinggang, apakah kebarangkalian bahawa antigen akan sepadan?

(ii) Jika ketiga-tiga adik-beradik menawarkan buah pinggang, apakah kebarangkalian bahawa ketiga-tiga antigen akan sepadan?

(iii) Jika ketiga-tiga adik-beradik menawarkan buah pinggang, apakah kebarangkalian bahawa tidak ada daripada antigen akan sepadan?

(iv) Ulang bahagian (ii) dan (iii), kali ini menganggap bahawa ketiga-tiga penderma yang diperoleh adalah daripada populasi pada umumnya.

[10 markah]

3. Researchers believed that one in every three women has been a victim of domestic abuse. This probability was obtained from a survey of nearly 2000 adult women residing in Baltimore, Maryland. Suppose we randomly sample 15 women and find that four have been abused.

(i) What is the probability of observing four or more abused women in a sample of 15 if the proportion p of women who are victims of domestic abuse is really $p = 1/3$?

(ii) Many experts on domestic violence believe that the proportion of women who are domestically abused is closer to $p = 0.10$. Calculate the probability of observing four or more abused women in a sample of 15 if $p = 0.10$.

(iii) Why might your answers to parts (i) and (ii) lead you to believe that $p = 1/3$? Explain.

[10 marks]

3. *Penyelidik percaya bahawa satu dalam setiap tiga wanita telah menjadi mangsa penderaan domestik. Kebarangkalian ini adalah diperolehi daripada kajian hampir 2000 wanita yang menetap di Baltimore, Maryland. Katakan kita secara rawak mengambil sampel 15 wanita dan mendapati bahawa empat telah didera.*

   (i) Apakah kebarangkalian memerhati empat atau lebih wanita yang telah dianiayai dalam sampel 15 orang jika kadaran wanita yang menjadi mangsa keganasan rumahtangga adalah benar-benar p = 1 / 3?
   (ii) Ramai pakar dalam keganasan rumah tangga percaya bahawa kadaran wanita yang didera adalah mendekati p = 0.10. Hitung kebarangkalian memerhatikan empat atau lebih wanita didera dalam sampel sebanyak 15 jika p = 0.10.
   (iii) Mengapa kemungkinan jawapan anda kepada bahagian (i) dan (ii) membawa anda untuk mempercayai bahawa p = 1 / 3? Jelaskan.

   [10 markah]

4. Life span for a type of battery produced by a plant is normally distributed with a mean of 29 months and a standard deviation of four months. A battery has been randomly selected. What is the probability that the battery has a life span

   (i) Between 20 and 36 months?
   (ii) More than 25 months?
   (iii) Less than 32 months?
   (iv) What is the minimum lifetime for the 25% most long-lasting battery?

   [10 marks]

4. *Masa hayat bagi suatu jenis bateri yang dihasilkan oleh sebuah kilang tertabur secara normal dengan min 29 bulan dan sisihan piawai 4 bulan. Suatu bateri itu dipilih secara rawak. Apakah kebarangkalian bateri itu mempunyai hayat*

   (i) Antara 20 dan 36 bulan?
   (ii) Lebih daripada 25 bulan?
   (iii) Kurang daripada 32 bulan?
   (iv) Apakah masa hayat minimum bagi 25% bateri yang tahan paling lama?

   [10 markah]

5. Postmortem interval (PMI) is the time elapsed between death and the performance of an autopsy on a brain specimen. The data below is the PMI of 22 randomly selected brain specimens obtained at autopsy.

5.5	14.5	6.0	5.5	5.3	5.8	11.0	6.4
7.0	14.5	10.4	4.6	4.3	7.2	10.5	6.5
3.3	7.0	4.1	6.2	10.4	4.9		

(i) Construct a 95% confidence interval for the true mean PMI of brain specimens obtained at autopsy.
(ii) Interpret the interval in part (i).
(iii) What assumption is required for the interval in part (i) to be valid?
(iv) What does it mean by the phrase "95% confidence"?

[10 marks]

5. Selang bedah siasat (PMI) adalah masa yang berlalu antara kematian dan prestasi bedah siasat ke atas suatu spesimen otak. Data di bawah adalah PMI bagi 22 spesimen otak yang dipilih secara rawak yang diperolehi di bedah siasat.

5.5	*14.5*	*6.0*	*5.5*	*5.3*	*5.8*	*11.0*	*6.4*
7.0	*14.5*	*10.4*	*4.6*	*4.3*	*7.2*	*10.5*	*6.5*
3.3	*7.0*	*4.1*	*6.2*	*10.4*	*4.9*		

(i) Bina selang keyakinan 95% bagi min sebenar PMI spesimen otak yang diperolehi semasa bedah siasat.
(ii) Tafsirkan selang dalam bahagian (i).
(iii) Apakah andaian yang diperlukan supaya selang keyakinan di bahagian (i) adalah sah?
(iv) Apa yang dimaksudkan oleh ungkapan "keyakinan 95%"?

[10 markah]

6. A study on the effectiveness of a skin cream on 33 women was done for a period of twenty two weeks. At the end of the study period, a dermatologist judged whether each woman exhibited skin improvement. The results for the 33 women (where I = improved skin and N = no improvement) are listed below:

I	I	N	I	N	N	I	I	I	I	I	I
N	I	I	I	N	I	I	I	N	I	N	I
I	I	I	I	I	N	I	I	N			

(i) Do the data provide sufficient evidence to conclude that the cream will improve the skin of more than 60% of women? Test using $\alpha = 0.05$.
(ii) Find and interpret the *p*-value of the test.

[10 marks]

6. *Satu kajian mengenai keberkesanan krim kulit pada 33 wanita telah dilakukan untuk tempoh 22 minggu. Pada akhir tempoh kajian, dermatologi menilai sama ada setiap wanita menunjukkan penambahbaikan kulit. Keputusan untuk 33 wanita (di mana I = penambahbaikan kulit dan N = tiada penambahbaikan) adalah seperti berikut:*

I	*I*	*N*	*I*	*N*	*N*	*I*	*I*	*I*	*I*	*I*	*I*
N	*I*	*I*	*I*	*N*	*I*	*I*	*I*	*N*	*I*	*N*	*I*
I	*I*	*I*	*I*	*I*	*N*	*I*	*I*	*N*			

(i) Adakah data memberi bukti yang mencukupi untuk membuat kesimpulan bahawa krim akan memperbaiki kulit lebih daripada 60% wanita yang dikaji? Uji dengan menggunakan $\alpha = 0.05$.

(ii) Cari dan tafsir nilai-p ujian tersebut.

[10 markah]

7. Old records have shown that the mean scores for the mathematics test for Form One in a school is 48 marks and the standard deviation is 15 marks. A mathematics teacher felt that the Form One students are more intelligent and he took a random sample of 40 students and found that their mean scores is 54.

(i) Test at 1% level of significance whether there is evidence to support the opinion of the teacher.

(ii) If the mean actual score was 56 points, calculate the Type II error.

[10 marks]

7. *Rekod lama menunjukkan bahawa min markah bagi ujian matematik untuk Tingkatan Satu di sebuah sekolah ialah 48 markah dan sisihan piawai ialah 15 markah. Seorang guru matematik berpendapat bahawa pelajar Tingkatan Satu ini lebih cerdik, maka dia mengambil satu sampel rawak 40 orang pelajar dan mendapati bahawa min markah mereka ialah 54.*

(i) Uji pada aras keertian 1% sama ada terdapat bukti untuk menyokong pendapat guru itu.

(ii) Jika min markah sebenar ialah 56 markah, hitung ralat Jenis II.

[10 markah]

8. Weight gain was recorded for two groups of animals which were given a diet with low protein or high protein. Group X received a high protein diet, while group Y received a low protein diet.

	Weight Gain							
Group *X*	0.94	0.79	0.96	0.98	1.02	1.02	1.09	0.91
Group *Y*	0.49	0.82	0.73	0.86	0.81	0.97	1.06	0.70

(i) Test the hypothesis whether the two are of the same variances. Use $\alpha = 0.05$.
(ii) Using the result of part (i), can we infer that both diets resulted in similar weight? Use $\alpha = 0.05$.
(iii) Using the result of part (i), build a 90% confidence interval for $\mu_X - \mu_Y$.

[10 marks]

8. *Peningkatan berat badan telah direkodkan bagi 2 kumpulan binatang yang telah diberikan diet dengan protein rendah atau protein tinggi. Kumpulan X menerima diet yang berprotein tinggi sementara kumpulan Y menerima diet yang berprotein rendah.*

	Peningkatan Berat Badan							
Kumpulan X	*0.94*	*0.79*	*0.96*	*0.98*	*1.02*	*1.02*	*1.09*	*0.91*
Kumpulan Y	*0.49*	*0.82*	*0.73*	*0.86*	*0.81*	*0.97*	*1.06*	*0.70*

(i) Uji hipotesis sama ada dua varians tersebut adalah sama. Gunakan α = 0.05.
(ii) Dengan menggunakan keputusan di bahagian (i), bolehkah kita membuat kesimpulan bahawa kedua-dua diet itu menghasilkan peningkatan berat badan yang sama. Gunakan α = 0.05.
(iii) Dengan menggunakan keputusan di bahagian (i), binakan selang keyakinan 90% bagi μ_X - μ_Y.

[10 markah]

9. Consider the following data obtained from a population

x	0	1	2	3	4	5
f	8	19	25	22	5	1

(i) Estimate the mean of the distribution of X above.
(ii) At 5% level of significance, test whether the binomial model is suitable for this data.

[10 marks]

9. *Pertimbangkan data berikut yang diperolehi daripada suatu populasi:*

x	*0*	*1*	*2*	*3*	*4*	*5*
f	*8*	*19*	*25*	*22*	*5*	*1*

(i) Anggarkan min bagi taburan X di atas
(ii) Pada aras keertian 5%, uji sama ada model binomial sesuai bagi data ini.

[10 markah]

10. A paired difference experiment with n = 30 pairs yielded $T_+ = 354$.

    (i) Specify the null and alternative hypotheses that should be used in conducting a hypothesis test to determine whether the probability distribution for population A is located to the right of that for population B.
    (ii) Conduct the test of part (i) using $\alpha = 0.05$.
    (iii) What is the approximate *p*-value of the test of part (ii)?
    (iv) What assumptions are necessary to ensure the validity of the test you performed in part (ii)?

    [10 marks]

10. Suatu ujikaji perbezaan yang berpasangan dengan n = 30 pasangan menghasilkan $T_+ = 354$.

    (i) Nyatakan hipotesis nol dan alternatif yang harus digunakan dalam menjalankan ujian hipotesis untuk menentukan sama ada taburan kebarangkalian bagi populasi A terletak di sebelah kanan populasi B.
    (ii) Jalankan ujian bahagian (i) dengan menggunakan $\alpha = 0.05$.
    (iii) Apakah nilai anggaran nilai-p untuk ujian di bahagian (ii)?
    (iv) Andaian apa yang perlu untuk memastikan kesahihan ujian dalam bahagian (ii)?

    [10 markah]

APPENDIX/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: $\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ $\overline{X} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$ $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $\frac{(n-1)s^2}{\chi^2_{\alpha/2}}$ *to* $\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$	$(\overline{X} - \overline{Y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_x^{\,2}}{n_x} + \frac{\sigma_y^{\,2}}{n_y}}$ $(\overline{X} - \overline{Y}) \pm z_{\alpha/2}\sqrt{\frac{s_x^{\,2}}{n_x} + \frac{s_y^{\,2}}{n_y}}$ $(\overline{X} - \overline{Y}) \pm t_{\alpha/2}\sqrt{S_p^{\,2}\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}$ $(\hat{p}_x - \hat{p}_y) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}$
Test Statistics: $Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ $T = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$	$T = \frac{\overline{d} - \mu_d}{\frac{s_d}{\sqrt{n_d}}}$ $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$

Test Statistics:

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

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

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

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

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

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

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

Nonparametric Statistics:

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

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

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

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

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