
UNIVERSITI SAINS MALAYSIA

Second Semester Examination
2011/2012 Academic Session

June 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. A small grocery store owner analyses his sales to get some information about the amount spent by a customer at each visit. The amount spent by a sample of his customers at a visit is shown in the following table:

Amount spent (RM)	Frequency
0 – 9	5
10 – 19	8
20 – 29	12
30 – 39	15
40 – 49	11
50 – 59	6

$$\sum_{i=1}^k x_i = 766.5, \quad \sum_{i=1}^k x_i^2 = 66444.25$$

- Calculate the mean and standard deviation of the amount spent by a customer at a visit.
- What proportion of the customers spent more than RM37 at a visit?
- Determine the mode and median of the above data, and hence comment on the symmetry of the distribution.

[20 marks]

1. Seorang pemilik kedai runcit menganalisis jualannya untuk mendapatkan sedikit maklumat tentang jumlah yang dibelanja oleh seorang pelanggan pada setiap kunjungan. Jumlah yang dibelanja oleh suatu sampel pelanggan pada satu kunjungan ditunjukkan dalam jadual yang berikut:

Jumlah belanja (RM)	Kekerapan
0 – 9	5
10 – 19	8
20 – 29	12
30 – 39	15
40 – 49	11
50 – 59	6

$$\sum_{i=1}^k x_i = 766.5, \quad \sum_{i=1}^k x_i^2 = 66444.25$$

- Hitung min dan sisihan piawai bagi jumlah yang dibelanja oleh seorang pelanggan pada setiap kunjungan.
- Berapakah kadar pelanggan yang membelanja lebih daripada RM37 pada satu kunjungan?
- Tentukan mod dan median bagi data di atas dan seterusnya, komen tentang simetri taburannya.

[20 markah]

2. Pairs of random numbers (x, y) are generated where the variables X and Y are integers between 2 and 6 inclusive. All outcomes are equally likely and a random variable W is defined as the absolute difference between X and Y .

- (a) Construct the probability distribution for W .
- (b) Calculate the mean and standard deviation of W .
- (c) Approximate the probability that a sample of 50 independent observations of W has a total sum of less than 120.

[15 marks]

2. Pasangan-pasangan nombor rawak (x, y) dijanakan yang mana pembolehubah X dan Y adalah integer antara 2 dan 6 terangkum. Semua kesudahan adalah sama boleh jadi dan pembolehubah rawak W ditakrifkan sebagai beza mutlak antara X dan Y .

- (a) Bina taburan kebarangkalian bagi W .
- (b) Hitung min dan sisihan piawai bagi W .
- (c) Anggarkan kebarangkalian bahawa suatu sampel 50 cerapan W yang tak bersandar mempunyai jumlah kurang daripada 120.

[15 markah]

3. A real estate agent has 10 master keys to open several new homes for prospective buyers to see. Only 1 master key will open any given new home which is locked and 40% of these new homes are usually left unlocked. If the agent randomly selects 3 master keys before leaving his office, what is the probability that he can open a new home for a prospective buyer to see?

[15 marks]

3. Seorang ejen hartanah mempunyai 10 kunci utama untuk membuka beberapa rumah baru untuk dilihat oleh bakal pembeli. Hanya satu kunci utama dapat membuka sesebuah rumah baru yang berkunci dan 40% daripada rumah-rumah baru ini selalunya dibiarkan tidak berkunci. Jika ejen tersebut memilih 3 kunci utama secara rawak sebelum meninggalkan pejabatnya, apakah kebarangkalian bahawa ia dapat membuka pintu sebuah rumah baru untuk dilihat oleh bakal pembeli?

[15 markah]

4. Suppose that A and B are non-empty, mutually exclusive events. Show that A and B must be dependent.

[10 marks]

4. Andaikan bahawa A dan B ialah peristiwa tak kosong dan saling eksklusif. Tunjukkan bahawa A dan B semestinya bersandar.

[10 markah]

5. The operational lifetimes of brand A calculators are normally distributed with mean 84 months. It is also known that 20% of these calculators have lifetimes exceeding 100 months.
- Calculate the standard deviation of the lifetimes. Give your answer correct to the nearest month.
 - Use an appropriate approximation to calculate the probability that in a random sample of 100 such calculators, at least 20 of them have lifetimes less than 100 months.
 - Suppose that the manufacturer of the calculator would like to offer a guarantee period so that calculators that malfunction within c months from the date of purchase can be returned for replacement without cost. If he is prepared to replace at most 0.5% of all calculators sold, what should the value of c be?
 - The operational lifetime of brand B calculators is Y months, where Y is a continuous random variable having probability density function f defined as follows:

$$f(y) = \begin{cases} \frac{2}{100}(90-y) & \text{for } 80 \leq y \leq 90 \\ 0 & \text{otherwise} \end{cases}$$

Between brand A and brand B calculators, which has (i) the larger mean lifetime? (ii) lifetimes with larger standard deviation?

[40 marks]

5. Masa hayat operasi mesinkira jenama A tertabur secara normal dengan min 84 bulan. Juga diketahui bahawa 20% daripada mesinkira ini mempunyai masa hayat melebihi 100 bulan,
- Hitung sisihan piawai masa hayatnya. Beri jawapan yang dibetulkan ke bulan terdekat.
 - Guna penghampiran yang sesuai untuk menghitung kebarangkalian bahawa dalam suatu sampel rawak 100 buah mesinkira tersebut, sekurang-kurangnya 20 buah daripadanya mempunyai masa hayat kurang dari 100 bulan.
 - Andaikan pengeluar mesinkira tersebut ingin menawarkan suatu tempoh jaminan supaya mesinkira yang tidak berfungsi dalam tempoh c bulan selepas tarikh pembelian dapat dipulangkan untuk diganti tanpa kos. Jika ia bersedia untuk menggantikan sebanyak-banyaknya 0.5% daripada semua mesinkira yang dijual, apakah nilai yang patut bagi c ?
 - Masa hayat operasi mesinkira jenama B ialah Y bulan, yang mana Y ialah pemboleh ubah rawak selanjar yang mempunyai fungsi taburan kebarangkalian f yang ditakrifkan seperti berikut:

$$f(y) = \begin{cases} \frac{2}{100}(90-y) & \text{for } 80 \leq y \leq 90 \\ 0 & \text{sebaliknya} \end{cases}$$

Di antara mesinkira jenama A dan mesinkira jenama B, jenama manakah yang mempunyai (i) min masa hayat yang lebih besar? (ii) masa hayat dengan sisihan piawai yang lebih besar?

[40 markah]

6. An educator claims that the passing rate for a high school major examination in rural areas is less than 65%. The result from a recent study shows that 120 out of 200 candidates who were randomly sampled from high schools in the rural area, passed the major examination.

- (a) At $\alpha = 0.05$, test whether there is enough evidence to reject the educator's claim.

Students' score in Mathematics in the major examination is known to be normally distributed and the educator wishes to estimate the average score in Mathematics for candidates in the rural area. She randomly selects a sample of 15 candidates and finds that their mean score in Mathematics is 60.5 with standard deviation 6.5.

- (b) Construct a 95% confidence interval for the mean score in Mathematics for of all candidates in the rural area.
- (c) If the educator desires a 0.90 probability that her sample estimate of the mean score lies within two units of the true value, what sample size should she take?

[30 marks]

6. Seorang pendidik mendakwa bahawa kadar lulus bagi suatu peperiksaan utama sekolah menengah di luar bandar adalah kurang daripada 65%. Hasil suatu kajian terkini menunjukkan bahawa 120 daripada 200 calon yang dipilih secara rawak daripada sekolah-sekolah menengah di luar bandar, telah lulus peperiksaan tersebut.

- (a) Pada $\alpha = 0.05$, uji sama ada terdapat bukti yang cukup untuk menolak dakwaan pendidik tersebut.

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

- (b) Bina suatu selang keyakinan 95% bagi min skor dalam Matematik bagi semua calon di kawasan luar bandar tersebut.
- (c) Jika pendidik tersebut inginkan kebarangkalian 0.90 bahawa anggarannya mengenai min skor berada dalam sekitar dua unit daripada nilainya yang sebenar, apakah saiz sampel yang patut diambilnya?

[30 markah]

7. A bag contains a very large number of identical marbles of different colours. Of these, an unknown proportion p are red marbles. Suppose you are required to test the null hypothesis $H_0 : p = 0.3$ against the alternative $H_1 : p < 0.3$. To perform the test, you are to take a random sample of 100 marbles and note X , the number of red marbles in the sample. Assume X is normally distributed.
- If the significance level of the test is 5%, determine the critical region in the form of $0 \leq X \leq c$, where c is an integer.
 - Calculate the probability of making a type II error when the proportion of red marbles in the bag is actually 0.2.

[20 marks]

7. Sebuah beg mengandungi sejumlah besar guli yang serupa tetapi berbeza warna. Daripada jumlah ini, suatu kadaran p yang tidak diketahui adalah berwarna merah. Andaikan anda dikehendaki menguji hipotesis nol $H_0 : p = 0.3$ berlawanan alternatif $H_1 : p < 0.3$. Untuk menjalankan ujian ini, anda perlu mengambil suatu sampel rawak 100 biji guli dan mencatat nilai X , iaitu bilangan guli merah dalam sampel. Andaikan X tertabur secara normal.
- Jika aras keertian ujian ialah 5%, tentukan kawasan genting dalam bentuk $0 \leq X \leq c$, yang mana c ialah suatu integer.
 - Hitung kebarangkalian melakukan ralat jenis II apabila kadaran guli merah di dalam beg tersebut ialah sebenarnya 0.2.

[20 markah]

8. A firm is presently using supplier A for its raw materials and will continue to do so provided that the mean delivery times are the same as or less than those of supplier B. If the firm finds otherwise, it will begin making purchases from supplier B. An independent sample that shows the delivery time characteristics of the two suppliers are as follows:

Supplier A	Supplier B
10, 10, 11, 12, 13, 13, 14, 14, 14, 15, 15, 16, 17, 18, 18	10, 10, 10, 11, 11, 12, 15, 15, 18, 18
$\bar{x}_A = 14$ days	$\bar{x}_B = 13$ days
$s_A = 3$ days	$s_B = 3.1$ days

Assume that the delivery times for both suppliers are normally distributed.

- At the 5% significance level, test whether there is a difference in the standard deviations of the delivery times for the two suppliers.
- Construct a 95% confidence interval for the mean difference in the mean delivery times between supplier A and supplier B. Based on the confidence interval, should the firm stay with supplier A or switch to supplier B for its raw materials?
- Suppose that the normality assumption cannot be made. Use the Wilcoxon signed-rank test to determine whether the firm should stay with supplier A or switch to supplier B for its raw materials.

[40 marks]

8. Pada masa ini, sebuah firma menggunakan pembekal A untuk bahan mentahnya dan akan terus menggunakannya dengan syarat min masa penghantarannya adalah sama atau kurang daripada min masa pembekal B. Jika firma tersebut mendapati hal sebaliknya, ia akan mula membeli daripada pembekal B. Suatu sampel tak bersandar yang menunjukkan ciri-ciri masa penghantaran kedua-dua pembekal adalah seperti berikut:

Pembekal A	Pembekal B
10, 10, 11, 12, 13, 13, 14, 14, 14, 15, 15, 16, 17, 18, 18	10, 10, 10, 11, 11, 12, 15, 15, 18, 18
$\bar{x}_A = 14$ hari	$\bar{x}_B = 13$ hari
$s_A = 2.5$ hari	$s_B = 3.1$ hari

Andaikan bahawa masa penghantaran bagi kedua-dua pembekal tertabur secara normal

- Pada aras keertian 5%, uji sama ada terdapat perbezaan antara sisihan piawai masa penghantaran pembekal A dan sisihan piawai masa penghantaran pembekal B.
- Bina suatu selang keyakinan 95% bagi perbezaan min masa penghantaran antara pembekal A dan pembekal B. Berdasarkan selang keyakinan tersebut, patutkah firma tersebut terus menggunakan pembekal A atau bertukar kepada pembekal B untuk bahan mentahnya?
- Andaikan anggapan kenormalan tidak dapat dibuat. Jalankan ujian pangkat bertanda Wilcoxon untuk menentukan sama ada firma tersebut patut terus menggunakan pembekal A atau bertukar kepada pembekal B untuk bahan mentahnya.

[40 markah]

9. A shop that repairs computers keeps a record of X , the number of computers brought in for repair each day. The numbers brought in during a random sample of 40 days are as follows:

```

4  0  0  0  2  1  1  1  0  1
0  1  1  0  3  0  1  1  1  1
3  1  1  1  0  0  2  0  1  1
0  0  0  1  1  1  0  2  0  2

```

- Group the data in a one-way table and calculate its mean and standard deviation. State why the results might suggest that X has a Poisson distribution.
- Perform a test, at the 5% significance table to determine whether the data could be a sample from a Poisson distribution.

[35 marks]

9. Sebuah kedai yang membaiki komputer menyimpan rekod X , iaitu bilangan komputer yang dibawa ke kedainya untuk dibaiki setiap hari. Bilangan yang dibawa bagi suatu sampel rawak 40 hari ialah seperti yang berikut:

4 0 0 0 2 1 1 1 0 1
 0 1 1 0 3 0 1 1 1 1
 3 1 1 1 0 0 2 0 1 1
 0 0 0 1 1 1 0 2 0 2

- (a) Kumpulkan data di atas dalam jadual satu-hala dan hitung min dan sisihan piawainya. Nyatakan kenapa hasilnya mencadangkan bahawa X mempunyai taburan Poisson.
 (b) Pada aras keertian 5%, uji bahawa angka-angka di atas ialah cerapan cerapan daripada suatu taburan Poisson.

[35 markah]

10. 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

Can you conclude that it takes longer to get pain relief with aspirin than with the new drug? Use an appropriate non-parametric test at the 5% significance level.

[25 marks]

10. 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

Bolehkan anda simpulkan bahawa tempoh masa yang lebih lama diambil untuk mendapat kelegaan menggunakan aspirin berbanding ubat baru tersebut? Jalankan suatu ujian tak berparameter yang sesuai pada aras keertian 5%.

[25 markah]

FORMULA

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

Nonparametric Statistics:

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

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

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

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