UNIVERSITI SAINS MALAYSIA

First Semester Examination
2011/2012 Academic Session

January 2012

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

Duration : 3 hours
[Masa : 3 jam]

Please check that this examination paper consists of SIXTEEN pages of printed material before you begin the examination.

[Sila pastikan bahawa kertas peperiksaan ini mengandungi ENAM BELAS muka surat yang bercetak sebelum anda memulakan peperiksaan ini.]

Instructions: Answer **all nine** [9] questions.

*[**Arahan:*** *Jawab **semua sembilan** [9] 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. For each of the following statements, state whether it is **TRUE** or **FALSE**:

   (a) A statistical test employs the technique of proof through the contradiction of the null hypothesis.
   (b) A Type I error is the rejection of null hypothesis, H_0 when H_0 is true and a Type II error is the acceptance of H_0 when the alternative hypothesis H_a is true.
   (c) The value of probability for a Type I error, $\beta\ \mu_a$ decreases as the distance from μ_a to μ_0 increases.
   (d) For a given sample size and value of the mean μ_a, α and $\beta\ \mu_a$ are inversely related.
   (e) For a data having a skewed distribution, the mean and standard deviation are appropriate measures of center and their variability respectively.
   (f) In statistical inference, we want tests having the largest possible power values because the power of a test is its ability to detect false null hypothesis.
   (g) If the population has a normal distribution, then the population mean and median are identical.

   [7 marks]

*1. Bagi setiap penyataan berikut, nyatakan sama ada ia **BENAR** atau **SALAH**.*

   (a) Ujian berstatistik menggunakan teknik pembuktian melalui penyangkalan hipotesis nol.
   (b) Ralat Jenis I ialah penolakan hipotesis nol, H_0 apabila H_0 benar dan ralat Jenis II ialah penerimaan H_0 apabila hipotesis alternatif adalah benar.
   (c) Nilai kebarangkalian bagi ralat Jenis I, $\beta\ \mu_a$ berkurangan apabila jarak dari μ_a ke μ_0 meningkat.
   (d) Bagi satu saiz sampel dan nilai min μ_a yang diberikan, α dan $\beta\ \mu_a$ adalah berhubungan secara songsang.
   (e) Bagi data yang mempunyai taburan terpencong, min dan sisihan piawai adalah ukuran memusat dan variabiliti yang bersesuaian.
   (f) Dalam pentakbiran statistik, kita berkehendakkan ujian yang mempunyai nilai kuasa yang terbesar mungkin kerana kuasa satu ujian ialah keupayaannya untuk mengesan hipotesis nol yang salah.
   (g) Jika populasi mempunyai taburan normal, maka min dan median populasi adalah serupa.

   [7 markah]

2. The manufacturer of an automatic control devise claims that the device will maintain a mean room humidity of 80%. The humidity in a controlled room was recorded for a period of 30 days, and the mean and the standard deviation were found to be 78.3% and 2.9%, respectively.

(a) Do the data present sufficient evidence to contradict the manufacturer's claim? Use $\alpha = 0.05$.

[5 marks]

(b) Determine the number of sample observations that would be required to estimate μ to within 1%, by using a 95% confidence interval.

[5 marks]

2. Pengeluar satu alat kawalan automatik mendakwa bahawa alat tersebut akan mengekalkan purata kelembapan bilik sebanyak 80%. Kelembapan dalam satu bilik kawalan dicatatkan untuk tempoh 30 hari, dan min dan sisihan piawaiannya didapati masing-masing ialah 78.3% dan 2.9%.

(a) Adakah data menunjukkan bukti yang cukup untuk menyangkal dakwaan pengeluar tersebut? Guna $\alpha = 0.05$

[5 markah]

(b) Tentukan bilangan sampel pemerhatian yang diperlukan untuk menganggar μ *sehingga antara 1% dengan menggunakan selang keyakinan 95%.*

[5 markah]

3. Improperly filled orders are a costly problem for mail-order houses. To estimate the mean loss per incorrectly filled order, a large firm plans to sample *n* incorrectly filled orders and to determine the added cost associated with each one. The firm estimates that the added cost is between RM40 and RM400. How many incorrectly filled orders must be sampled to estimate the mean additional cost using a 95% confidence interval of width RM20?

[5 marks]

3. Pesanan yang disilap isi adalah masalah yang berkos bagi rumah pesanan-pos. Bagi menganggar purata kerugian bagi setiap pesanan yang silap diisi, satu firma yang besar bercadang untuk membuat pensampelan n pesanan silap isi bagi menentukan kos tambahan yang berkaitan bagi setiap satu. Firma tersebut menganggarkan bahawa kos tambahan ialah antara RM40 dan RM400. Berapa banyakkah pesanan silap isi yang mesti disampel bagi menganggar min kos tambahan menggunakan selang keyakinan 95% dengan kelebaran RM20?

[5 markah]

4. An experiment was conducted to evaluate the effectiveness of a treatment for tapeworm in the stomachs of sheep. A random sample of 24 worm-infected sheep of approximately the same age and health was randomly divided into two groups. Twelve of the sheep were injected with the drug and the remaining twelve were left untreated. After a 6-month period, the sheep were slaughtered and the following worm counts were recorded.

Drug-treated sheep	18	43	28	50	16	32	13	35	38	33	6	7
Untreated sheep	40	54	26	63	21	37	39	23	48	58	28	39

With the following data:

Drug-treated sheep: $n_1 = 12, \bar{y}_1 = 26.58, s_1 = 14.36$
Untreated sheep : $n_1 = 12, \bar{y}_2 = 39.67, s_2 = 13.86$

(a) Test whether the mean number of tapeworms in the stomachs of the treated sheep is less than the mean for untreated sheep. Use an $\alpha = 0.05$.

[6 marks]

(b) Find a 95% confidence interval on $\mu_1 - \mu_2$ to assess the size of the difference in the two means.

[5 marks]

4. Satu ujikaji telah dijalankan bagi menilai keberkesanan satu rawatan cacing dalam perut biri-biri. Satu sampel rawak terdiri 24 ekor biri-biri yang dijangkiti cacing yang dianggarkan sama umur dan tahap kesihatannya telah dibahagi secara rawak kepada dua kumpulan. Dua belas biri-biri disuntik dengan ubat dan baki yang dua belas ekor lagi dibiarkan tampa rawatan. Selepas tempoh 6 bulan, biri-biri tersebut disembelih dan berikut merupakan bilangan cacing yang direkodkan:

Biri-biri dirawat ubat	*18*	*43*	*28*	*50*	*16*	*32*	*13*	*35*	*38*	*33*	*6*	*7*
Biri-biri tidak dirawat	*40*	*54*	*26*	*63*	*21*	*37*	*39*	*23*	*48*	*58*	*28*	*39*

Dengan data berikut:

Biri-biri dirawat ubat: $n_1 = 12, \bar{y}_1 = 26.58, s_1 = 14.36$
Biri-biri tidak dirawat : $n_1 = 12, \bar{y}_2 = 39.67, s_2 = 13.86$

(a) Uji sama ada min bilangan cacing dalam perut biri-biri yang dirawat adalah kurang daripada min bagi biri-biri yang tidak dirawat. Guna $\alpha = 0.05$.

[6 markah]

(b) Cari selang keyakinan 95% bagi $\mu_1 - \mu_2$ *untuk menilai saiz perbezaan dalam dua min tersebut.*

[5 markah]

5. (a) Suppose we have a random sample of 30 measurements from a population having median M. We want to test $H_0 : M \leq M_0$ versus $H_1 : M > M_0$ at level $\alpha = 0.05$. Identify the rejection region for testing these hypotheses.

[5 marks]

(b) Then, use the large-sample approximation to identify the rejection region and compare your results to the rejection region obtained in (a).

[5 marks]

5. (a) Katakan kita mempunyai sampel rawak bagi 30 ukuran daripada satu populasi yang mempunyai median M. Kita mahu menguji $H_0 : M \leq M_0$ lawan $H_1 : M > M_0$ pada aras $\alpha = 0.05$. Kenalpastikan kawasan penolakan bagi menguji hipotesis ini.

[5 markah]

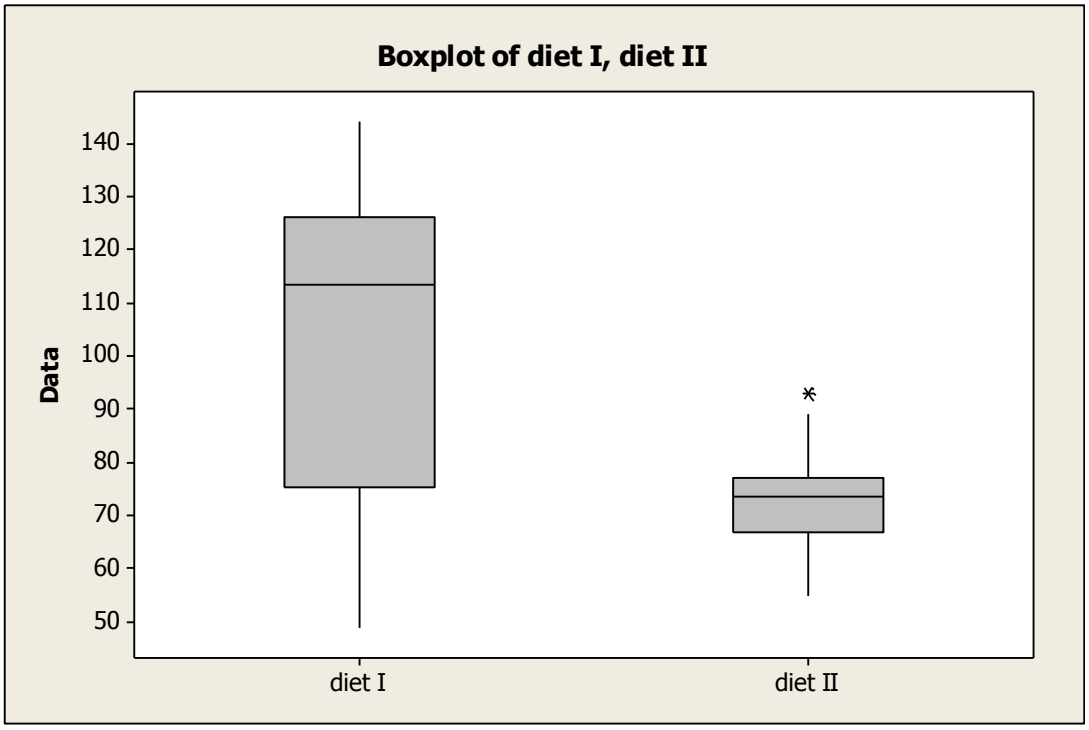
(b) Kemudian, gunakan penganggaran sampelan-besar untuk kenalpasti kawasan penolakan dan bandingkan keputusan anda dengan kawasan penolakan yang diperolehi dalam (a).

[5 markah]

6. An experiment was conducted to compare the weights of the wool of sheep feds two different vitamin-supplemented diets. Twenty-eight healthy sheep were randomly divided into two groups, with one group receiving diet I and the other receiving diet II. After the study period, the wool weight (in milligrams) was recorded for each sheep. The data are given as:

Diet I	73	130	115	144	127	112	76	68	101	126	49	110	123	126
Diet II	80	72	73	60	55	74	67	89	75	66	93	75	68	76

The boxplot of the data as follows:



with their descriptive statistics as

Descriptive Statistics: diet I

```
Variable     N    Mean  SE Mean  StDev  Variance  Minimum     Q1  Median
diet I      14  105.71     7.56  28.27    799.14    49.00  75.25  113.50


Variable      Q3  Maximum    IQR
diet I    126.25   144.00  51.00
```

Descriptive Statistics: diet II

```
Variable     N    Mean  SE Mean  StDev  Variance  Minimum     Q1  Median    Q3
diet II     14  73.07     2.70  10.10    102.07    55.00  66.75   73.50  77.00


Variable  Maximum    IQR
diet II     93.00  10.25
```

…7/-

(a) Use the appropriate t procedure to determine whether there is a difference in the distributions of wool weights for the two groups. Use $\alpha = 0.05$.

[5 marks]

(b) Use the Wilcoxon rank sum test to determine whether there is a difference in the distributions of wool weights for the two groups. Use $\alpha = 0.05$.

[10 marks]

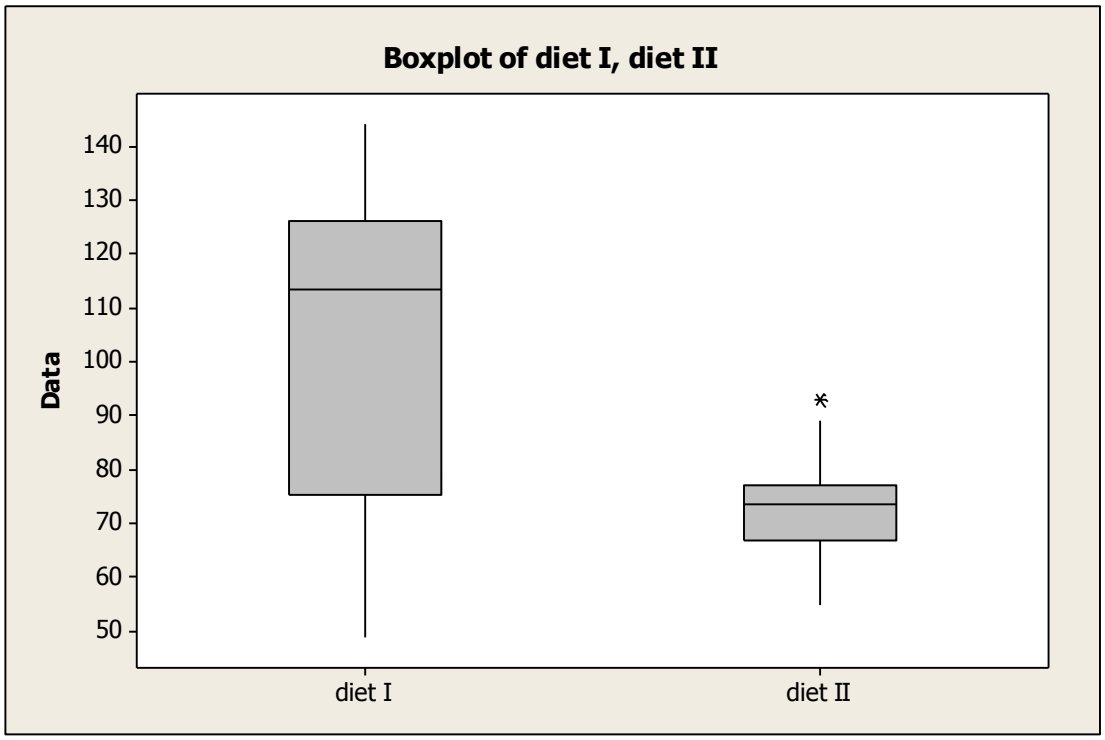
(c) Which procedure, Wilcoxon or t, seems most appropriate for evaluating the results of this experiment? Explain your answer.

[5 marks]

6. *Satu ujikaji dijalankan bagi membandingkan berat bulu biri-biri yang telah diberi makan dua diet dengan vitamin tambahan yang berbeza. Dua puluh lapan ekor biri-biri sihat dibahagikan secara rawak kepada dua kumpulan, dengan satu kumpulan menerima diet I dan yang lainnya menerima diet II. Selepah tempoh kajian, berat bulu (dalam milligram) dicatatkan bagi setiap biri-biri. Data tersebut adalah seperti berikut:*

Diet I	*73*	*130*	*115*	*144*	*127*	*112*	*76*	*68*	*101*	*126*	*49*	*110*	*123*	*126*
Diet II	*80*	*72*	*73*	*60*	*55*	*74*	*67*	*89*	*75*	*66*	*93*	*75*	*68*	*76*

Plot kotak bagi data tersebut seperti berikut:



dengan data deskriptif sebagai;

Descriptive Statistics: diet I

```
Variable     N    Mean  SE Mean  StDev  Variance  Minimum     Q1  Median
diet I      14  105.71     7.56  28.27    799.14    49.00  75.25  113.50


Variable      Q3  Maximum    IQR
diet I    126.25   144.00  51.00
```

Descriptive Statistics: diet II

```
Variable     N    Mean  SE Mean  StDev  Variance  Minimum     Q1  Median     Q3
diet II     14  73.07     2.70  10.10    102.07    55.00  66.75   73.50  77.00


Variable  Maximum    IQR
diet II     93.00  10.25
```

…9/-

(a) *Gunakan prosedur t yang bersesuaian bagi menentukan sama ada terdapat perbezaan dalam taburan berat bulu bagi dua kumpulan tersebut. Guna* $\alpha = 0.05$.

[5 markah]

(b) *Gunakan ujian hasil tambah pangkat Wilcoxon bagi tentukan sama ada terdapat perbezaan dalam taburan berat bulu bagi dua kumpulan tersebut.*

[10 markah]

(c) *Prosedur yang mana, Wilcoxon atau t, kelihatan bersesuaian bagi menilai keputusan ujikaji ini? Huraikan jawapan anda.*

[5 markah]

7. The following values were computed from the length of life of two brands of light bulbs (in hours):

	Brand A	**Brand B**
n	9	16
$\bar{y}$	1560	1573
$\sum y-\bar{y}^{2}$	440	1860

(a) Is there a difference in the variability of lifetimes for the two brands of bulbs? State your assumption. (Use $\alpha = 0.02$).

[5 marks]

(b) Find a 98% confidence interval on the ratio of the two variability.

[5 marks]

7. Nilai-nilai berikut dikira daripada panjang hayat bagi dua jenama lampu bulat (dalam jam)

	Jenama A	***Jenama B***
n	*9*	*16*
$\bar{y}$	*1560*	*1573*
$\sum y-\bar{y}^{2}$	*440*	*1860*

(a) Adakah terdapat perbezaan dalam variabiliti bagi jangka hayat kedua-dua jenama lampu bulat tersebut? Nyatakan andaian anda. (Guna $\alpha = 0.02$)

[5 markah]

(b) Cari selang keyakinan 98% terhadap nisbah kedua variabiliti tersebut.

[5 markah]

8. An experiment is designed to study the side effects of two drugs used as treatments for a certain ailment. A group of 90 subjects are assigned to two drug groups. After the specified drug is given, the side effects are classified as follows:

	Side Effects			
	Major	Minor	None	Total
Drug A	13	15	17	45
Drug B	8	21	16	45

Are the side effects distributed the same for the two drugs? State your hypotheses accordingly.

[7 marks]

8. Satu ujikaji direka untuk mengkaji kesan sampingan dua jenis ubat yang digunakan sebagai rawatan bagi penyakit tertentu. Satu kumpulan 90 orang diagihkan kepada dua kumpulan ubat tersebut. Selepas ubat tertentu diberikan, kesan sampingan dikelaskan seperti berikut:

	Kesan sampingan			
	Utama	*Kecil*	*Tiada*	*Jumlah*
Ubat A	*13*	*15*	*17*	*45*
Ubat B	*8*	*21*	*16*	*45*

Adakah kesan sampingan bertaburan yang sama bagi kedua ubat tersebut? Nyatakan hipotesis anda yang bersesuaian.

[7 markah]

9. A forest entomologist has isolated 7 insecticides that are reasonably safe to the rest of the environment when used to control gypsy moths. She wants to determine whether any one of them produces significantly greater mortality than the others. Using standard technique, she applies a given insecticide to the abdoment of each of 100 moths. This procedure is repeated 5 times for each insecticide, with new solutions being prepared each time. Percent mortality is recorded after 24 hours for each insecticide trial. Assume that the data, although distributed in binomial fashion, will approximate the normal distribution adequately for ANOVA procedures.

(a) Although 3500 moths are used, why are there only 34 degrees of freedom associated with this experiment?

[2 marks]

(b) In using the y_{ij} notation, does the j subscript refer to the insecticide or the trial?

[2 marks]

(c) What are the assumptions for an ANOVA?

[3 marks]

(d) Use this information to complete the accompanying ANOVA table:

$$\frac{\left(\sum_i \sum_j y_{ij}\right)^2}{an} = 143,360 \quad ; \quad \sum_i \sum_j y_{ij}^2 = 144,334$$

Source	df	SS	MS
Insecticides			55
Trials within insecticides			

[5 marks]

(e) Give the null and alternative hypotheses.

[2 marks]

(f) Give the critical value ($\alpha = 0.05$) for a test of the above hypothesis and draw conclusions about the experiment.

[6 marks]

9. *Seorang pakar serangga hutan telah mengasingkan 7 racun serangga yang agak selamat kepada alam sekitar apabila digunakan untuk mengawal kupu-kupu gipsi. Beliau mahu tentukan sama ada salah satu daripadanya menyebabkan kadar kematian yang lebih signifikan daripada yang selainnya. Menggunakan teknik piawai, beliau menyapu racun serangga tertentu pada perut bagi setiap 100 ekor kupu-kupu. Prosedur ini diulang 5 kali bagi setiap racun serangga dengan larutan yang baru disediakan pada setiap masa. Peratusan kematian dicatatkan selepas 24 jam bagi setiap cubaan racun serangga. Andaikan bahawa data walaupun tertabur secara binomial, akan menghampiri taburan normal bersesuaian untuk prosedur ANOVA.*

(a) *Walaupun 3500 ekor kupu-kupu diguna, mengapa hanya 34 darjah kebebasan digunakan bagi ujikaji ini?*

[2 markah]

(b) *Dalam menggunakan pelambangan y_{ij} , adakah subskrip j merujuk kepada racun serangga atau cubaan?*

[2 markah]

(c) *Apakah andaian-andaian untuk satu ANOVA?*

[3 markah]

(d) *Guna maklumat ini untuk melengkapkan jadual ANOVA yang disertakan:*

$$\frac{\left(\sum_i \sum_j y_{ij}\right)^2}{an} = 143{,}360 \quad ; \quad \sum_i \sum_j y_{ij}^2 = 144{,}334 .$$

Source	*df*	*SS*	*MS*
Insecticides			*55*
Trials within insecticides			

[5 markah]

(e) *Berikan hipotesis nol dan alternatif.*

[2 markah]

(f) *Berikan nilai genting ($\alpha = 0.05$) untuk menguji hipotesis di atas dan nyatakan kesimpulan mengenai ujikaji tersebut.*

[6 markah]

Appendix/Lampiran

Useful formulations:

1. $\frac{n-1\ s^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{n-1\ s^2}{\chi^2_{1-\alpha/2}}$.

2. $\frac{s_1^2}{s_2^2} F_{1-\alpha/2,\nu_2,\nu_1} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{s_1^2}{s_2^2} F_{\alpha/2,\nu_2,\nu_1}$.

3. $t' = \frac{\bar{y}_1 - \bar{y}_2\ - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ with degrees of freedom $w = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$.

4. $n = \frac{2z^2_{\alpha/2}\sigma^2}{E^2}$.

5. $n = 2\sigma^2 \frac{z_{\alpha/2} + z_\beta\ ^2}{\Delta^2}$.

6. $n_1 = \left(\frac{m+1}{m}\right)\sigma^2 \frac{Z_{\alpha/2} + Z_\beta\ ^2}{\Delta^2}$.

7. $\sigma_T^2 = \frac{n_1 n_2}{12}\left(n_1 + n_2 + 1\ - \frac{\sum_{j=1}^{k} t_j\ \ t_j^2 - 1}{n_1 + n_2\quad n_1 + n_2 - 1} \right)$.

Table 1

…16/-

Table 2