
UNIVERSITI SAINS MALAYSIA

Second Semester Examination
Academic Session 2007/2008

April 2008

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 four** [4] questions.

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

1. (a) A store owner analyses his sales to get some information about the amount spent by each of his customers at each visit. The amount spent by a sample of his customers at a visit is given in the following table:

| Amount spent (RM X) | Frequency, f |
|------------------------|----------------|
| 0 – 9 | 8 |
| 10 – 19 | 10 |
| 20 – 29 | 7 |
| 30 – 39 | 5 |
| 40 – 49 | 3 |
| 50 – 59 | 2 |

$$\sum xf = 767.50, \quad \sum x^2 f = 24,298.75$$

(Give all answers in two decimal places.)

- (i) Find the mean and standard deviation of the amount spent by a customer at each visit.
 - (ii) What proportion of the customers spent between RM15 and RM35 at each visit?
 - (iii) Using Chebyshev's Theorem, obtain an interval of the amount spent by at least 45% of the customers.
- (b) Pairs of random numbers (x, y) are generated where the variables X and Y are integers between 1 and 6 inclusive. All outcomes are equally likely and a random variable W is defined as the absolute difference between X and Y .
- (i) Construct the probability distribution for W .
 - (ii) Calculate the mean and standard deviation of W .
 - (iii) Approximate the probability that a sample of 50 independent observations of W has a total sum of less than 110.
- (c) A real estate agent has 8 master keys to open several new homes. He selects 3 master keys at random before leaving the office. Only 1 master key will open any given house. If 40% of these homes are usually left unlocked, what is the probability that the real estate agent can get into a new home.

[100 marks]

1. (a) Seorang pemilik kedai menganalisis jualanannya untuk mendapatkan maklumat tentang amaun yang dibelanja oleh setiap pelanggan pada setiap kunjungan. Amaun yang dibelanja oleh suatu sampel pelanggan pada suatu kunjungan ialah seperti dalam jadual yang berikut:

| Amaun yang dibelanja (RM X) | Kekerapan, f |
|--------------------------------|----------------|
| 0 – 9 | 8 |
| 10 – 19 | 10 |
| 20 – 29 | 7 |
| 30 – 39 | 5 |
| 40 – 49 | 3 |
| 50 – 59 | 2 |

$$\sum xf = 767.50, \quad \sum x^2 f = 24,298.75$$

(Berikan semua jawapan dalam dua tempat perpuluhan.)

- (i) Dapatkan min dan sisihan piawai bagi amaun yang dibelanja oleh seorang pelanggan pada setiap kunjungan.
- (ii) Berapakah kadaran pelanggan yang membelanja antara RM15 dan RM35 pada setiap kunjungan.
- (iii) Dengan menggunakan Teorem Chebyshev, dapatkan suatu selang amaun yang dibelanja oleh sekurang-kurangnya 45% daripada pelanggan-pelanggan kedai tersebut.
- (b) Pasangan-pasangan nombor rawak (x, y) dijanakan di mana pembolehubah X dan Y adalah integer antara 1 dan 6 terangkum. Semua kesudahan adalah sama bolehjadi dan pembolehubah rawak W ditakrifkan sebagai beza mutlak antara X dan Y .
- (i) Bina taburan kebarangkalian bagi W .
- (ii) Hitung min W dan sisihan piawai W .
- (iii) Anggarkan kebarangkalian bahawa suatu sampel 50 cerapan rawak W mempunyai jumlah kurang daripada 110.
- (c) Seorang ejen hartanah mempunyai 8 kunci utama untuk membuka pintu beberapa rumah baru. Hanya satu kunci utama dapat membuka sebuah rumah tertentu. Jika 40% daripada rumah-rumah ini selalunya dibiarkan tidak berkunci, apakah kebarangkalian bahawa ejen hartanah tersebut dapat masuk ke dalam sebuah rumah baru?

[100 markah]

2. (a) Let A and B be nonempty, mutually exclusive events. Show that they must be dependent.
- (b) It is known that the life of a calculator manufactured by a company has a normal distribution with a mean of 54 months and a standard deviation of 8 months.
- The company guarantees that calculators that malfunction within c months of the date of purchase will be replaced. If the company is willing to replace at most 1% of all calculators sold, what should c be?
 - From past data, it is known that 5% of calculators manufactured by this company malfunction within two years of purchase. If the company recently sent 500 such calculators to its customers, what is the probability that more than 28 will be returned for replacement within a two-year period?
 - Determine the median life of calculators that malfunction within two years of purchase.
- (c) An educator claims that the passing rate for a major examination in rural areas of a state is at least 65%. The result from a recent study shows that 120 out of 200 candidates randomly sampled from rural high schools in the state, passed the major examination.
- At $\alpha = 0.05$, test whether there is enough evidence to reject the educator's claim.

Students' score in Mathematics is known to be normally distributed. The educator wishes to estimate the average score in Mathematics for the rural high school candidates in the state. She randomly samples 12 candidates and finds that their mean score in Mathematics is 60.5 with standard deviation 6.5.

- Construct a 95% confidence interval for the mean score in Mathematics for all rural high school candidates in the state.
- If the educator wishes to ensure that the probability is about 0.90 that her sample estimate of the mean score lies within two percentage points of the population mean, what sample size should she take?

[100 marks]

2. (a) Andaikan A dan B ialah peristiwa-peristiwa bukan kosong yang saling eksklusif. Tunjukkan bahawa A dan B semestinya bersandaran.
- (b) Hayat sebuah kalkulator yang dikeluarkan oleh sebuah syarikat diketahui tertabur secara normal dengan min 54 bulan dan sisihan piawai 8 bulan.
- (i) Syarikat tersebut memberi jaminan bahawa kalkulator yang tidak berfungsi dengan betul dalam tempoh c bulan selepas tarikh dibeli akan diganti. Jika syarikat tersebut sanggup mengganti sebanyak-banyaknya 1% daripada semua kalkulator yang dijual, apakah nilai c yang sepatutnya?
 - (ii) Daripada data lepas, diketahui bahawa 5% daripada kalkulator yang dihasilkan oleh syarikat tersebut tidak berfungsi dengan betul dalam tempoh dua tahun selepas dibeli. Jika syarikat tersebut baru-baru ini menghantar 500 buah kalkulator kepada pembelinya, apakah kebarangkalian bahawa lebih daripada 28 buah akan dikembalikan untuk diganti dalam tempoh dua tahun?
 - (iii) Hitung hayat median bagi kalkulator yang tidak berfungsi dengan betul dalam tempoh dua tahun selepas dibeli.
- (c) Seorang pendidik mendakwa bahawa kadar lulus dalam suatu peperiksaan utama di kawasan luar bandar sebuah negeri ialah sekurang-kurangnya 65%. Hasil daripada suatu kajian terkini menunjukkan bahawa 120 daripada 200 calon yang disampelkan secara rawak daripada sekolah-sekolah menengah luar bandar di negeri tersebut lulus peperiksaan tersebut.
- (i) Pada $\alpha = 0.05$, uji sama ada terdapat bukti yang cukup untuk menolak dakwaan pendidik tersebut.

Skor pelajar dalam Matematik diketahui tertabur secara normal. Pendidik tersebut ingin menganggar skor purata dalam Matematik bagi calon-calon di sekolah luar bandar di negeri tersebut. Ia mengambil suatu sampel rawak 12 orang calon dan mendapati bahawa skor purata mereka dalam Matematik ialah 60.5 dengan sisihan piawai 6.5.

- (ii) Bina suatu selang keyakinan 95% bagi min skor dalam Matematik bagi semua calon di sekolah luar bandar di negeri tersebut.
- (iii) Jika pendidik tersebut ingin memastikan dengan kebarangkalian 0.90 bahawa anggaran sampelnya mengenai min skor berada dalam sekitar dua peratus daripada min populasi, apakah saiz sampel yang patut diambilnya?

[100 markah]

3. (a) Two methods are proposed to determine whether a coin is biased. In the first method, the coin will be tossed 10 times and it will be considered biased if at least 8 heads or at least 8 tails are obtained.
- State a suitable null and alternative hypotheses for the problem.
 - State the critical region and determine the significance level of the test.
 - Determine the probability of making a type II error when the probability of obtaining a head on each toss is actually 0.6.

In the second method, the coin will be tossed 100 times and it will be considered biased if at least 60 heads or at least 60 tails are obtained.

- Approximate the significance level of the test.

- (b) The manager of bank *A* claims that its customer's waiting time to be served is shorter than customer's waiting time at bank *B*. Random samples of customer's waiting time at both banks result in the following summarized data:

| Bank <i>A</i> | Bank <i>B</i> |
|-------------------------------------|--------------------------------------|
| $n = 25$ | $n = 20$ |
| $\sum_{i=1}^{25} (x_i - 8) = 0$ | $\sum_{i=1}^{20} (y_i - 10) = 0$ |
| $\sum_{i=1}^{25} (x_i - 8)^2 = 165$ | $\sum_{i=1}^{20} (y_i - 10)^2 = 135$ |

Assume that the waiting times at both banks are normally distributed.

- At the 1% significance level, test whether there is a difference in the standard deviations of the customer's waiting time at both banks.
- Obtain an unbiased estimate of $\mu_{\bar{X}-\bar{Y}}$ and $\sigma_{\bar{X}-\bar{Y}}$.
- At the 1% significance level, test whether the claim made by the manager of bank *A* can be refuted.

3. (a) Dua kaedah dicadangkan untuk menentukan sama ada sekeping syiling adalah saksama atau tidak. Dalam kaedah pertama, syiling tersebut akan dilambung 10 kali dan akan dianggap tak saksama jika sekurang-kurangnya 8 kepala atau 8 bunga diperoleh.
- Nyatakan hipotesis nol dan hipotesis alternatif yang sesuai untuk masalah di atas
 - Nyatakan rantau genting dan tentukan aras keertian ujian.
 - Hitung kebarangkalian melakukan ralat jenis II apabila kebarangkalian memperoleh kepala pada setiap lambungan ialah sebenarnya 0.6.

Dalam kaedah kedua, syiling tersebut akan dilambung 100 kali dan akan dianggap tak saksama jika sekurang-kurangnya 60 kepala atau 60 bunga diperoleh.

- Anggarkan aras keertian ujian.

- (b) Pengurus bank A mendakwa bahawa masa menunggu untuk dilayan bagi pelanggannya adalah lebih pendek daripada masa menunggu bagi pelanggan bank B. Sampel-sampel rawak masa menunggu pelanggan di kedua-dua bank menghasilkan ringkasan data yang berikut:

| Bank A | Bank B |
|-------------------------------------|--------------------------------------|
| $n = 25$ | $n = 20$ |
| $\sum_{i=1}^{25} (x_i - 8) = 0$ | $\sum_{i=1}^{20} (y_i - 10) = 0$ |
| $\sum_{i=1}^{25} (x_i - 8)^2 = 165$ | $\sum_{i=1}^{20} (y_i - 10)^2 = 135$ |

Andaikan bahawa masa menunggu di kedua-dua bank tertabur secara normal.

- Pada aras keertian 1%, uji sama ada terdapat perbezaan antara sisihan piawai masa menunggu pelanggan di bank A dan di bank B.
- Dapatkan anggaran saksama bagi $\mu_{\bar{X}-\bar{Y}}$ dan $\sigma_{\bar{X}-\bar{Y}}$.
- Pada aras keertian 1%, uji sama ada dakwaan yang dibuat oleh pengurus bank A dapat disangkal.

- (c) A botanist examining the distribution of daisies in a field counts the number of daisies in 200 randomly chosen, non-overlapping, small areas, each one metre square. The results are given in the table below:

| | | | | | | | | | |
|------------------------|----|----|----|----|----|----|---|---|----------|
| Number of daisies, X | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | ≥ 8 |
| Number of squares | 37 | 49 | 49 | 32 | 15 | 12 | 5 | 1 | 0 |

- (i) Find the mean number of daisies in a one metre square area.
(ii) At the 5% significance level, decide whether the above result suggest that the distribution of daisies is Poisson.

[100 marks]

- (c) Seorang ahli botani yang memeriksa taburan bunga daisi di sebuah padang menghitung bilangan daisi dalam 200 petak kecil yang tak bertindih, setiap satu seluas satu meter persegi. Hasilnya diberikan dalam jadual yang berikut:

| | | | | | | | | | |
|---------------------------|----|----|----|----|----|----|---|---|----------|
| Bilangan bunga daisi, X | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | ≥ 8 |
| Bilangan petak | 37 | 49 | 49 | 32 | 15 | 12 | 5 | 1 | 0 |

- (i) Dapatkan bilangan purata bunga daisi dalam satu petak satu meter persegi.
- (ii) Pada aras keertian 5%, tentukan sama ada hasil di atas mencadangkan bahawa taburan bunga daisi tertabur secara Poisson.

[100 markah]

4. (a) An accounting firm has hired two temporary employees, A and B , to prepare individual federal income tax returns during the tax season. Clients who have relatively simple tax situations are randomly assigned to either A or B . The firm randomly selected 18 income tax returns prepared by each of these two employees and recorded the times taken to prepare these tax returns. After these times taken to prepare tax returns were ranked, the sum of the ranks for A was found to be 298 and the sum of the ranks for B was equal to 368. Based on these results, can you conclude that there is a difference in the median times taken to prepare such income tax returns by A and B ? Test at the 2.5% significance level.
- (b) The length of time required for a human subject to respond to a new drug was tested by a pharmacist. Ten 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 | 8 | 9 | 10 |
|----------|----|----|----|----|----|----|----|----|----|----|
| Aspirin | 15 | 20 | 12 | 20 | 17 | 14 | 17 | 16 | 19 | 18 |
| New Drug | 7 | 14 | 13 | 11 | 10 | 16 | 11 | 14 | 19 | 16 |

- (i) If the pharmacist wants to determine whether it takes longer to get pain relief with aspirin than with the new drug, which of the nonparametric tests could she use? (There could be more than one answer.)
- (ii) Using an appropriate test from your answer in part (i), can you conclude that the new drug gives pain relief faster than aspirin does? Test at the 5% significance level.
- (c) Explain
- (i) the conditions under which the binomial distribution $Bin(n, p)$ can be approximated by a Poisson distribution and a normal distribution.
- (ii) the meaning of a population distribution and a sampling distribution.
- (iii) the difference between parametric and nonparametric methods.

[100 marks]

4. (a) Sebuah firma perakaunan mengambil dua pekerja sementara, A dan B, untuk menyediakan pulangan cukai pendapatan individu semasa musim membayar cukai. Pelanggan yang mempunyai situasi cukai pendapatan yang agak mudah dirujuk kepada A atau B secara rawak. Firma itu memilih 18 pulangan cukai pendapatan yang disediakan oleh setiap pekerja ini dan merekodkan masa yang digunakan mereka untuk menyediakan pulangan-pulangan cukai. Setelah masa-masa yang digunakan untuk menyediakan pulangan cukai dipangkatkan, hasiltambah pangkat bagi A didapati ialah 298 dan hasiltambah pangkat bagi B bersamaan dengan 368. Berdasarkan hasil ini, bolehkah anda simpulkan bahawa terdapat perbezaan dalam masa median yang digunakan oleh A dan B untuk menyediakan pulangan cukai pendapatan? Uji pada aras keertian 2.5%.
- (b) Tempoh masa yang diperlukan oleh seorang subjek manusia untuk memberi tindak balas terhadap sejenis ubat baru diuji oleh seorang ahli farmasi. Sepuluh 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 menunjukkan kelegaan dicatatkan bagi kedua-dua jenis ubat. Data kajian ditunjukkan dalam jadual yang berikut.

| Subjek | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------|----|----|----|----|----|----|----|----|----|----|
| Aspirin | 15 | 20 | 12 | 20 | 17 | 14 | 17 | 16 | 19 | 18 |
| Ubat baru | 7 | 14 | 13 | 11 | 10 | 16 | 11 | 14 | 19 | 16 |

- (i) Jika ahli farmasi tersebut ingin menentukan sama ada tempoh masa yang diperlukan untuk mendapat kelegaan daripada aspirin adalah lebih lama daripada ubat baru tersebut, apakah ujian tak berparameter yang boleh ia gunakan? (Kemungkinan terdapat lebih daripada satu jawapan.)
- (ii) Dengan menggunakan ujian yang sesuai daripada jawapan anda dalam bahagian (i), bolehkah anda simpulkan bahawa ubat baru tersebut memberi kelegaan dengan lebih cepat daripada aspirin? Uji pada aras keertian 5%.
- (c) Terangkan
- (i) keadaan-keadaan di mana taburan binomial $\text{Bin}(n, p)$ dapat dihampiri dengan suatu taburan Poisson dan suatu taburan normal.
- (ii) makna taburan populasi dan taburan pensampelan.
- (iii) perbezaan antara kaedah berparameter dan kaedah tak berparameter.

[100 markah]

FORMULA

| | | |
|---|---|--|
| $\bar{x} = \frac{\sum xf}{\sum f}$ $s^2 = \frac{\sum (x^2 f) - \frac{(\sum xf)^2}{\sum f}}{\sum f - 1}$ | $S_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$ $\bar{p} = \frac{X + Y}{n_x + n_y}$ | |
| <p>Confidence Intervals:</p> $\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ $\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$ $\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $\frac{(n-1)s^2}{\chi_{\alpha/2}^2} \text{ to } \frac{(n-1)s^2}{\chi_{1-\alpha/2}^2}$ | $(\bar{X} - \bar{Y}) \pm z_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$ $(\bar{X} - \bar{Y}) \pm t_{\alpha/2} \sqrt{S_p^2 \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}$ $(\hat{p}_x - \hat{p}_y) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}$ | |
| <p>Test Statistics:</p> $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ $T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$ $T = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n_d}}}$ $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$ | $Z = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$ $T = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{S_p^2 \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}}$ $Z = \frac{(\hat{p}_x - \hat{p}_y) - (p_x - p_y)}{\sqrt{\frac{p_x(1-p_x)}{n_x} + \frac{p_y(1-p_y)}{n_y}}}$ | $Z = \frac{(\hat{p}_x - \hat{p}_y) - (p_x - p_y)}{\sqrt{\bar{p}(1-\bar{p}) \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}}$ $F = \frac{s_1^2}{s_2^2}$ $\chi^2 = \sum \frac{(O - E)^2}{E}, \quad E = np$ |

Nonparametric Statistics:

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

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

Wilcoxon Rank Sum Test: $T = \sum R$

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

- 000 O 000 -