
UNIVERSITI SAINS MALAYSIA

Second Semester Examination
Academic Session 2006/2007

April 2007

MGM 561 – Statistical Methods For Research
[Kaedah Statistik Untuk Penyelidikan]

Duration : 3 hours
[Masa : 3 jam]

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

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

Instructions: Answer **all four** [4] questions.

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

...2/-

1. (a) In a survey of the prevalence of drug use, randomly sampled secondary school students are to be asked if they have ever used any abused drugs, such as marijuana, heroine or ecstasy. Their responses are to be completely confidential.
- (i) If the proportion of abused-drug users among secondary school students is believed to be about 3%, how many students should be included in the sample if the investigator wishes to be 95% certain that the tolerable error (W) of estimation of the proportion of abused-drug users does not exceed 0.04?
- (ii) Suppose that a random sample of 300 students is actually taken and that 12 of them report they have used abused drugs. Based on this information, construct a 90% confidence interval for the proportion of abused-drug users among secondary school students.

[20 marks]

- (b) It is estimated that lead poisoning may affect as many as a quarter of a million children each year, causing them to suffer from severe, irreversible retardation. To determine the presence of lead poisoning, a blood sample is taken and the lead levels in blood are measured. A journal article suggests that testing lead in hair samples is an acceptable screening method and avoid the need to take a blood sample. A study is designed where a sample of 1600 children of age 1-5 years in an urban area is taken and the standard method (test from blood samples) to get the actual blood lead levels is performed. It is found that lead poisoning presents in 90 of them. The hair-screening test is then conducted to all of these children and the following table (Table 1) shows the results of the test.

Table 1

	Lead poisoning	
	Present	Absent
Positive hair screening	78	15
Negative hair screening	12	1495

It is believed that 5% of the children in the area have lead poisoning.

.../3-

1. (a) Dalam suatu kaji selidik mengenai penyalahgunaan dadah, pelajar sekolah menengah yang dipilih secara rawak ditanya sama ada mereka pernah terlibat dalam penyalahgunaan dadah terlarang seperti ganja, heroin atau pil ekstasi. Maklumbalas mereka adalah sulit sepenuhnya.
- (i) Jika kadar penyalahguna dadah dalam kalangan pelajar sekolah menengah dipercayai sekitar 3%, berapa ramaiakah pelajar yang perlu dimasukkan dalam sampel jika penyelidik ingin 95% yakin bahawa ralat toleransi (W) bagi penganggaran kadar penyalahguna dadah tidak melebihi 0.04?
- (ii) Andaikan suatu sampel rawak 300 pelajar diambil dan didapati 12 daripadanya melaporkan mereka pernah terlibat dengan penyalahgunaan dadah. Berdasarkan maklumat ini, bina 90% selang keyakinan bagi kadar penyalahguna dadah dalam kalangan pelajar sekolah menengah.

[20 markah]

- (b) Dianggarkan keracunan plumbum boleh berlaku kepada hampir suku juta kanak-kanak setiap tahun dan menyebabkan mereka menderita akibat kerencatan otak yang teruk dan berkekalan. Bagi menentukan kewujudan keracunan plumbum, sampel darah diambil dan paras plumbum dalam darah diukur. Sebuah artikel jurnal mencadangkan bahawa menguji plumbum dalam sampel rambut adalah kaedah penyaringan yang boleh diterima dan mengelak keperluan untuk pengambilan sampel darah. Satu kajian direka yang mana suatu sampel yang terdiri daripada 1600 kanak-kanak berumur 1-5 tahun di suatu bandar diambil dan kaedah standard (uji daripada sampel darah) bagi mendapatkan paras sebenar plumbum dalam darah dijalankan. Keputusan mendapati keracunan plumbum dalam darah terdapat dalam 90 daripada kanak-kanak itu. Ujian penyaringan rambut dijalankan kemudiannya ke atas semua kanak-kanak itu dan jadual di bawah (Jadual 1) menunjukkan keputusan ujian tersebut.

Jadual 1

	Keracunan plumbum	
	Ada	Tiada
Penyaringan rambut positif	78	15
Penyaringan rambut negatif	12	1495

Dipercayai 5% daripada kanak-kanak di bandar itu mempunyai keracunan plumbum.

.../4-

- (i) Find the sensitivity and specificity of the hair-screening method.

A child of age 1-5 years in the area is randomly chosen and the hair screening test for lead levels is performed on her.

- (ii) Find the probability that the child truly has lead poisoning given that the hair-screening test is positive.
- (iii) Find the probability that the child truly does not have lead poisoning given that the hair-screening test is positive.
- (iv) Find the probability that the child truly does not have lead poisoning given that the hair-screening test is negative.

[30 marks]

2. (a) An experiment was conducted to compare the filling capability of packaging equipment at two different manufacturing companies of mineral water. Twenty bottles of mineral water from company A were randomly selected and measured, along with 20 bottles of mineral water from company B. The data are as follows (fill volume is in ml).

<u>Company A</u>				<u>Company B</u>			
254	256	255	256	252	253	254	251
257	245	256	256	253	252	253	253
256	255	256	252	251	254	250	255
253	256	254	256	255	253	253	252
255	254	256	256	253	251	252	253

$$\sum x_1 = 5094$$

$$\sum (x_1 - \bar{x}_1)^2 = 128.20$$

$$\sum x_2 = 5053$$

$$\sum (x_2 - \bar{x}_2)^2 = 32.55$$

.../5-

- (i) Dapatkan sensitiviti dan spesifikasi bagi kaedah penyaringan rambut.

Seorang kanak-kanak berumur 1-5 tahun dipilih secara rawak dalam bandar itu dan ujian penyaringan rambut dijalankan ke atasnya.

- (ii) Dapatkan kebarangkalian yang kanak-kanak itu sebenarnya mempunyai keracunan plumbum jika ujian penyaringan rambutnya adalah positif.
- (iii) Dapatkan kebarangkalian yang kanak-kanak itu sebenarnya tidak mempunyai keracunan plumbum jika ujian penyaringan rambutnya adalah positif.
- (iv) Dapatkan kebarangkalian yang kanak-kanak itu sebenarnya tidak mempunyai keracunan plumbum jika ujian penyaringan rambutnya adalah negatif.

[30 markah]

2. (a) Suatu eksperimen dijalankan untuk membandingkan keupayaan mengisi bagi peralatan pembungkusan di dua syarikat pengeluaran air mineral. Dua puluh botol air mineral daripada Syarikat A dipilih secara rawak dan kandungannya disukat, bersama dengan 20 botol air mineral daripada Syarikat B. Datanya adalah seperti berikut (isipadu dalam ml).

<u>Syarikat A</u>				<u>Syarikat B</u>			
254	256	255	256	252	253	254	251
257	245	256	256	253	252	253	253
256	255	256	252	251	254	250	255
253	256	254	256	255	253	253	252
255	254	256	256	253	251	252	253

$$\sum x_1 = 5094$$

$$\sum (x_1 - \bar{x}_1)^2 = 128.20$$

$$\sum x_2 = 5053$$

$$\sum (x_2 - \bar{x}_2)^2 = 32.55$$

.../6-

- (i) Do the data support the claim that both companies fill bottles to the same mean volume? Use $\alpha = 0.05$ and assume that the normality assumption is satisfied and both population standard deviation are equal in drawing conclusions.
- (ii) Calculate p -value for the statistical test in part (i).
- (iii) Construct box-plot for the two samples. Does it seem that the assumptions of normality and equal variances made in part (i) are reasonable? Provide a practical interpretation for these plots.

[30 marks]

- (b) A school administrator wishes to introduce a new core curriculum activity for the students in the school to enhance their soft skills. Before implementing the activity, he needs to obtain students' opinion regarding this matter. A random sample of 200 consisting of Form 1 to Form 5 students is asked about their opinion on a proposed core curriculum activity. The results are as follows.

Table 2

Class	Opinion	
	Favouring	Opposing
Form 1	32	5
Form 2	38	12
Form 3	12	13
Form 4	30	18
Form 5	25	15

- (i) Test the null hypothesis that opinion on the proposed activity is independent of class standing. Can the null hypothesis be rejected at $\alpha = 0.05$?
- (ii) Compute the p -value of this test.

[20 marks]

.../7-

- (i) Adakah data tersebut menyokong bahawa kedua-dua syarikat mengisi botol dengan min isipadu yang sama? Gunakan $\alpha = 0.05$ dan andaikan bahawa andaian kenormalan dipenuhi dan sisihan piawai kedua-dua populasi adalah sama dalam membuat kesimpulan.
- (ii) Kira nilai-p bagi ujian statistik di bahagian (i).
- (iii) Bina gambarajah plot-kotak untuk kedua-dua sampel. Adakah andaian kenormalan dan kesamaan varians yang dibuat di bahagian (i) munasabah? Berikan tafsiran untuk plot-plot ini.

[30 markah]

- (b) Pentadbir sebuah sekolah berhasrat untuk memperkenalkan satu aktiviti ko-kurikulum kepada pelajar bagi meningkatkan kemahiran insaniah mereka. Sebelum menjalankan aktiviti tersebut, beliau perlu mendapatkan pandangan pelajar berkenaan cadangan itu. Sampel rawak seramai 200 pelajar yang terdiri daripada pelajar Tingkatan 1 hingga Tingkatan 5 ditanya pendapat mereka mengenai cadangan aktiviti ko-kurikulum itu. Keputusannya adalah seperti berikut.

Jadual 2

Kelas	Pendapat	
	Menyokong	Menentang
Tingkatan 1	32	5
Tingkatan 2	38	12
Tingkatan 3	12	13
Tingkatan 4	30	18
Tingkatan 5	25	15

- (i) Uji hipotesis nol bahawa pendapat mengenai akitiviti yang dicadangkan itu tidak bersandar dengan kedudukan kelas pelajar. Bolehkah hipotesis nol ditolak pada $\alpha = 0.05$?
- (ii) Kira nilai-p bagi ujian ini.

[20 markah]

3. (a) Pack of a certain brand of snacks carry the specification that the average composition of fat is 7 grams per pack. A government testing agency chemically analyzes a random sample of 100 of these packs and finds that the sample average and the standard deviation are 7.25 grams and 0.5 grams, respectively.
- (i) Assuming $\alpha = 0.05$, construct a test to determine if these results provide strong evidence that the average composition of fat in the pack is higher than the specified value.
- (ii) State your assumptions in (i).
- (iii) What is the smallest level of significance that will lead to a rejection of the null hypothesis on the basis of the data?

[25 marks]

- (b) Suppose that, as a teacher, you want to compare the effectiveness of three methods of teaching a topic in mathematics for secondary school students. Method 1 is the conventional method, which is what you are doing now; Method 2 involves using the graphing calculator whereas Method 3 is using computer software to solve the problems. Random samples of size 10 are taken from large groups of students taught by the three methods and an appropriate test is given to each of them. The summary of their scores is given as follows.

	<u>Method 1</u>	<u>Method 2</u>	<u>Method 3</u>
Sample size	10	10	10
Sample mean	78	82	83
Sample standard deviation	4.8	3.9	3.2

- (i) Construct an analysis-of-variance table for these data.
- (ii) Test the null hypothesis that the three methods of teaching are equally effective against the alternative hypothesis that they are not equally effective at 0.01 level of significance.
- (iii) If the level of significance is increased to 0.05, will your conclusion in (ii) change?
- (iv) State the assumptions of the test.

[25 marks]

.../9-

3. (a) *Bungkusan sejenis makanan ringan mempunyai spesifikasi yang menyatakan komposisi purata lemak ialah 7 gram bagi setiap bungkus. Sebuah agensi pengujian kerajaan menganalisis secara kimia suatu sampel rawak 100 bungkusan makanan ringan tersebut dan mendapati purata dan sisihan piawai bagi sampel itu ialah masing-masing 7.25 gram dan 0.5 gram.*
- (i) *Andaikan $\alpha = 0.05$, bina suatu ujian untuk menentukan jika keputusan ini memberi bukti yang kukuh bahawa komposisi purata lemak dalam bungkusan itu lebih tinggi daripada nilai yang ditetapkan.*
- (ii) *Nyatakan andaian-andaian anda dalam bahagian (i).*
- (iii) *Apakah aras keertian terendah yang membawa kepada penolakan hipotesis nol berasaskan data tersebut?*

[25 markah]

- (b) *Andaikan, sebagai guru, anda ingin membandingkan keberkesanan tiga kaedah mengajar suatu topik dalam matematik untuk pelajar sekolah menengah. Kaedah 1 ialah kaedah biasa iaitu kaedah yang anda amalkan sekarang; Kaedah 2 melibatkan penggunaan kalkulator grafik manakala Kaedah 3 ialah dengan menggunakan suatu perisian komputer untuk menyelesaikan masalah. Sampel rawak bersaiz 10 diambil daripada kumpulan besar pelajar yang diajar kaedah-kaedah ini dan suatu ujian bersesuaian diberi kepada setiap pelajar ini. Ringkasan markah mereka diberi seperti berikut.*

	<u>Kaedah 1</u>	<u>Kaedah 2</u>	<u>Kaedah 3</u>
<i>Saiz sampel</i>	10	10	10
<i>Min sampel</i>	78	82	83
<i>Sisihan piawai sampel</i>	4.8	3.9	3.2

- (i) *Bina jadual analisis varians bagi data ini.*
- (ii) *Uji hipotesis nol bahawa ketiga-tiga kaedah pengajaran ini adalah sama berkesan melawan hipotesis alternatif bahawa kaedah-kaedah ini bukan sama berkesan pada aras keertian 0.01.*
- (iii) *Jika aras keertian ditingkatkan kepada 0.05, adakah kesimpulan anda di bahagian (ii) akan berubah?*
- (iv) *Nyatakan andaian-andaian bagi ujian ini.*

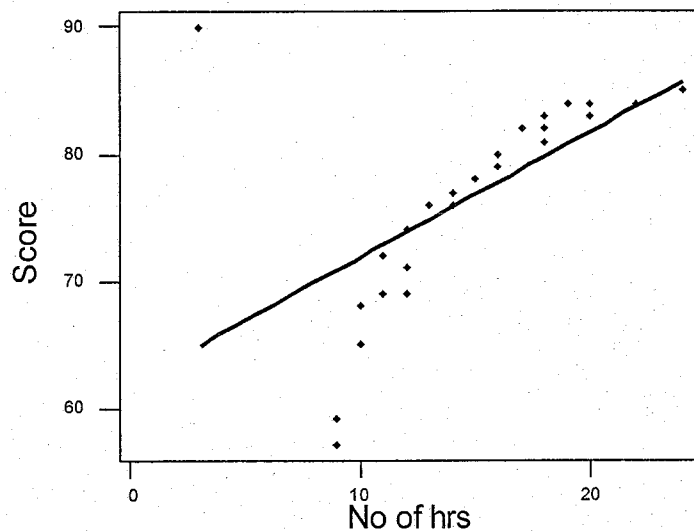
[25 markah]

.../10-

4. (a) Suppose that you are interested in estimating the annual mean of expenses for books for the primary school students in Penang.
- Identify the population and variable of interest.
 - How might you select a sample from this population?
 - Now, suppose that you have the information that the annual expenses for books for the primary school students in Penang in the year 2005 was RM30.00. State several hypotheses that might be of interest to you.

[10 marks]

- (b) A teacher wishes to determine the relationship between the number of hours the students study for a certain achievement test and their score on the test. He randomly selects 25 students of Form 5 in the school. The students are asked about the number of hours they study for a mathematics test and their score on the test. The teacher believes that there is a linear relationship between the score on the test and the number of hours the students study for the test. A scatter plot of the data and the regression analysis results are given as follows.

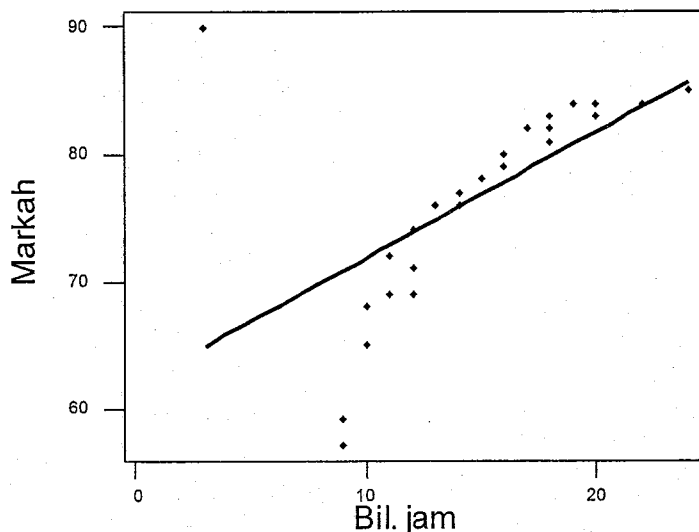


.../11-

4. (a) *Andaikan anda berminat untuk menganggar purata perbelanjaan tahunan untuk buku bagi pelajar sekolah rendah di Pulau Pinang.*
- (i) *Kenalpasti populasi dan pembolehubah yang berkaitan.*
 - (ii) *Bagaimana anda akan memilih sampel daripada populasi ini?*
 - (iii) *Sekarang, katakan anda mempunyai maklumat bahawa purata perbelanjaan untuk buku bagi pelajar sekolah rendah di Pulau Pinang pada tahun 2005 ialah RM30.00. Nyatakan beberapa hipotesis yang mungkin menarik minat anda.*

[10 markah]

- (b) *Seorang guru ingin menentukan perhubungan antara bilangan jam belajar yang diperuntukkan oleh seorang pelajar bagi suatu ujian pencapaian dan markah yang diperolehi dalam ujian tersebut. Pelajar ditanya mengenai bilangan jam yang diperuntukkan untuk suatu ujian matematik dan markah yang diperolehi untuk ujian tersebut. Guru tersebut percaya bahawa terdapat perhubungan linear antara markah yang diperolehi dan bilangan jam yang diperuntukkan pelajar untuk belajar bagi menghadapi ujian tersebut. Plot taburan dan keputusan analisis regresi diberi seperti berikut.*



.../12-

Table 3

The regression equation is					
Score = a + b. No of hrs					
Predictor	Coef	SE Coef	T	P	
Constant	61.941	4.537	13.65	0.000	
No of hr	0.9903	0.2972	3.33	0.003	
S = 6.997		R-Sq = 32.6%	R-Sq(adj) = 29.6%		
Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	1	543.49	543.49	D	0.003
Residual Error	A	B	C		
Total	24	1669.44			

- (i) Obtain the values of A, B, C and D in Table 3.
- (ii) Write the regression equation relating the score on the test to the numbers of hours the student studies for the test and interpret the slope coefficient.
- (iii) Is there a statistically significant relation between the score on the test and the number of hours the student studies for the test?
- (iv) If the point at the upper left to the scatter plot is deleted, will the slope of regression line increase or decrease? Do you expect a substantial change?
- (v) Do you think the assumption of linear relationship is appropriate for this data? Suppose that the point at the upper left to the scatter plot is deleted. Suggest a better relationship to relate the score on the test to the number of hours the student studies for the test.

[40 marks]

Jadual 3

The regression equation is					
Markah = a + b. Bil. jam					
Predictor	Coef	SE Coef	T	P	
Constant	61.941	4.537	13.65	0.000	
Bil. jam	0.9903	0.2972	3.33	0.003	
S = 6.997		R-Sq = 32.6%		R-Sq(adj) = 29.6%	
Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	1	543.49	543.49	D	0.003
Residual Error	A	B	C		
Total	24	1669.44			

- (i) Dapatkan nilai A, B, C dan D dalam Jadual 3.
- (ii) Tuliskan persamaan regresi yang menghubungkan markah ujian kepada bilangan jam pelajar belajar untuk ujian tersebut dan tafsirkan pekali kecerunan.
- (iii) Adakah terdapat perhubungan bererti secara statistik antara markah ujian dan bilangan jam pelajar belajar untuk ujian itu?
- (iv) Jika titik pada bahagian kiri paling atas dalam gambarajah plot itu dibuang, adakah kecerunan garis regresi menokok atau menyusut? Adakah anda menjangkakan suatu perubahan yang ketara?
- (v) Adakah anda rasa anggapan hubungan linear ini bersesuaian? Andaikan titik pada bahagian kiri paling atas dalam gambarajah plot itu dibuang. Cadangkan suatu hubungan lain yang anda rasa lebih baik untuk menghubungkan markah ujian kepada bilangan jam pelajar belajar untuk ujian tersebut.

[40 markah]

.../14-

**MGM 561 – STATISTICAL METHODS FOR RESEARCH
SOME USEFUL FORMULA**

Confidence Interval

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$$

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$df = \frac{(n_1 - 1)(n_2 - 1)}{(1 - c^2)(n_1 - 1) + c^2(n_2 - 1)},$$

$$c = \frac{s_1^2/n_1}{s_1^2/n_1 + s_2^2/n_2}$$

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_2)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_1)}{n_2}}$$

$$\left(\frac{(n-1)s^2}{\chi_{\alpha/2, n-1}^2}, \frac{(n-1)s^2}{\chi_{1-\alpha/2, n-1}^2} \right)$$

$$\left(\frac{s_1^2}{s_2^2} F_{1-\alpha/2, df_2, df_1}, \frac{s_1^2}{s_2^2} F_{\alpha/2, df_2, df_1} \right)$$

Test Statistic

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

$$T = \frac{\bar{d} - D_0}{s_d/\sqrt{n}}$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - D_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$T' = \frac{(\bar{X}_1 - \bar{X}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_2)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_1)}{n_2}}}$$

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

$$F = \frac{s_1^2}{s_2^2}; F = \frac{s_{\max}^2}{s_{\min}^2}$$

$$\chi^2 = \sum_i \frac{(n_i - E_i)^2}{E_i}; \chi^2 = \sum_{i,j} \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

-0000000-