UNIVERSITI SAINS MALAYSIA

Peperiksaan Kursus Semasa Cuti Panjang
Sidang Akademik 2006/2007

Jun 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.]*

1. (a) The following frequency distribution shows the salary ranges (in thousands of RM) for all middle management personnel in a large company.

Salary of personnel	f
10 – 20	10
20 – 30	25
30 – 40	20
40 – 50	15
50 – 60	5

(i) Calculate the mean salary and its standard deviation.
(ii) Draw a histogram of the salary distribution and describe its shape.
(iii) Find the minimum salary of the top 10% of middle management personnel in the company.
(iv) Using Chebyshev's Theorem, find an interval of salaries of at least 70% of middle management personnel in the company.

[Give your answers correct to two decimal places.]

(b) A consumer testing service is commissioned to rank the top three brands of laundry detergent. A total of 10 brands are to be included in the study.

(i) In how many ways can the consumer testing service arrive at the final ranking?
(ii) If the testing service cannot distinguish the difference between the brands and therefore arrives at the final ranking by random selection, what is the probability that brand X (included in the study) is ranked first?

(c) (i) A sample of 500 primary school students were chosen at random and administered a test to measure aggressiveness. Each student was classified according to one of four categories as shown in the table below:

	Firstborn	Not Firstborn
Aggressive	15%	15%
Not Aggresive	25%	45%

If a student is picked at random and A: {Student is aggressive} and B: {Student is firstborn}, are A and B independent? Explain.

(ii) Let E and F be two independent events. Show that $P(\overline{E} \cap \overline{F}) = P(\overline{E})P(\overline{F})$.

[100 marks]

1. (a) *Taburan kekerapan yang berikut menunjukkan julat pendapatan (dalam ribuan RM) bagi semua personel pentadbiran pertengahan di sebuah syarikat besar.*

Pendapatan personel	*f*
10 – 20	10
20 – 30	25
30 – 40	20
40 – 50	15
50 – 60	5

(i) *Hitung min pendapatan dan sisihan piawainya.*

(ii) *Lukiskan suatu histogram bagi taburan pendapatan di atas dan perihalkan bentuknya.*

(iii) *Berapakah pendapatan minimum bagi 10% atasan personel pentadbiran pertengahan di syarikat tersebut?*

(iv) *Dengan menggunakan Teorem Chebyshev, dapatkan suatu selang pendapatan bagi sekurang-kurangnya 70% daripada personel pentadbiran pertengahan di syarikat tersebut?*

[Berikan jawapan anda dalam dua tempat perpuluhan].

(b) *Sebuah perkhidmatan penguji pengguna diberikan kuasa untuk memberi pangkat kepada tiga jenama pencuci pakaian terbaik. Sejumlah 10 jenis jenama dimasukkan dalam kajian tersebut.*

(i) *Dalam berapa carakah perkhidmatan penguji pengguna tersebut dapat menghasilkan pangkat terakhir?*

(ii) *Jika perkhidmatan pengujian tersebut tidak dapat menentukan perbezaan di antara jenama-jenama dan maka itu menghasilkan pangkat terakhir dengan cara pemilihan rawak, berapakah kebarangkalian bahawa jenama X (termasuk dalam kajian) diberi pangkat pertama?*

(c) (i) *Suatu sampel 500 pelajar sekolah rendah dipilih secara rawak dan diberikan suatu ujian untuk mengukur tahap keagresifan mereka. Setiap pelajar dikelaskan dalam salah satu daripada empat kategori yang ditunjukkan dalam jadual yang berikut:*

	Anak Sulong	*Bukan Anak Sulong*
Agresif	15%	15%
Tidak Agresif	25%	45%

Jika seorang pelajar dipilih secara rawak dan A:{Pelajar adalah agresif} dan B:{Pelajar adalah anak sulong}, adakah A dan B tak bersandar? Jelaskan.

(ii) *Andaikan E dan F adalah dua peristiwa tak bersandar. Tunjukkan bahawa* $P(\overline{E} \cap \overline{F}) = P(\overline{E})P(\overline{F})$.

[100 markah]

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

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

(i) Determine the value of k.
(ii) Find $E(X)$ and $Var(X)$.
(iii) The median for a continuous probability distribution $f(x)$ 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 the above $f(x)$.

(b) A medical journal reported that about 70% of the individuals needing a kidney transplant find a suitable donor when they turn to registries of unrelated donors.

(i) In a group of ten individuals needing a kidney transplant, find the probability that no more than five will find a suitable donor among the registries of unrelated donors.
(ii) Approximate the probability that in a group of 50 individuals needing a kidney transplant, at least 50% of them find a suitable donor among the registries of unrelated donors.

(c) The waiting time, X, at a fast-food restaurant during lunch time is known to be approximately normally distributed with a mean of 4.5 minutes and a standard deviation of 1.2 minutes.

(i) Find the probability that in a sample 10 randomly selected customers, the mean waiting time is not more than 5.0 minutes.
(ii) The restaurant has a policy that a customer who has to wait for more than c minutes gets a free meal. If 0.1% of their customers get a free meal, what is the value of c?
(iii) Find the probability that in a random sample of 10 customers, at most only one will get a free meal.

[100 marks]

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

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

(i) *Tentukan nilai k.*

(ii) *Dapatkan E(X) dan Var(X).*

(iii) *Median bagi suatu taburan kebarangkalian selanjar f(x) ialah suatu nilai m yang membahagikan taburan tersebut kepada dua keluasan yang sama, iaitu:*

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

Dapatkan median bagi f(x di atas.

(b) *Sebuah jurnal perubatan melaporkan bahawa 70% daripada individu yang memerlukan pemindahan buah pinggang mendapat penderma yang sesuai daripada daftar penderma yang tiada pertalian.*

(i) *Dalam suatu kumpulan 10 individu yang memerlukan pemindahan buah pinggang, dapatkan kebarangkalian bahawa tidak lebih daripada lima orang akan mendapat penderma yang sesuai daripada daftar penderma yang tiada pertalian.*

(ii) *Anggarkan bahawa dalam suatu kumpulan 50 individu yang memerlukan pemindahan buah pinggang, sekurang-kurangnya 50% daripada mereka akan mendapat penderma yang sesuai daripada daftar penderma yang tiada pertalian.*

(c) *Masa menunggu, X, di sebuah restoran makanan segera semasa waktu makan tengahari diketahui tertabur secara normal dengan min 4.5 minit dan sisihan piawai 1.2 minit.*

(i) *Dapatkan kebarangkalian bahawa dalam suatu sampel 10 orang pelanggan yang dipilih secara rawak, min masa menunggunya adalah tidak lebih daripada 5.0 minit.*

(ii) *Restoran tersebut mempunyai polisi bahawa seorang pelanggan yang menunggu lebih daripada c minit akan mendapat makanan percuma. Jika 0.1% daripada pelanggan mereka mendapat makanan percuma, apakah nilai c?*

(iii) *Dapatkan kebarangkalian bahawa dalam suatu sampel rawak 10 orang pelanggan, paling banyak hanya seorang sahaja yang akan mendapat makanan percuma.*

[100 markah]

...6/-

3. (a) Suppose that the amount of time teenagers spend on the internet per day is normally distributed with a standard deviation of 1.5 hours. A sample of 100 teenagers is selected at random, and the sample mean computed is 6.5 hours.

(i) Find a 95% confidence interval of the population mean.
(ii) Determine the minimum number of teenagers that need to be sampled if you would like to be 99% confident that your estimate of the population mean is within 30 minutes of its true value.
(iii) Over a period of several months, it was found that in a sample of 1500 teenagers, 105 spent more than 8 hours on the internet per day which is considered too long. Calculate a 95% confidence interval for the proportion of teenagers who spend too much time on the internet.

(b) A microbiologist claims that 30% of pre-cooked chicken sold in supermarkets are contaminated with the listeria germ. A popular supermarket chain ran a test on a random sample of 20 pre-cooked chickens from its supplier. If fewer than 3 of the chickens proved to be contaminated with listeria, it will contest the microbiologist's claim.

(i) State a suitable null and alternative hypotheses for the above problem.
(ii) State the critical region and determine the significance level of the test.
(iii) Determine the power of the test if 15% of the supermarket chain's chicken are contaminated with listeria.

(c) A test concerning some of the fundamental facts about AIDS was administered to two groups; one consisting of college graduates and the other consisting of high school graduates. A summary of the test scores are as follows:

College graduates (X): $n = 13, \quad \bar{x} = 80.5, \quad s^2 = 25.0$

High school graduates (Y): $n = 25, \quad \bar{y} = 60.5, \quad s^2 = 30.0$

(i) At $\alpha = 0.05$, test whether there is a difference in the variance of the scores of the two groups. What assumption is necessary?
(ii) Construct a 95% confidence interval for $\mu_X - \mu_Y$.
(iii) Using the p-value approach, test at the 5% significance level, that college graduates' scores are higher than high school graduate' scores.

[100 marks]

3. (a) *Andaikan amaun masa yang dihabiskan oleh remaja dengan melayari internet setiap hari tertabur secara normal dengan sishan piawai of 1.5 jam. Suatu sampel 100 orang remaja dipilih secara rawak dan min sampel yang dihitung ialah 6.5 jam.*

   (i) *Dapatkan selang keyakinan 95% bagi min populasinya.*
   (ii) *Tentukan bilangan remaja yang minimum yang patut disampelkan jika anda ingin 99% yakin bahawa anggaran anda mengenai min populasi adalah dalam sekitar 30 minit daripada nilainya yang sebenar.*
   (iii) *Dalam suatu tempoh beberapa bulan, didapati bahawa dalam suatu sampel 1500 orang remaja, 105 orang menghabiskan lebih daripada 8 jam dengan melayari internet yang dianggap terlalu lama. Dapatkan suatu selang keyakinan 95% bagi kadaran remaja yang menghabiskan terlalu banyak masa melayari internet.*

(b) *Seorang ahli mikrobiologi mendakwa bahawa 30% daripada ayam pra-masak yang dijual di pasaraya-pasaraya tercemar dengan kuman listeria. Sebuah rangkaian pasaraya terkemuka menjalankan suatu ujian ke atas suatu sampel rawak 20 ekor ayam pra-masak daripada pembekalnya. Jika kurang daripada 3 ekor terbukti tercemar dengan listeria, ia akan menentang dakwaan ahli mikrobiologi tersebut.*

   (i) *Nyatakan hipotesis nol dan hipotesis alternatif yang sesuai bagi masaalah di atas.*
   (ii) *Nyatakan kawasan genting ujian dan tentukan aras keertiannya.*
   (iii) *Tentukan kuasa ujian jika 15% daripada ayam di rangkaian pasaraya tersebut tercemar dengan listeria.*

(c) *Suatu ujian tentang beberapa fakta asas mengenai AIDS diberikan kepada dua kumpulan; satu kumpulan terdiri daripada graduan kolej dan satu lagi kumpulan terdiri daripada graduan sekolah menengah. Ringkasan skor ujian tersebut adalah seperti yang berikut:*

   Graduan kolej (X): $n = 13$, $\bar{x} = 80.5$, $s^2 = 25.0$

   Graduan sekolah menengah (Y): $n = 25$, $\bar{y} = 60.5$, $s^2 = 30.0$

   (i) *Pada* $\alpha = 0.05$, *uji sama ada terdapat perbezaan dalam varians skor ujian kedua-dua kumpulam. Apakan andaian yang diperlukan?*
   (ii) *Bina suatu selang keyakinan 95% bagi* $\mu_X - \mu_Y$.
   (iii) *Menggunakan kaedah nilai-p, uji pada aras keertian 5% bahawa skor graduan kolej adalah lebih tinggi daripada skor graduan sekolah menengah.*

[100 markah]

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 20 samples each day to determine whether the median level differs from the target.

(i) Suppose that one day's samples result in 15 values that exceed 0.75 mg/L. Conduct a hypothesis test using $\alpha = 0.10$ and state the appropriate conclusion.

(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) The length of time required for a human subject to respond to a new drug was tested by a pharmacist. Seven randomly selected subjects were given both aspirin and the new drug. The two treatments were spaced in time and assigned in random order. The length of time (in minutes) required for a subject to indicate pain relief is recorded for both drugs. The data is shown in the following table.

Subject	1	2	3	4	5	6	7
Aspirin	15	20	12	20	17	14	17
New Drug	7	14	13	11	10	16	11

Do the data provide evidence that it takes longer to get pain relief with aspirin than with the new drug? Test using the Wilcoxon signed-rank test at the 5% significance level.

(c) The null hypothesis of a goodness-of-fit test states that a discrete random variable X has a 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 of 72 observations of X is summarized in the table below:

Value of X	1	2	3	4	5
Frequency	5	7	15	20	25

Is the above data sufficient evidence to reject the the null hypothesis? Test at the 1% significance level.

[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 menjadi merbahaya. Andaikan suatu pusat rawatan mensasar 0.75 mg/L bagi airnya. Pusat tersebut menguji 20 sampel setiap hari untuk menentukan sama ada aras median berbeza daripada aras sasaran.*

(i) *Andaikan sampel-sampel dalam satu hari menghasilkan 15 nilai yang melebihi 0.75 mg/L. Jalankan suatu ujian hipotesis menggunakan $\alpha = 0.10$ dan nyatakan kesimpulan yang bersesuaian.*

(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 grafik maksud penyelia tersebut dan terangkan kenapa jawapannya ialah sebab ujian tanda lebih dicenderungi daripada ujian-t.*

(b) *Tempoh masa yang diperlukan oleh seorang subjek manusia untuk memberi tindak balas terhadap sejenis ubat baru diuji oleh seorang ahli farmasi. Tujuh orang subjek yang dipilih secara rawak diberikan aspirin dan ubat baru tersebut. Kedua-dua rawatan dijarakkan masanya dan diberikan dalam tertib rawak. Tempoh masa (dalam minit) yang diperlukan untuk seorang subjek menandakan kelegaan dicatatkan bagi kedua-dua jenis ubatan. Datanya ditunjukkan dalam jadual yang berikut.*

Subjek	*1*	*2*	*3*	*4*	*5*	*6*	*7*
Aspirin	*15*	*20*	*12*	*20*	*17*	*14*	*17*
Ubat baru	*7*	*14*	*13*	*11*	*10*	*16*	*11*

Adakah data di atas memberikan bukti bahawa tempoh masa yang diperlukan untuk mendapat kelegaan daripada aspirin adalah lebih lama daripada ubat baru tersebut? Uji dengan menggunakan ujian pangkat bertanda Wilcoxon pada aras keertian 5%.

(c) *Hipotesis nol suatu ujian kebaikan penyuaian menyatakan bahawa pembolehubah rawak diskrit X mempunyai taburan kebarangkalian yang diberikan oleh:*

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

Suatu sampel rawak 72 cerapan X diringkaskan dalam jadual yang berikut:

Nilai X	*1*	*2*	*3*	*4*	*5*
Kekerapan	*5*	*7*	*15*	*20*	*25*

Adakah data di atas bukti yang cukup untuk menolak hipotesis nol tersebut? Uji pada aras keertian 1%.

[100 markah]

FORMULA

1. $\bar{x} = \dfrac{\sum xf}{\sum f}$ 2. $s^2 = \dfrac{\sum(x^2 f) - \dfrac{(\sum xf)^2}{\sum f}}{\sum f - 1}$	3. $S_p^{\,2} = \dfrac{(n_x - 1)s_x^{\,2} + (n_y - 1)s_y^{\,2}}{n_x + n_y - 2}$ 4. $\bar{p} = \dfrac{X + Y}{n_x + n_y}$
Confidence Intervals: 1. $\bar{X} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ or $\bar{X} \pm z_{\alpha/2}\dfrac{s}{\sqrt{n}}$ 2. $\bar{X} \pm t_{\alpha/2}\dfrac{s}{\sqrt{n}}$ 3. $\hat{p} \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ 4. $(\bar{X} - \bar{Y}) \pm z_{\alpha/2}\sqrt{\dfrac{\sigma_x^{\,2}}{n_x} + \dfrac{\sigma_y^{\,2}}{n_y}}$	5. $(\bar{X} - \bar{Y}) \pm t_{\alpha/2}\sqrt{S_p^{\,2}\left(\dfrac{1}{n_x} + \dfrac{1}{n_y}\right)}$ 6. $(\hat{p}_x - \hat{p}_y) \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \dfrac{\hat{p}_y(1-\hat{p}_y)}{n_y}}$ 7. $\dfrac{(n-1)s^2}{\chi^2_{\alpha/2}}$ *to* $\dfrac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$
Test Statistics: 1. $Z = \dfrac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$ or $Z = \dfrac{\bar{X} - \mu}{\dfrac{s}{\sqrt{n}}}$ 2. $T = \dfrac{\bar{X} - \mu}{\dfrac{s}{\sqrt{n}}}$ 3. $T = \dfrac{\bar{d} - \mu_d}{\dfrac{s_d}{\sqrt{n_d}}}$ 4. $Z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$	5. $Z = \dfrac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{\sigma_x^{\,2}}{n_x} + \dfrac{\sigma_y^{\,2}}{n_y}}}$ 6. $T = \dfrac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{S_p^{\,2}\left(\dfrac{1}{n_x} + \dfrac{1}{n_y}\right)}}$ 7. $T = \dfrac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{s_x^{\,2}}{n_x} + \dfrac{s_y^{\,2}}{n_y}}}$ with $df = \dfrac{\left(\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}\right)^2}{\dfrac{\left(\dfrac{s_x^2}{n_x}\right)^2}{n_x - 1} + \dfrac{\left(\dfrac{s_y^2}{n_y}\right)^2}{n_y - 1}}$

8. $Z = \dfrac{(\hat{p}_x - \hat{p}_y) - (p_x - p_y)}{\sqrt{\dfrac{p_x(1-p_x)}{n_x} + \dfrac{p_y(1-p_y)}{n_y}}}$	10. $\chi^2 = \dfrac{(n-1)s^2}{\sigma^2}$
	11. $F = \dfrac{s_1^2}{s_2^2}$
9. $Z = \dfrac{(\hat{p}_x - \hat{p}_y) - (p_x - p_y)}{\sqrt{\bar{p}(1-\bar{p})\left(\dfrac{1}{n_x} + \dfrac{1}{n_y}\right)}}$, where	12. $\chi^2 = \sum \dfrac{(O-E)^2}{E}, \quad E = np$

Nonparametric Statistics:

1. Sign Test:
   Small sample: $T =$ Number of (+) signs

   Large sample: $Z = \dfrac{2T - n}{\sqrt{n}}$

2. Wilcoxon Signed-rank:
   Small sample: $T = \sum R^+$, $T = \sum R^-$

   Large sample: $Z = \dfrac{T - \mu_T}{\sigma_T}$, $\mu_T = \dfrac{n(n+1)}{4}$, $\sigma_T = \sqrt{\dfrac{n(n+1)(2n+1)}{24}}$

3. Wilcoxon Rank Sum Test:
   Small sample: $U = R - \dfrac{n(n+1)}{2}$

   Large sample: $Z = \dfrac{T - \mu_T}{\sigma_T}$, $\mu_T = \dfrac{n_1(n_1 + n_2 + 1)}{2}$, $\sigma_T = \sqrt{\dfrac{n_1 n_2 (n+1)(n_1 + n_2 + 1)}{12}}$

- 000 O 000 -