

---

UNIVERSITI SAINS MALAYSIA

Second Semester Examination  
2009/2010 Academic Session

April/May 2010

**MGM 561 – Statistical Methods for Research**  
***[Kaedah Statistik untuk Penyelidikan]***

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 eight** [8] questions.

**Arahan:** Jawab **semua lapan** [8] 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. Consider the following questions:-

(a) A random sample of 25 observations,  $X$ , gave the following results

$$\sum x = 2500, \sum x^2 = 250096$$

Calculate the standard error of the sample mean.

(b) The sizes of claims under a certain type of policy are distributed normally about a mean of RM 1800 and with a standard deviation of RM 400. Find the probability that the sizes of two randomly selected claims differ by more than RM 500.

[10 marks]

2. Consider the following questions:-

(a) The claim amounts (RM, to the nearest RM 10) for a random sample of 23 recent claims for storm damage to private houses in a particular town are as follows:

2810 3230 4170 670 2940 5410 2390 3520 2160 4670 3880 1240  
1490 6780 3710 1590 4630 3110 2510 8320 3980 920 2530

(i) Construct a stem-and-leaf display of the claim amounts.

(ii) Find the interquartile range of the claim amounts.

(b) The table below shows a grouped frequency distribution for 100 claim amounts on a certain class of insurance policy.

<i>Claim Amount(RM)</i>	<i>Frequency</i>
under 100	4
100–149.99	10
150–199.99	25
200–249.99	30
250–299.99	15
300–349.99	12
350–399.99	4
400 or over	0

Determine approximate values for the median and the interquartile range of these claims amounts.

[10 marks]

1. Pertimbangkan soalan-soalan berikut:-

(a) Diberi suatu sampel rawak dengan 25 cerapan,  $X$ , memberi keputusan berikut

$$\sum x = 2500, \sum x^2 = 250096$$

Kira ralat piawai bagi min sampel tersebut.

(b) Saiz-saiz tuntutan di bawah suatu jenis polisi adalah bertaburan normal dengan min RM 1800 dan sisihan piawai RM 400. Cari kebarangkalian bahawa saiz –saiz bagi dua tuntutan yang dipilih secara rawak berbeza lebih daripada RM 500.

[10 markah]

2. Pertimbangkan soalan-soalan berikut:-

(a) Amaun-amaun tuntutan (RM, sehingga angka terdekat RM 10) bagi suatu sampel rawak 23 tuntutan terkini untuk kerosakan bencana kilat (ribut) terhadap rumah-rumah persendirian dalam suatu Bandar diberikan seperti berikut:

2810 3230 4170 670 2940 5410 2390 3520 2160 4670 3880 1240  
1490 6780 3710 1590 4630 3110 2510 8320 3980 920 2530

(i) Bina plot “tangkai-dan-daun” bagi amaun tuntutan

(ii) Cari julat antara kuartil bagi amaun tuntutan tersebut

(b) Jadual di bawah menunjukkan suatu kumpulan kekerapan bagi taburan 100 amaun tuntutan bagi sesuatu kelas polisi insurans

Amaun Tuntutan (RM)	Kekerapan
Bawah 100	4
100–149.99	10
150–199.99	25
200–249.99	30
250–299.99	15
300–349.99	12
350–399.99	4
400 ke atas	0

Kira nilai-nilai anggaran bagi median dan julat antara kuartil (interquartile range) bagi amaun-amaun tuntutan tersebut.

[10 markah]

3. An insurance company has clients for its household contents policies in two areas A and B. The company believes that the average claim amount in area B is lower than that in area A. A sample of twelve houses for which claims have arisen is selected at random from each of the two areas. The claim amounts are as follows:

House No.	1	2	3	4	5	6	7	8	9	10	11	12
Area A <i>x</i> :	165	185	195	200	205	210	215	220	225	235	245	263
Area B <i>y</i> :	148	125	105	160	157	167	184	200	203	217	243	265

Given,  $\sum x = 2563$ ,  $\sum x^2 = 555269$ ,  $\sum y = 2174$  and  $\sum y^2 = 418120$

- Perform a test to investigate whether there is evidence of a difference between the variability of claim amounts within the two areas.
- Perform a *t*-test to investigate whether the average claim amount for area B is lower than that for area A by an amount which exceeds RM 15, and comment on the validity of such a test.

[10 marks]

4. Consider the following questions:-

- Independent random samples of size  $n_1 = 10$  and  $n_2 = 25$ , respectively, are taken from normal distributions with the same variance. The sample variances are  $S_1^2$  and  $S_2^2$ , respectively. What is the value of  $k$  such that  $P\left(\frac{S_1^2}{S_2^2} > k\right) = 0.01$ ?
- The amounts (RM) of pocket-money received by a random sample of 100 of 13-year old girls had sample mean RM 4.56 and sample standard deviation RM 1.21. Calculate a 95% confidence interval for the mean amount received in the population of all such girls.

[10 marks]

5. Consider the following questions:-

- A sample of 12 paired observations  $x_i, y_i$  yielded a sample correlation coefficient,  $r = 0.91$ . Test the population correlation coefficient,  $\rho$  for the following hypothesis:-

$$H_0 : \rho = 0.8 \text{ versus } H_1 : \rho > 0.8$$

3. Sebuah syarikat insurans mempunyai pelanggan-pelanggan yang menduduki rumah di dua kawasan, iaitu kawasan A dan B. Syarikat ini mempercayai bahawa purata tuntutan di kawasan B lebih rendah daripada purata tuntutan bagi kawasan A. Suatu sampel dua belas buah rumah yang mana tuntutan itu telah dikemukakan, dipilih secara rawak dari dua kawasan tersebut. Amaun-amaun tuntutan diberikan seperti berikut:

Rumah No.	1	2	3	4	5	6	7	8	9	10	11	12
Kawasan A x:	165	185	195	200	205	210	215	220	225	235	245	263
Kawasan B y:	148	125	105	160	157	167	184	200	203	217	243	265

Diberi,  $\sum x = 2563$ ,  $\sum x^2 = 555269$ ,  $\sum y = 2174$  dan  $\sum y^2 = 418120$

- (a) Jalankan suatu ujian untuk menyiasat sama ada terdapat bukti perbezaan antara variasi amaun-amaun tuntutan dalam dua kawasan tersebut
- (b) Jalankan suatu ujian-t untuk menyiasat sama ada purata amaun tuntutan bagi kawasan B lebih rendah daripada kawasan A dengan amaun melebihi RM 15, serta berikan komen tentang kesahihan model tersebut.

[10 markah]

4. Pertimbangkan soalan-soalan berikut:-

- (a) Sampel-sampel rawak yang tidak bersandar bersaiz  $n_1 = 10$  dan  $n_2 = 25$ , masing-masing, diambil daripada taburan normal dengan variansnya yang sama nilai. Varians-variens sampel adalah  $S_1^2$  dan  $S_2^2$ , masing-masing.

Apakah nilai  $k$  supaya  $P\left(\frac{S_1^2}{S_2^2} > k\right) = 0.01$ ?

- (b) Amaun (RM) bagi wang saku yang diterima oleh suatu sampel rawak sejumlah 100 remaja perempuan berusia 13 tahun mempunyai min RM 4.56 dan sampel sisihan piawai RM 1.21. Kira 95% selang keyakinan bagi amaun min yang diterima di dalam populasi remaja perempuan tersebut.

[10 markah]

5. Pertimbangkan soalan-soalan berikut:-

- (a) Suatu sampel 12 pasangan cerapan  $x_i, y_i$  mempunyai pekali korelasi,  $r = 0.91$ . Uji pekali korelasi populasi,  $\rho$  untuk hipotesis berikut:-

$$H_0 : \rho = 0.8 \text{ lawan } H_1 : \rho > 0.8$$

- (b) The time spent waiting for a bus at a particular stop may be assumed to follow a distribution with mean 10 minutes and standard deviation 2 minutes. A sample of 25 independent observations from this distribution has mean  $\bar{T}$ . Calculate an approximate value of the probability that  $\bar{T} < 11$  minutes. [Hint: Use Central Limit Theorem to overcome this question]

[10 marks]

6. A random sample of size 20 is taken from the normal distribution,  $N(\mu, \sigma^2)$  and the following values are recorded:

-0.31   -0.90   -0.22   1.00   0.12   -0.01   -0.16   -1.31   0.38   -0.38  
 -1.07   1.65   -1.02   -0.06   -0.47   -0.10   0.71   0.94   -0.29   -0.57

Also given  $\sum x = -2.07$  and  $\sum x^2 = 10.9385$

- (a) Test the hypothesis  $H_0 : \sigma^2 = 1$  against the alternative,  $H_1 : \sigma^2 \neq 1$ .  
 (b) Determine the 95% confidence intervals for  $\mu$  by assuming that  
 (i)  $\sigma^2 = 1$ , and  
 (ii) no assumption about the value of  $\sigma^2$ .

[15 marks]

7. In the surgical treatment of duodenal ulcers there are three different operations corresponding to the removal of various amounts of the stomach. The three operations are denoted A, B and C, with A being the least traumatic and C the most traumatic.

It is known that these operations have an undesirable side-effect for some patients. In cases where the side effect is present, it can be classified as being of “slight degree” or of “moderate degree”. The data in the following table relate to a group of 417 patients and specify the operation received and, where relevant, the degree of the side effects suffered.

Operation	Existence/degree of side effects			Total
	None	Slight	Moderate	
A	63	26	7	96
B	126	63	25	214
C	51	40	16	107
Total	240	129	48	417

Perform a  $\chi^2$  test on this table to investigate the independence between level of operation and existence/degree of side-effects

[15 marks]

- (b) Masa yang digunakan untuk menunggu bas di sesebuah perhentian bas boleh diandaikan mengikuti taburan dengan min 10 minit dan sisihan piawai 2 minit. Suatu sampel 25 cerapan yang tidak bersandar daripada taburan itu mempunyai min  $\bar{T}$ . Kira nilai anggaran kebarangkalian  $\bar{T} < 11$  minit. [Petunjuk: Gunakan Teori Had Memusat untuk menyelesaikan soalan ini]

[10 markah]

6. Suatu sampel rawak bersaiz 20 diambil daripada taburan normal,  $N_{\mu, \sigma^2}$  dan nilai-nilai tersebut direkodkan seperti berikut:

-0.31   -0.90   -0.22   1.00   0.12   -0.01   -0.16   -1.31   0.38   -0.38  
 -1.07   1.65   -1.02   -0.06   -0.47   -0.10   0.71   0.94   -0.29   -0.57

Juga diberi  $\sum x = -2.07$  dan  $\sum x^2 = 10.9385$

- (a) Uji hipotesis  $H_0: \sigma^2 = 1$  lawan hipotesis alternatif,  $H_1: \sigma^2 \neq 1$ .  
 (b) Cari selang keyakinan 95% untuk  $\mu$  dengan mengandaikan  
 (i)  $\sigma^2 = 1$ , dan  
 (ii) Tiada andaian berkenaan nilai  $\sigma^2$ .

[15 markah]

7. Dalam rawatan pembedahan duodenal ulcers, terdapat tiga jenis pembedahan berlainan yang sepadan dengan penyingkiran pelbagai amaun di dalam perut. Tiga pembedahan tersebut ditandakan sebagai A, B dan C, dengan A mewakili sedikit trauma dan C paling banyak trauma.

Diketahui bahawa pembedahan-pembedahan ini mempunyai kesan sampingan ke atas sesetengah pesakit. Bagi kes yang mana kesan sampingan itu wujud, ia boleh diklasifikasikan sebagai kes "slight degree" atau "moderate degree". Data-data dalam jadual di bawah berkait dengan kumpulan 417 pesakit dan menyatakan pembedahan yang diterima, dan di mana relavan, darjah kesan sampingan yang dideritainya.

Pembedahan	Kehadiran / darjah kesan sampingan			Jumlah
	Tiada	Kecil	Sederhana	
A	63	26	7	96
B	126	63	25	214
C	51	40	16	107
Jumlah	240	129	48	417

Jalankan ujian  $\chi^2$  bagi jadual ini untuk menyiasat ketidakbarsandaran aras kelas pembedahan dan kehadiran / darjah kesan-kesan sampingan.

[15 markah]

8. The following paired observations give the body weight  $x$ , and the water intake  $y$ , for 7 subjects under a controlled experiment:

<i>Subject</i>	1	2	3	4	5	6	7
<i>X</i>	95	52	73	50	82	68	60
<i>Y</i>	3.94	1.03	1.71	1.75	1.76	2.01	0.97

- (a) Draw a scatter plot for these data and comment.
- (b) Determine the fitted regression line in which the water intake is modelled as the response and the body weight as an explanatory variable.
- (c) What proportion of the total variability in the responses is “explained” by the model? Comment on your answer.
- (d) By making appropriate assumptions, which should be stated, construct:
  - (i) A 95% confidence interval for the predicted water intake for an individual whose body weight is 80.

[20 marks]



8. Jadual di bawah merupakan pasangan cerapan bagi berat badan  $x$ , dan amaun pengambilan air  $y$ , bagi 7 subjek yang dikawal dalam sesebuah eksperimen:

Subjek	1	2	3	4	5	6	7
$X$	95	52	73	50	82	68	60
$Y$	3.94	1.03	1.71	1.75	1.76	2.01	0.97

- (a) Lukiskan plot scatter bagi data-data ini dan berikan komen.
- (b) Cari garis regresi yang tersuai yang mana pengambilan air dimodelkan sebagai pemboleh ubah respons manakala berat badan sebagai pemboleh ubah penerang.
- (c) Berapakah peratusan jumlah variasi dalam respons diterangkan daripada model ini? Berikan komen anda.
- (d) Dengan membuat andaian yang sesuai, bina:
- (i) Suatu selang keyakinan 95% untuk jangkaan pengambilan air bagi individu yang mempunyai berat badan 80.

[20 markah]

## APPENDIX

Binomial Distribution,  $X \sim \text{binomial } n, p$

$$\Pr X = x = \binom{n}{x} p^x (1-p)^{n-x}$$

$$E X = np$$

$$\text{Var } X = np(1-p)$$

$$M_X t = 1 - p + pe^t$$

For independent  $i = 1, \dots, n$  and  $Y_i \sim N(\mu, \sigma^2)$

$$\begin{aligned} \frac{Y_i - \mu}{\sigma} &\sim N(0, 1), & \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} &\sim N(0, 1), & \left(\frac{Y_i - \mu}{\sigma}\right)^2 &\sim \chi_1^2 \\ \left(\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}\right)^2 &\sim \chi_1^2, & s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}, & \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sigma^2} &\sim \chi_{n-1}^2 \\ n-1 \frac{s^2}{\sigma^2} &\sim \chi_{n-1}^2 \end{aligned}$$

For independent  $k = 1, \dots, m$ ;  $i = 1, \dots, n_k$  and  $Y_{ik} \sim N(\mu_k, \sigma^2)$ , the pooled variance is computed as follows

$$s_p^2 = \frac{\sum_{k=1}^m (n_k - 1) s_k^2}{\left(\sum_{k=1}^m n_k\right) - m} \text{ and } \left[\left(\sum_{k=1}^m n_k\right) - m\right] \frac{s_p^2}{\sigma^2} \sim \chi_{\left(\sum_{k=1}^m n_k\right) - m}^2$$

Additional notes:-

$$\text{If } Z \sim N(0, 1) \text{ and } A \sim \chi_a^2, \text{ then } \frac{Z}{\sqrt{A/a}} \sim t_a$$

$$\text{If } A \sim \chi_a^2 \text{ and } B \sim \chi_b^2, \text{ then } \frac{A/a}{B/b} \sim F_{a,b}$$