

UNIVERSITI SAINS MALAYSIA

First Semester Examination
2016/2017 Academic Session

December 2016/January 2017

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 frequency distribution shows the blood glucose levels (in milligrams per deciliter) for 50 patients at a medical facility.

Blood Glucose Level	Frequency, f_i
60 - 64	2
65 - 69	1
70 - 74	5
75 - 79	12
80 - 84	18
85 - 89	6
90 - 94	5
95 - 99	1

Given that $\sum_{i=1}^k x_i f_i = 4030$ and $\sum_{i=1}^k x_i^2 f_i = 327,420$

- (a) Calculate the mean of the blood glucose level and its standard deviation.
- (b) The patients with glucose level exceed 87 mg/dl will be admitted in the hospital. What is the percentage of the patients who will be admitted in the hospital?
- (c) Use Chebyshev's Theorem to obtain an interval of the blood glucose levels of at least 75% of the patients at a medical facility.

[20 marks]

1. *Taburan frekuensi menunjukkan tahap glukosa darah (dalam milligram per desiliter) bagi 50 pesakit di sebuah pusat perubatan.*

Tahap Glukosa Darah	Frekuensi, f_i
60 - 64	2
65 - 69	1
70 - 74	5
75 - 79	12
80 - 84	18
85 - 89	6
90 - 94	5
95 - 99	1

Diberikan bahawa $\sum_{i=1}^k x_i f_i = 4030$ dan $\sum_{i=1}^k x_i^2 f_i = 327,420$

- (a) *Kira min bagi tahap glukosa darah dan sisihan piawainya.*
- (b) *Pesakit dengan tahap glukosa melebihi 87 mg/dl akan dimasukkan ke hospital. Berapakah peratusan pesakit yang akan dimasukkan ke hospital?*
- (c) *Guna Teorem Chebyshev bagi mendapatkan selang bagi tahap glukosa darah untuk sekurang-kurangnya 75% pesakit di pusat perubatan tersebut.*

[20 markah]

2. The following table shows the sentiments of 2500 wage-earning employees at Company XYZ on a proposal to emphasize fringe benefits rather than wage increases during their impending contract discussions.

Employee	Opinion			Total
	Favor	Neutral	Opposed	
Male	800	200	500	1500
Female	400	100	500	1000
Total	1200	300	1000	2500

- (a) Calculate the probability that an employee selected at random from this group will oppose the proposal.
- (b) Calculate the probability that an employee selected at random from this group will be female.
- (c) Calculate the probability that an employee selected at random from this group will be opposed, given that the person is male.
- (d) Are the events “opposed” and “female” independent? Explain.

[20 marks]

2. Jadual di bawah menunjukkan sentimen bagi 2500 pekerja berkaitan cadangan untuk memberi keistimewaan pekerja berbanding kenaikan gaji semasa perbincangan kontrak.

Pekerja	Pendapat			Jumlah
	Setuju	Neutral	Tidak Setuju	
Lelaki	800	200	500	1500
Perempuan	400	100	500	1000
Jumlah	1200	300	1000	2500

- (a) Kira kebarangkalian pekerja yang dipilih secara rawak daripada kumpulan ini tidak setuju dengan cadangan tersebut.
- (b) Kira kebarangkalian pekerja yang dipilih secara rawak daripada kumpulan ini merupakan perempuan.
- (c) Kira kebarangkalian pekerja yang dipilih secara rawak daripada kumpulan ini berpendapat tidak setuju, diketahui bahawa pekerja tersebut adalah lelaki.
- (d) Adakah peristiwa “Tidak Setuju” dan “Perempuan” tidak bersandar? Jelaskan.

[20 markah]

3. The probability density function of a continuous variable X is given by

$$f(x) = \begin{cases} \frac{x+1}{6}, & 1 < x < 3 \\ 0, & \text{otherwise} \end{cases}$$

- (a) Find $E(X)$ and $\text{Var}(X)$.
- (b) The median for a continuous probability distribution 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 $f(x)$.

[20 marks]

3. *Fungsi ketumpatan kebarangkalian bagi pemboleh ubah selanjar X dinyatakan sebagai*

$$f(x) = \begin{cases} \frac{x+1}{6}, & 1 < x < 3 \\ 0, & \text{selainnya} \end{cases}$$

- (a) *Cari $E(X)$ dan $\text{Var}(X)$.*
- (b) *Median bagi taburan kebarangkalian selanjar ialah nilai m yang membahagikan taburan kepada dua kawasan yang sama iaitu*

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

Dapatkan median bagi $f(x)$.

[20 markah]

4. In a certain area, the probability of a randomly selected cow dying from 'mad cow' disease is 0.04.

- (a) Calculate the probability that in a random sample of 18 cows exactly 2 will die from the disease.
- (b) Find the probability that in a random sample of 50 cows between 3 and 5 (inclusive) will die from the disease. [Use exact distribution]
- (c) Use a distributional approximation to find the probability that in a random sample of 300 cows, fewer than 12 will die from the disease.

[20 marks]

4. Di dalam sebuah kawasan, kebarangkalian seekor lembu yang dipilih secara rawak mati disebabkan penyakit lembu gila adalah 0.04.

- (a) Kira kebarangkalian dalam sampel rawak bagi 18 ekor lembu, terdapat 2 ekor yang akan mati disebabkan penyakit.
- (b) Cari kebarangkalian dalam sampel rawak sebanyak 50 ekor lembu, di antara 3 hingga 5 (termasuk) akan mati disebabkan penyakit. [Gunakan taburan tepat]
- (c) Guna taburan penghampiran bagi mencari kebarangkalian dalam sampel rawak 300 ekor lembu, kurang daripada 12 ekor akan mati disebabkan penyakit.

[20 markah]

5. A random variable Y has a normal distribution with mean μ and variance σ^2 . Given that 10% of the values of Y exceed 17.24 and that 25% of the values of Y are less than 14.37. Find the values of μ and σ^2 .

[20 marks]

5. *Pemboleh ubah rawak Y mempunyai taburan Normal dengan min μ dan varians σ^2 . Diberikan bahawa 10% daripada nilai Y melebihi 17.24 dan 25% daripada nilai Y kurang daripada 14.37. Dapatkan nilai μ dan σ^2 .*

[20 markah]

6. It is claimed that the performance of the first year undergraduate students in Statistics course in one university can be studied in their ratios of grades. The ratios of distributions to grade A, B and C are 3:5:2 respectively. A total of 150 students are surveyed and the survey reveals the following result:

Grade A	Grade B	Grade C
51	81	18

- (a) If the null hypothesis (3:5:2) is true, what are the expected number of students that score A, B and C respectively in Statistics?
- (b) Complete the hypothesis test using the 0.01 significance level.

[20 marks]

6. *Terdapat dakwaan yang menyatakan bahawa prestasi pelajar ijazah tahun pertama dalam statistik boleh dikaji dari nisbah gred mereka. Taburan untuk gred A, B dan C adalah dalam nisbah 3:5:2. Seramai 150 mahasiswa ditinjau dan hasil tinjauan adalah seperti yang berikut:*

Gred A	Gred B	Gred C
51	81	18

- (a) *Jika hipotesis nol (3:5:2) adalah benar, berapakah bilangan pelajar yang mendapat gred A, B dan C?*
- (b) *Lengkapkan pengujian hipotesis dengan menggunakan aras keertian 0.01.*

[20 markah]

...6/-

7. A scientist is studying the paramecium, a one-celled organism, under a microscope. There are 1500 paramecia in the slide he is studying, and the standard deviation of their lengths is 0.12mm. He views a sample of 30 paramecia and finds that the mean length of these 30 specimens is 0.25mm.
- What is the point estimate of the mean length of the entire paramecium population in the slide?
 - Construct a 95% confidence interval for the mean length of the population.
 - What is the maximum error for the estimate?

[20 marks]

7. *Seorang saintis mengkaji paramesium, organisma satu-sel di bawah mikroskop. Sebanyak 1500 paramesia dalam slaid yang dikaji dan sisihan piawai panjangnya ialah 0.12mm. Beliau mengambil sampel 30 paramesia dan mendapati min panjang bagi 30 spesimen tersebut ialah 0.25mm.*

- Apakah anggaran titik bagi min panjang bagi seluruh populasi paramesium di dalam slaid?*
- Bina selang keyakinan 95% bagi panjang min populasi.*
- Apakah ralat maksimum bagi anggaran tersebut?*

[20 markah]

8. (a) Briefly describe Type I and Type II errors.
- (b) In a poll of college athletes in 2003, 76% of the athletes declared that their favourite brand of shoe was Brand X. The manufacturer of Brand X is interested to know whether this is still the case. Therefore, 120 athletes are polled, and 73.3% declared that they prefer Brand X. The manufacturer then performs a hypothesis test at the 10% level of significance. What is the manufacturer's conclusion?

[20 marks]

8. (a) *Terangkan secara ringkas ralat jenis I dan ralat jenis II.*
- (b) *Dalam undian atlet kolej pada 2003, 76% daripada atlet menyatakan jenama kasut kegemaran mereka adalah Jenama X. Pengeluar Jenama X berminat untuk mengetahui sama ada kegemaran mereka masih sama. Oleh yang demikian, daripada 120 atlet mengundi, 73.3% mengistiharkan bahawa mereka gemar Jenama X. Pengeluar kemudiannya menjalankan ujian hipotesis pada aras keertian 10%. Apakah kesimpulan pengeluar?*

[20 markah]

9. A clinic tested two medicines on patients to find the relief time from pain. Brand A medicine was given to 22 patients, and it produced a mean relief time of 14 minutes with a standard deviation of 4 minutes. Brand B medicine was given to 25 patients, and it produced a mean relief time of 12 minutes with a standard deviation of 3 minutes. Assume that the relief times for both medicines for all patients have normal distributions with equal standard deviations.
- (a) Construct a 95% confidence interval for the difference between the mean relief times for these two medicines.
- (b) Test at the 5% significance level if the mean relief times for two brand of medicines are different.

[20 marks]

9. *Sebuah klinik menguji tempoh kesan dua jenis ubat penahan sakit terhadap pesakit-pesakitnya. Ubat jenama A diberikan kepada 22 pesakit dan min masa kelegaan adalah 14 minit dengan sisihan piawai 4 minit. Ubat jenama B pula diberikan kepada 25 pesakit, dan min masa kelegaan adalah 12 minit dengan sisihan piawai 3 minit. Andaikan masa kelegaan bagi kedua-dua jenis ubat terhadap semua pesakit bertaburan normal dengan sisihan piawai yang sama.*

- (a) *Bina selang keyakinan 95% bagi perbezaan min masa kelegaan untuk kedua-dua jenis ubat.*
- (b) *Uji pada aras keertian 5% sama ada min masa kelegaan bagi kedua-dua jenama ubat adalah berbeza.*

[20 markah]

10. Jenifer has decided to change the girls physical education at Central High School to include a volunteer weight reduction and fitness program for the high school seniors who want to reduce their body weight. The 15 girls who signed up for the program weighed in at the onset and again one month later. The data of the 15 girls involved in this program are given in table below:

Weight Before	Weight After	Weight Before	Weight After	Weight Before	Weight After
146	140	135	136	151	150
159	151	178	170	132	131
139	132	146	141	149	141
175	165	153	150	129	129
152	143	166	168	139	133

Using the 5% level of significance, can we conclude that the new physical education program helped reduced the median body weight of high school senior females?

[20 marks]

10. Jenifer membuat keputusan untuk mengubah pendidikan jasmani pelajar perempuan di sekolah menengah bagi memasukkan sukarelawati dalam program kecergasan dan penurunan berat badan dalam kalangan pelajar senior yang ingin menurunkan berat badan mereka. Sebanyak 15 pelajar perempuan menyertai program ini dan berat badan mereka diambil pada masa program bermula dan selepas sebulan kemudian. Data 15 pelajar perempuan yang terlibat dalam program ini diberikan dalam jadual di bawah:

<i>Berat Sebelum</i>	<i>Berat Selepas</i>	<i>Berat Sebelum</i>	<i>Berat Selepas</i>	<i>Berat Sebelum</i>	<i>Berat Selepas</i>
146	140	135	136	151	150
159	151	178	170	132	131
139	132	146	141	149	141
175	165	153	150	129	129
152	143	166	168	139	133

Guna aras keertian 5%, bolehkah disimpulkan bahawa program pendidikan jasmani baru membantu dalam menurunkan median berat badan pelajar senior perempuan sekolah menengah?

[20 markah]

FORMULA

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

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

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