UNIVERSITI SAINS MALAYSIA

First Semester Examination
Academic Session 2005/2006

November 2005

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

Duration : 3 hours
[Masa : 3 jam]

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

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

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

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

...2/-

1. (a) A hotel manager would like to know the average length of stay of the guests in the hotel. It would be impossible to look through all of the past records and average the lengths of stay. Instead, 100 guests over the last year were randomly selected and their lengths of stay were averaged.

(i) Describe the population of interest.

(ii) What is the sample in this problem?

(iii) What variable should be measured?

(iv) Give your comment about this sampling technique.

[20 marks]

(b) A medical diagnostic test for a certain disease will yield either a positive or a negative reaction. If you have the disease, there is 0.95 chance that the test result will be positive, while if you do not have the disease, there is 0.90 chance that the test will be negative. It is estimated that 2% of the population has the disease.

(i) Find the probability that you have the disease, given that you have a positive reaction.

(ii) Find the probability that you do not have the disease, given that you have a negative reaction.

[20 marks]

(c) Suppose that 10% of the students in a school are infected with a certain type of virus.

(i) What is the probability that the number of students having the virus in a random sample of 250 will be between 20 and 30?

(ii) What is your assumption in i)?

(iii) After sampling 250 students, suppose that 35 students are found to have the virus. Does this contradict the hypothesis that the population proportion is 10%? Justify your answer by using the terminologies of hypotheses testing.

[35 marks]

(d) Complete the following statements (more than one word may be needed):

(i) If we take all possible samples (of a given large sample size) from a population, then the distribution of sample means tend to be_____.

(ii) The larger the sample size, other things remaining equal, the _____ the confidence interval.

...3/-

(iii) The larger the confidence coefficient, other things remaining equal, the _____ the confidence interval.

(iv) The statement "If random samples of a fixed size are drawn from any population (regardless of the form of the population distribution), as n becomes larger, the distribution of samples means approaches normality", is known as the _____.

(v) By failing to reject a null hypothesis that is false, one makes a _____ error.

[25 marks]

1. (a) *Seorang pengurus hotel berminat untuk mengetahui purata tempoh penginapan tetamu yang menginap di hotelnya. Adalah mustahil untuk memeriksa kesemua rekod penginapan, maka beliau telah memilih secara rawak 100 tetamu yang menginap di hotel tersebut semenjak setahun lalu dan mendapatkan purata tempoh penginapan mereka.*

(i) Terangkan mengenai populasi yang berkaitan dengan masalah ini.

(ii) Apakah sampelnya?

(iii) Apakah pembolehubah yang mesti diukur?

(iv) Berikan komen anda mengenai teknik persampelan ini.

[20 markah]

(b) *Ujian diagnostik perubatan bagi suatu jenis penyakit akan menghasilkan samada reaksi positif atau negatif. Jika anda mengidap peyakit tersebut, terdapat 0.95 peluang yang keputusannya akan positif sementara jika anda tidak mengidap penyakit tersebut, terdapat 0.90 peluang yang ujian itu akan negatif.*

(i) Dapatkan kebarangkalian yang anda mengidap penyakit tersebut, diberi bahawa anda memberikan reaksi positif.

(ii) Dapatkan kebarangkalian yang anda tidak mengidap penyakit tersebut, diberi bahawa anda memberikan reaksi negatif.

[20 markah]

(c) *Andaikan bahawa 10% daripada pelajar di sebuah sekolah telah dijangkiti sejenis virus.*

(i) Apakah kebarangkalian bahawa bilangan pelajar yang dijangkiti virus tersebut adalah antara 20 orang hingga 30 orang daripada sampel rawak bersaiz 250?

(ii) Apakah anggapan anda di bahagian i)?

...4/-

(iii) *Selepas mensampel 250 pelajar, didapati 35 pelajar telah dijangkiti virus. Adakah ini bercanggah dengan hipotesis bahawa kadaran populasi 10%? Sahkan jawapan anda dengan menggunakan terminologi-terminologi pengujian hipotesis.*

[35 markah]

(d) *Lengkapkan penyataan-penyataan berikut (mungkin memerlukan lebih daripada satu perkataan):*

(i) *Jika kita mengambil kesemua sampel yang mungkin (bagi saiz sampel tertentu yang besar) daripada suatu populasi, maka taburan min-min sampel itu cenderung menjadi* _____.

(ii) *Semakin besar saiz sampel, yang lainnya kekal, semakin* _____ *selang keyakinan.*

(iii) *Semakin besar koefisien keyakinan, yang lainnya kekal, semakin* _____ *selang keyakinan.*

(iv) *Penyataan "Jika sampel-sampel rawak dengan saiz tertentu diambil daripada populasi (tanpa mengambil kira bentuk taburan populasi), semakin besar n, taburan min-min sampel menghampiri kenormalan", dikenali sebagai* _____.

(v) *Dengan tidak menolak hipotesis nol yang salah, seseorang membuat ralat* _____.

[25 markah]

2. (a) A scientist wishes to estimate the mean weight of rats when injected with a specified dosage of serum based on the weight measurements from four rats, X_1, X_2, X_3 and X_4. He considers three estimators:

$$M_1 = \frac{X_1 + X_2 + X_3 + X_4}{4},$$

$$M_2 = \frac{0.5X_1 + X_2 + X_3 + 0.5X_4}{3}$$

and $$M_3 = \frac{2X_1 + X_2 + X_3 + 2X_4}{6}.$$

(i) Determine if these estimators are unbiased.

(ii) Compute the variances of these estimators. State your assumption.

(iii) Which estimator should be preferred? Why?

[35 marks]

...5/-

(b) The city housing department wants to estimate the average rent for rent-controlled apartments. From past results, the rent for controlled apartments ranged from RM300 to RM1200 per month. They need to determine the number of renters to include in the survey.

(i) In order to estimate the average rent to within RM50 using a 95% confidence interval, what sample size is required?

(ii) If the level of confidence is increased to 99% with the average rent estimated to within RM50, what sample size is required?

(iii) If the level of confidence is increased to 99% with the average rent estimated to within RM25, is sample size required be twice the size in ii)?

[30 marks]

(c) From the label of the packet, the minimum weight of a chocolate bun is claimed to be 200gm. You are not satisfied with the factory's claim and decide to conduct a test. From 12 packets of bun, you obtain the following weights:

204 201 196 200 198 199 201 207 195 198
205 202

(i) Do the data support the factory's claim? Test at $\alpha = 0.05$.
State the assumption you make about the population distribution.

(ii) Construct a 95% confidence interval for the mean weight of the chocolate bun.

(iii) Provide an interpretation of the confidence interval computed in ii).

[35 marks]

2. (a) *Seorang saintis ingin menganggarkan berat purata tikus selepas disuntik dengan serum (mengikut dos tertentu) berdasarkan berat empat ekor tikus, X_1, X_2, X_3 dan X_4. Beliau mempertimbangkan tiga penganggar:*

$$M_1 = \frac{X_1 + X_2 + X_3 + X_4}{4},$$

$$M_2 = \frac{0.5X_1 + X_2 + X_3 + 0.5X_4}{3}$$

dan $$M_3 = \frac{2X_1 + X_2 + X_3 + 2X_4}{6}.$$

...6/-

(i) *Tentukan samada penganggar-penganggar tersebut saksama.*

(ii) *Kira varians bagi setiap penganggar. Nyatakan anggapan yang telah anda buat.*

(iii) *Penganggar mana yang patut dipilih? Kenapa?*

[35 markah]

(b) *Jabatan Perumahan sebuah bandar ingin menganggarkan purata sewa bagi pangsapuri-pangsapuri yang dikawal-sewa. Daripada rekod lepas, sewa bagi pangsapuri-pangsapuri ini adalah antara RM300 hingga RM1200 sebulan. Mereka perlu tentukan bilangan penyewa untuk dimasukkan dalam kajian ini.*

(i) *Supaya anggaran purata sewa dalam lingkungan RM50 menggunakan selang keyakinan 95%, berapakah saiz sampel yang diperlukan?*

(ii) *Jika aras keyakinan ditingkatkan kepada 99% dengan anggaran purata sewa dalam lingkungan RM50, berapakah saiz sampel yang diperlukan?*

(iii) *Jika aras keyakinan ditingkatkan kepada 99% dengan anggaran purata sewa dalam RM25, adakah saiz sampel yang diperlukan dua kali ganda saiz sampel dalam ii)?*

[30 markah]

(c) *Daripada label sebungkus roti cokelat, berat minimum adalah 200gm. Anda tidak berpuas hati dengan dakwaaan kilang tersebut dan bercadang untuk menjalankan satu ujian. Daripada 12 bungkus roti cokelat, anda menimbang dan mendapati beratnya adalah seperti berikut:*

204 201 196 200 198 199 201 207 195 198 205 202

(i) *Adakah data tersebut menyokong dakwaaan kilang? Uji pada α = 0.05. Nyatakan anggapan yang anda buat terhadap taburan populasi.*

(ii) *Bina selang keyakinan 95% bagi berat purata roti cokelat tersebut.*

(iii) *Beri interpretasi terhadap selang keyakinan dalam ii).*

[35 markah]

...7/-

3. (a) It is anticipated that using mental arithmetic will more effectively improve the understanding of elementary school children in Mathematics. To test this conjecture, 14 children are divided at random into two groups of 7 each. One group is instructed using the standard method and the other group is instructed using mental arithmetic. The children's scores on a Mathematics test are shown in the following table:

	Mathematics Test Scores						
Standard Method	80	72	65	68	53	70	75
Mental Arithmetic	75	65	80	85	79	90	70

(i) Do the data provide strong evidence that mental arithmetic improve the performance of children in Mathematics? Test with $\alpha = 0.05$.

(ii) What is the p-value for this test?

(iii) State the assumption you make for the population distribution.

[40 marks]

(b) Grades in a Statistics course and an Operations Research course taken simultaneously by a group of students were as follows:

		Operations Research Grade			
		A	B	C	Others
Statistics Grade	A	25	6	17	13
	B	17	16	15	6
	C	18	4	18	10
	Others	10	8	61	49

(i) Are the grades in Statistics and Operations Research related? Use $\alpha = 0.01$ in reaching your conclusion.

(ii) What is the p-value for this test?

[40 marks]

(c) A study was conducted on 90 adult male patients following a new treatment for congestive heart failure. One of the variables measured on the patients was the increase in exercise capacity (in minutes) over a 4-week treatment period. The previous treatment regime had produced an average increase of $\mu = 2$ minutes. The researchers wanted to evaluate whether the new treatment had increased the value of μ in comparison to the previous treatment. The data yielded $\bar{y} = 2.17$ and $s = 1.05$.

(i) What are the Type I and Type II errors related to this problem?

(ii) What is the probability of making a Type II error if the actual value of μ is 2.1?

[20 marks]

...8/-

3. (a) *Penggunaan aritmetik mental dikatakan lebih berkesan dalam meningkatkan kefahaman pelajar dalam Matematik. Bagi menguji penyataan ini, 14 orang pelajar dibahagikan kepada dua kumpulan yang terdiri daripada 7 orang setiap kumpulan. Satu kumpulan diajar dengan kaedah biasa dan satu kumpulan lagi dengan aritmetik mental. Markah pelajar-pelajar ini dalam ujian Matematik ditunjukkan dalam jadual di bawah:*

	Markah ujian Matematik						
Kaedah biasa	*80*	*72*	*65*	*68*	*53*	*70*	*75*
Aritmetik mental	*75*	*65*	*80*	*85*	*79*	*90*	*70*

(i) *Adakah data ini memberikan bukti kukuh bahawa aritmetik mental meningkatkan pencapaian pelajar dalam Matematik? Uji dengan* $\alpha = 0.05$.

(ii) *Apakah nilai-p bagi ujian ini?*

(iii) *Nyatakan anggapan yang anda buat bagi taburan populasi.*

[40 markah]

(b) *Gred bagi kursus Statistik dan kursus Penyelidikan Operasi yang diambil secara serentak oleh sekumpulan pelajar adalah seperti berikut:*

		Gred Penyelidikan Operasi			
		A	B	C	Lain-lain
Gred	A	25	6	17	13
Statistik	B	17	16	15	6
	C	18	4	18	10
	Lain-lain	10	8	61	49

(i) *Adakah terdapat perkaitan antara gred-gred Statistik dan Penyelidikan Operasi? Gunakan* $\alpha = 0.01$ *dalam memberikan kesimpulan.*

(ii) *Apakah nilai-p bagi ujian ini?*

[40 markah]

(c) *Satu kajian ke atas 90 lelaki dewasa berkaitan kaedah rawatan terkini untuk kegagalan jantung kongestif telah dijalankan. Satu daripada pembolehubah yang diukur adalah pertambahan dalam kapasiti latihan (dalam minit) bagi tempoh empat minggu rawatan. Kaedah rawatan lama menghasilkan pertambahan purata sebanyak* $\mu = 2$ *minit. Penyelidik berhasrat untuk menilai samada kaedah rawatan terkini ini meningkatkan nilai* μ *berbanding kaedah rawatan lama. Daripada data,* $\bar{y} = 2.17$ *dan* $s = 1.05$.

...9/-

(i) *Apakah ralat jenis I dan ralat jenis II bagi masalah ini?*

(ii) *Apakah kebarangkalian membuat ralat jenis II jika nilai sebenar μ ialah 2.1?*

[20 markah]

4. (a) The compression strength of concrete is being studied and four different mixing techniques are being investigated. The following data have been collected.

Mixing Technique	Compressive	Strength		(psi)
1	3129	3000	2865	2890
2	3200	3300	2975	3150
3	2800	2900	2985	3050
4	2600	2700	2600	2765

An experimenter computed the SS_T and SS_{TR} and obtained that

$SS_T = 643648.438$

and $SS_{TR} = 489740.1875$.

(i) Present the ANOVA table for these data.

(ii) With $\alpha = 0.05$, test the hypothesis that mixing techniques affect the strength of the concrete.

[30 marks]

(b) In an experiment designed to determine the relationship between the dose of a compost fertilizer x and the yield of a crop y, the following summary statistics are recorded:

$$n = 15 \qquad \bar{x} = 10.8 \qquad \bar{y} = 122.7$$
$$S_{xx} = 70.6 \qquad S_{yy} = 98.5 \qquad S_{xy} = 68.3 .$$

Assume there is a linear relationship.

(i) Find the equation of the least squares regression line.

(ii) Compute the error sum of squares and estimate σ^2.

(iii) Predict the yield of crop when the dose of a compost fertilizer is 12 and obtain a 95% prediction interval.

[35 marks]

(c) A scientist wishes to examine the relationship between the weight and systolic blood pressure. To perform the test, 26 males in the age group 25 to 30 are randomly selected and their weights and systolic blood pressures are recorded (Assume that weight and blood pressure are jointly normally distributed). A regression analysis is performed using the Minitab package, yielding the following output:

...10/-

Regression Analysis

The regression equation is
Systolic BP = *a* + *b*.Weight

Predictor	Coef	StDev	T	P
Constant	69.10	12.91	5.35	0.000
Weight	0.41942	0.07015	5.98	0.000

S = 8.681 R-Sq = *c* R-Sq(adj) = 58.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	A	2693.6	E	G	0.000
Residual Error	B	D	F		
Total	C	4502.2			

(i) Obtain the values of A, B, C, D, E, F and G.

(ii) Write the regression equation relating systolic blood pressure to weight.

(iii) Interpret the slope coefficient.

(iv) Find the coefficient of determination. What does it indicate about the predictive value of systolic blood pressure?

[35 marks]

4. *(a) Suatu kajian berkenaan kekuatan tekanan konkrit dilakukan dan empat teknik campuran diselidiki. Data berikut telah diperolehi.*

Teknik Campuran	*Kekuatan Tekanan (psi)*			
1	*3129*	*3000*	*2865*	*2890*
2	*3200*	*3300*	*2975*	*3150*
3	*2800*	*2900*	*2985*	*3050*
4	*2600*	*2700*	*2600*	*2765*

Penyelidik mengira nilai SS_T *dan* SS_{TR} *dan memperolehi*

$SS_T = 643648.438$

dan $SS_{TR} = 489740.1875.$

(i) Bina jadual ANOVA bagi data ini.

(ii) Dengan $\alpha = 0.05$*, uji hipotesis bahawa teknik campuran mempengaruhi kekuatan konkrit.*

[30 markah]

...11/-

(b) *Dalam eksperimen untuk menentukan hubungan antara dos sejenis baja x dan hasil tanaman y, ringkasan statistik berikut direkodkan:*

$$n = 15 \qquad \bar{x} = 10.8 \qquad \bar{y} = 122.7$$
$$S_{xx} = 70.6 \qquad S_{yy} = 98.5 \qquad S_{xy} = 68.3.$$

Anggapkan hubungan adalah linear.

(i) *Dapatkan persamaan garis regresi kuasa dua terkecil.*

(ii) *Kira ralat hasil tambah kuasa dua dan anggarkan* σ^2.

(iii) *Ramalkan hasil tanaman bila dos baja adalah 12 dan dapatkan selang ramalan 95%.*

[35 markah]

c) *Seorang saintis ingin mengkaji hubungan antara berat dan tekanan darah sistolik. Bagi menjalankan ujian, 26 lelaki berumur antara 25 tahun hingga 30 tahun dipilih secara rawak dan berat serta tekanan darah sistolik mereka direkodkan (anggap bahawa taburan berat dan tekan darah tercantum normal). Analisis regresi dilakukan menggunakan perisian Minitab dan berikut adalah outputnya:*

Regression Analysis

The regression equation is
Systolic BP = *a* + *b*.Weight

Predictor	Coef	StDev	T	P
Constant	69.10	12.91	5.35	0.000
Weight	0.41942	0.07015	5.98	0.000

S = 8.681 R-Sq = *c* R-Sq(adj) = 58.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	A	2693.6	E	G	0.000
Residual Error	B	D	F		
Total	C	4502.2			

(i) *Dapatkan nilai A, B, C, D, E, F dan G.*

(ii) *Tuliskan persamaan regresi menghubungkan tekanan darah sistolik dengan berat.*

(iii) *Interpretasikan pekali kecerunan.*

(iv) *Dapatkan pekali penentuan. Apakah yang ditunjukkannya tentang nilai ramalan tekanan darah sistolik?*

[35 markah]

- 000 O 000 -