UNIVERSITI SAINS MALAYSIA

First Semester Examination
Academic Session 2004/2005

October 2004

MST 567 – CATEGORICAL DATA ANALYSIS
[ANALISIS DATA BERKATEGORI]

Duration : 3 hours

Masa : 3 jam

Please check that this examination paper consists of **TWENTY ONE [21]** pages of printed material before you begin the examination.

Answer **all FIVE** questions.

*[Sila pastikan bahawa kertas peperiksaan ini mengandungi **DUA PULUH SATU [21]** muka surat yang bercetak sebelum anda memulakan peperiksaan ini].*

Jawab **semua lima** soalan.

...2/-

1. (a) The following data show the numbers of ignitions with two types of explosives fired in a gas chamber.Use an exact conditional test treating the total number of ignitions as fixed, and approxomate tests based on the X^2 statistic with and without the use of Yate's continuity correction, to determine whether there is real evidence that the probability of ignition differs between the two explosives. Contrast the observed significance level of the exact conditional test with the approximating significance levels of the tests based on the X^2 statistic.

	Type of Explosives		Totals
	Explosive A	Explosive B	
Number of ignitions	3	12	15
Number of non ignitions	7	3	10
Totals	10	15	25

[55 marks]

(b) In an investigation of a new treatment for the common cold two randomly selected groups of 50 male sufferers and two randomly selected groups of 50 female sufferers were used as subjects. One from each pair of groups was used as a control group, receiving a placebo, and the other groups received the new treatment. After 36 hours, the condition of the subjects was noted, with the following results.

	Male			Female		
	Improved	No improvement	Total	Improved	No improvement	Total
Treatment	34	16	50	28	22	50
Control	29	21	50	23	27	50
Total	63	37	100	51	49	100

A log-linear model
$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{23(jk)}$,
(where i denotes sex, j treatment and k subject condition) is fitted to these data.
(i) Explain whether such a model fits the data satisfactorily,
(ii) Interpret the parameters in the model ,
based on the given computer output **(see Appendix A 1(b)).**

[45 marks]

1. (a) Data berikut menunjukkan bilangan nyalaan oleh dua jenis peledak yang dibakar dalam suatu bilik gas. Gunakan suatu ujian bersyarat tepat yang mengambilkira jumlah bilangan nyalaan sebagai tetap, dan ujian-ujian penghampiran berdasarkan statistik X^2 dengan dan tanpa menggunakan pembetulan keselanjaraan Yate's, untuk menentukan sama ada wujud bukti nyata bahawa kebarangkalian nyalaan berbeza antara dua jenis peledak itu. Kontraskan aras kesignifikanan tercerap ujian bersyarat tepat dengan aras kesignifikanan hampiran ujian-ujian berdasarkan kepada statistik X^2.

	Jenis Peledak		*Jumlah*
	Peledak A	*Peledak B*	
Bilangan Nyalaan	*3*	*12*	*15*
Bilangan Gagal Nyalaan	*7*	*3*	*10*
Jumlah	*10*	*15*	*25*

[55 markah]

(b) *Dalam suatu kajian tentang rawatan baru terhadap selsema dua kumpulan 50 orang pengidap lelaki dan dua kumpulan 50 orang pengidap perempuan telah dipilih secara rawak sebagai unit kajian. Seorang dari setiap pasangan kumpulan itu telah digunakan sebagai kumpulan kawalan, menerima placebo, dan kumpulan satu lagi menerima rawatan baru. Selepas 36 jam, keadaan unit kajian itu telah diperhatikan, menghasilkan keputusan berikut.*

	Lelaki		Jumlah
	Ada Peningkatan	Tiada Peningkatan	
Rawatan Baru	34	16	50
Kawalan	29	21	50
Jumlah	63	37	100

Perempuan		Jumlah
Ada Peningkatan	Tiada Peningkatan	
28	22	50
23	27	50
51	49	100

Suatu model log-linear

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{23(jk)},$$

(dengan i menandai jantina, j rawatan and k keadaan subjek) telah disuaikan kepada data.

(i) Terangkan sama ada model sedemikian secocok betul dengan data,

(ii) Tafsirkan parameter-parameter dalam model,

*berasaskan output komputer yang diberi (**Lihat lampiran A 1(b)**).*

[45 markah]

2. (a) Describe the two probability models for the observations in a 2 x 2 contingency table with entries $\{ n_{ij} \}$, which are approriate

(i) when the values in both margins of the table are random, and

(ii) when the values in one margin are fixed.

For each model give a brief description of circumstances in which it might be employed and define the approriate 'no association' hypothesis.

[35 marks]

(b) In a certain district 466 men were graded according to their drinking habits and the amount of overcrowding in their homes as follows.

Amount of overcrowding	Drinking Habits		
	Does no drink (x_i)	Drinks ($n_i - x_i$)	Total (n_i)
Under 0.5	1	0	1
0.5 and under 1.0	69	21	90
1.0 and under 1.5	117	48	165
1.5 and under 2.0	76	51	127
2.0 and under 2.5	35	28	63
2.5 and under 3.0	5	5	10
3.0 and under 3.5	3	5	8
3.5 and under 4.0	0	2	2
Total (n_i)	306	160	466

(i) Plot the observed sample proportions $p_i = (x_i/n_i)$ against the midpoint z_i of the class intervals of the average number of people per room.

(ii) Assuming a linear model of the form $P_i = \alpha + \beta z_i$, where P_i is the probability that a man in the ith group does not drink, obtain the ordinary least squares estimates of α and β and contrast the fitted probabilities

$\hat{P}_i = \hat{\alpha} + \hat{\beta} z_i$ with the observed sample proportions.

(iii) Make a formal test of fit of the linear model.

[65 marks]

2. *(a)* *Perihalkan dua model kebarangkalian untuk cerapan dalam jadual kekontingenan 2 x 2 dengan pemasukan $\{ n_{ij} \}$, yang sesuai*

(i) bilamana nilai-nilai sut pada jadual adalah rawak, dan

(ii) bila mana nilai-nilai pada satu sut tetap.

Untuk setiap model itu berikan penjelasan secara ringkas situasi yang boleh ianya digunakan dan takrifkan hipotesis sesuai untuk 'tidak wujud kesekutuan'

[35 markah]

(b) *Dalam sebuah daerah 466 orang lelaki telah digredkan menurut tabiat minum dan bilangan lebihan penghuni dalam rumah mereka seperti berikut.*

lebihan penghuni	*Tabiat minum*		
	Tidak minum (x_i)	*Minum ($n_i - x_i$)*	*Jumlah (n_i)*
Dibawah 0.5	*1*	*0*	*1*
0.5 dan dibawah 1.0	*69*	*21*	*90*
1.0 dan dibawah 1.5	*117*	*48*	*165*
1.5 dan dibawah 2.0	*76*	*51*	*127*
2.0 dan dibawah 2.5	*35*	*28*	*63*
2.5 dan dibawah 3.0	*5*	*5*	*10*
3.0 dan dibawah 3.5	*3*	*5*	*8*
3.5 dan dibawah 4.0	*0*	*2*	*2*
Jumlah (n_i)	*306*	*160*	*466*

(i) *Plotkan kadaran sampel tercerap* $p_i = (x_i/n_i)$ *melawan titik tengah* z_i *bagi selang kelas purata bilangan orang per bilik.*

(ii) *Dengan menganggapkan suatu model linear berbentuk* $P_i = \alpha + \beta z_i$, *dengan* P_i *ialah kebarangkalian seorang lelaki dalam kumpulan ke-i bukan peminum, dapatkan anggaran kuasa terkecil biasa bagi* α *dan* β *dan kontraskan anggaran kebarangkalian* $\hat{P}_i = \hat{\alpha} + \hat{\beta} z_i$ *dengan kadaran sampel tercerap.*

(iii) *Jalankan suatu ujian formal terhadap kesesuaian model linear itu.*

[65 markah]

3. (a) A single random sample of n individuals is drawn from a population in which the individuals are categorised according to two factors A and B with r and s levels respectively. If N_{ij} and P_{ij} represents the frequency and probability of individuals with the classification (A_i, B_j) respectively,

(i) What is the joint distribution of N_{ij} ?

(ii) For such table in (a), we may construct (r-1)x(s-1) 2x2 tables in which cell (r,s) is always present. Associated with these tables, we may define (r-1)(s-1) odds ratios given by

$$\theta_{ij} = \frac{P_{ij}P_{rs}}{P_{is}P_{rj}} \quad , \quad i = 1,2,3,\ldots,r-1, \quad j = 1,2,3,\ldots,s-1 \; .$$

Show that with an appropriate substitution the joint distribution of N_{ij} is given by

$$n!\exp\left\{nK + \sum_{i=1}^{r-1} n_{i.}K_i^A + \sum_{j=1}^{s-1} n_{.j}K_j^B + \sum_{i=1}^{r-1}\sum_{j=1}^{s-1} n_{ij}\lambda_{ij}\right\} / \prod_{(i,j)=(1,1)}^{(r,s)} n_{ij}!$$

where $\lambda_{ij} = \log\left(\frac{P_{ij}P_{rs}}{P_{is}P_{rj}}\right)$, $i = 1,2, \ldots, r-1,\; j = 1,2, \ldots s-1$,

$K_i^A = \log\left(\frac{P_{is}}{P_{rs}}\right)$, $i = 1,2, \ldots, r-1$, $K_j^B = \log\left(\frac{P_{rj}}{P_{rs}}\right)$, $j = 1,2, \ldots, s-1$

and $K = \log P_{rs}$.

(iii) Hence, show that the joint distribution of $\{N_{ij}\}$ conditional on $\{N_{i+} = n_{i+}\}$ and $\{N_{+i} = n_{+i}\}$, where $\sum_{i=1}^{r} N_{ij} = N_{+j}$ *and* $\sum_{j=1}^{s} N_{ij} = N_{i+}$ with n_{+j} and n_{i+} defined similarly, is given by

$$\frac{\exp\left(\sum_{i=1}^{r-1}\sum_{j=1}^{s-1} n_{ij}\lambda_{ij}\right) / \prod_{(i,j)=(1,1)}^{(r,s)} n_{ij}!}{\sum_{\{n_{ij}\}}^{*}\left\{\exp\left(\sum_{i=1}^{r-1}\sum_{j=1}^{s-1} n_{ij}\lambda_{ij}\right) / \prod_{(i,j)=(1,1)}^{(r,s)} n_{ij}!\right\}}$$

where $\sum_{\{n_{ij}\}}^{*}$ denotes the sum over all non-negative integers { n_{ij} } satisfying

$$\sum_{j=1}^{s} n_{ij} = n_{i+} \quad , i = 1,2,\ldots, r \qquad \text{and} \quad \sum_{i=1}^{r} n_{ij} = n_{+j} \quad , j = 1,2,\ldots, s \; .$$

[55 marks]

(b) A sample of patients with a condition requiring treatment are classified by age and the degree of success of the treatment. The following results are obtained.

Success of Treatment	**Age of Patients (years)** Under 30	30 to 50	over 50 years	Totals
Full Recovery	52	90	20	162
Partial Recovery	40	47	10	97
No Effect	56	118	48	222
Totals	148	155	78	481

(i) Perform a large sample overall tests of the hypothesis of total independence of the two factors using X^2 and G statistics.

(ii) Partition the table into four 2x2 tables and obtain the modified unconditional maximum likelihood estimates of the log odds ratios for the four tables together with their approximate estimated standard deviations. Discuss the interpretation of your results.

[45 marks]

3. (a) *Satu sampel rawak n individu telah dipilih dari suatu populasi dengan individu-individu itu dikategorikan menurut dua faktor A and B masing-masing dengan aras r dan s . Jika N_{ij} dan P_{ij} masing-masing mewakili kekerapan dan kebarangkalian individu mempunyai kelasifikasi (A_i , B_j) ,*

(i) *Apakah taburan tercantum N_{ij} ?*

(ii) *Untuk jadual dalam (a) , kita boleh bina (r-1)x(s-1) jadual 2x2 dengan sel (r,s) sentiasa wujud. Sehubungan dengan jadual ini, kita boleh takrifkan (r-1)(s-1) nisbah ods diberi oleh*

$$\theta_{ij} = \frac{P_{ij}P_{rs}}{P_{is}P_{rj}} \quad , \quad i = 1,2,3,\ldots,r-1, \quad j = 1,2,3,\ldots,s-1 \; .$$

Dengan menggunakan penggantian tertentu, tunjukkan taburan tercantum bagi N_{ij} diberi oleh

$$n\,!\exp\left\{nK+\sum_{i=1}^{r-1}n_{i.}K_i^A+\sum_{j=1}^{s-1}n_{.j}K_j^B+\sum_{i=1}^{r-1}\sum_{j=1}^{s-1}n_{ij}\lambda_{ij}\right\}/\prod_{(i,j)=(1,1)}^{(r,s)}n_{ij}!$$

dengan $\lambda_{ij}=\log\left(\dfrac{P_{ij}P_{rs}}{P_{is}P_{rj}}\right)$, $i=1,2,\ \ldots,\ r-1,\quad j=1,2,\ \ldots\ s-1$,

$K_i^A=\log\left(\dfrac{P_{is}}{P_{rs}}\right)$, $i=1,2,\ \ldots,r-1$, $K_j^B=\log\left(\dfrac{P_{rj}}{P_{rs}}\right)$, $j=1,2,\ \ldots,s-1$

dan $K=\log P_{rs}$.

(iii) *Oleh itu tunjukkan taburan tercantum* $\{N_{ij}\}$ *dengan syarat* $\{N_{i+}=n_{i+}\}$ *dan* $\{N_{+i}=n_{+i}\}$, *dimana* $\sum_{i=1}^{r}N_{ij}=N_{+j}$ *and* $\sum_{j=1}^{s}N_{ij}=N_{i+}$ *dengan* n_{+j} *dan* n_{i+} *ditakrifkan dengan cara yang sama diberi oleh*

$$\frac{\exp\left(\sum_{i=1}^{r-1}\sum_{j=1}^{s-1}n_{ij}\lambda_{ij}\right)/\prod_{(i,j)=(1,1)}^{(r,s)}n_{ij}\,!}{\sum_{\{n_{ij}\}}^{*}\left\{\exp\left(\sum_{i=1}^{r-1}\sum_{j=1}^{s-1}n_{ij}\lambda_{ij}\right)/\prod_{(i,j)=(1,1)}^{(r,s)}n_{ij}\,!\right\}}$$

dimana $\sum_{\{n_{ij}\}}^{*}$ *menandai hasil tambah terhadap semua integer tak negatif* $\{n_{ij}\}$ *yang memenuhi*

$\sum_{j=1}^{s}n_{ij}=n_{i+}$, $i=1,2,\ldots,r$ *dan* $\sum_{i=1}^{r}n_{ij}=n_{+j}$, $j=1,2,\ldots,s$.

[55 markah]

(b) *Satu sampel pesakit dalam keadaan memerlukan rawatan dikelasifikasikan mengikut umur dan tahap kejayaan rawatan tersebut. Keputusan berikut telah diperoleh.*

Kejayaan Rawatan	***Umur Pesakit (tahun)***			*Jumlah*
	Dibawah 30 tahun	*30 to 50 tahun*	*melebihi 50 tahun*	
Sembuh Sepenuhnya	*52*	*90*	*20*	*162*
Separa Sembuh	*40*	*47*	*10*	*97*
Tiada Kesan	*56*	*118*	*48*	*222*
Jumlah	*148*	*255*	*78*	*481*

(i) *Jalankan suatu ujian sampel besar untuk menguji hipotesis ketakbersandaran keseluruhan bagi dua faktor itu dengan menggunakan statistik X^2 dan G.*

(ii) *Partisikan jadual itu kepada empat jadual 2x2 dan dapatkan anggaran-anggaran kebolehjadian maksimum tak bersyarat terpinda masing-masing bagi log nisbah odsnya bersama dengan anggaran hampiran sisihan piawai masing-masing. Bincangkan tafsiran keputusan yang anda perolehi.*

[45 markah]

4. (a) In an investigation into the effectiveness of a new processing treatment, independent trials were made in two factories comparing the new processing treatment with the standard method of processing. After processing, the units were classified as either defective or satisfactory and the following data were obtained.

Type of Units	Factory 1		Factory 2	
	New Treatment	Old Treatment	New Treatment	Old Treatment
Defective Units	4	11	6	9
Ssatisfactory Units	26	14	24	16

Suppose that the probability $P_{1,k}$ of a unit being defective, which is treated with the new method in factory k is given by

$$\log\left\{\frac{P_{1,k}}{1-P_{1,k}}\right\} = F_k + \Delta$$

and that the corresponding probability $P_{2,k}$ for a unit having the standard processing treatment is given by

$$\log\left\{\frac{P_{2,k}}{1-P_{2,k}}\right\} = F_k - \Delta$$

If X_k and Y_k are random variables representing the numbers of defectives in the kth factory in the groups of units processed by the new and standard methods respectively, show that

$$\Pr(X_k = x | X_k + Y_k = t) = \frac{\binom{30}{x}\binom{25}{t-x}}{\sum_{all\ u}\binom{30}{u}\binom{25}{t-u}e^{2\Delta u}} .$$

(b) If $X = X_1 + X_2$ denotes the total number of defectives for the new processing treatment, show that conditional on $X_1 + Y_1 = t_1$, $X_2 + Y_2 = t_2$,

$$E(X \mid t_1, t_2 ; \Delta = 0) = 6(t_1 + t_2)/11 .$$

(You may use the result that $\sum_{all\ u}\binom{a}{u}\binom{b}{t-u} = \binom{a+b}{t}$)

(c) Describe how in principle an exact conditional test based on X could be made to test the hypothesis of no difference between treatments (no detailed calculations are required).

[100 marks]

4. (a) *Dalam satu kajian terhadap keberkesanan suatu pemprosesan rawatan baru, ujian-ujian tak bersandar telah dijalankan dalam dua kilang untuk membandingkan rawatan pemprosesan baru dengan kaedah pemprosesan piawai. Selepas pemprosesan, unit-unit itu telah dikelasifikasikan sama ada cacat atau memuaskan dan data berikut telah diperoleh.*

Jenis Unit	*Kilang 1*		*Kilang 2*	
	Rawatan Baru	*Rawatan Lama (Piawai)*	*Rawatan Baru*	*Rawatan Lama (Piawai)*
Cacat	*4*	*11*	*6*	*9*
Memuaskan	*26*	*14*	*24*	*16*

Andaikan $P_{1,k}$ *ialah kebarangkalian suatu unit itu cacat, yang menerima rawatan baru dalam kilang k diberi oleh*

$$\log\left\{\frac{P_{1,k}}{1 - P_{1,k}}\right\} = F_k + \Delta$$

dan kebarangkalian sepadan $P_{2,k}$ *untuk unit yang menerima rawatan pemprosesan piawai pula diberi oleh*

$$\log\left\{\frac{P_{2,k}}{1 - P_{2,k}}\right\} = F_k - \Delta$$

Jika X_k *dan* Y_k *adalah masing-masing pemboleh ubah rawak mewakili bilangan cacat dalam kilang ke-k bagi kumpulan unit yang diproses oleh kaedah baru dan kaedah lama, tunjukkan*

$$\Pr(X_k=x|X_k+Y_k=t)=\frac{\binom{30}{x}\binom{25}{t-x}}{\sum_{all\,u}\binom{30}{u}\binom{25}{t-u}e^{2\Delta u}}\ .$$

(b) *Jika* $X = X_1 + X_2$ *menandai jumlah bilangan cacat untuk rawatan pemprosesan baru, tunjukkan bahawa dengan syarat* $X_1 + Y_1 = t_1$, $X_2 + Y_2 = t_2$,

$$E(X \mid t_1, t_2 ; \Delta = 0) = 6(t_1 + t_2)/11\ .$$

(Anda boleh menggunakan keputusan bahawa $\sum_{all\,u}\binom{a}{u}\binom{b}{t-u}=\binom{a+b}{t}$ *)*

(c) *Perihalkan bagaimana secara perinsipnya ujian bersyarat tepat ber dasarkan kepada X boleh dijalankan untuk menguji hipotesis bahawa tiada wujud perbezaan antara rawatan-rawatan itu (pengiraan terperinci tidak perlu)*

[100 markah]

5. (a) In a study on alcohol consumption conducted by a government agency, a random sample of 3603 individuals above the age of 20 years were selected for interviewing. Four variables were considered in the study:

Alcohol consumption – 1 unit a day , 2 units a day

Social group -- Top class, Middle class, Lower class, Bottom class

Marriage status -- Married , Unmarried

Age -- 20 – 39 years , 40 – 69 years

(1 unit is roughly one 33 cl bottle of beer of strength 4% of its volume or its equivalent).

The resulting four-way contingency table is shown below.

Alcohol Consumption	Social group	Marriage Status	Age 20 – 39	Age 40 – 69
1 unit a day	Top	Married	43	38
		Unmarried	11	6
	Middle	Married	112	158
		Unmarried	35	33
	Lower	Married	195	169
		Unmarried	76	44
	Bottom	Married	135	216
		Unmarried	72	61
2 units a day	Top	Married	149	134
		Unmarried	37	28
	Middle	Married	200	225
		Unmarried	80	45
	Lower	Married	296	238
		Unmarried	195	50
	Bottom	Married	131	212
		Unmarried	130	49

(i) Describe the process of obtaining the final model using the backward elimination and state the final model based on the computer output in **Appendix B 5(a)**

(ii) Interpret the highest interaction parameter estimates in the final model.

[50 marks]

(b) For 2 x 2 tables, a measure of association similar to a correlation coefficient is Yule's Q which is defined as

$$Q = \frac{p_{11}\,p_{22} - p_{12}p_{21}}{p_{11}p_{22} + p_{12}p_{21}}\ .$$

Find Q in terms of the odds ratio. Show that Q must lie between –1 and 1 .

[25 marks]

(c) Consider an r x 2 table formed when independent samples of size $n_{.1}$ and $n_{.2}$ are drawn from populations having the same r classifying levels. Let N_{ij} be a random variable representing the frequency of individuals in the cell (i, j) , i = 1 , 2 , j = 1 , 2 . Show that for the case equal sample sizes with $n_{.1} = n_{.2}$, the X^2 statistic takes the form

$$X^2 = \sum_{i=1}^{r} \frac{(N_{i1} - N_{i2})^2}{N_{i1} + N_{i2}}$$

[25 marks]

5. (a) *Dalam satu kajian terhadap pengambilan alkohol yang dijalankan oleh sebuah agensi kerajaan , satu sampel rawak yang terdiri dari 3603 individu yang berumur 20 tahun ke atas telah dipilih untuk ditemubual. Empat pemboleh ubah telah dipertimbangkan dalam kajian itu:*
Pengambilan alkohol – 1 unit dalam satu hari, 2 unit dalam satu hari
Kumpulan sosial -- Kelas atas, Kelas Pertengahan, Kelas Rendah, Kelas Bawahan.
Status Perkahwinan -- Berkahwin, Bujang
Umur -- 20 – 39 tahun , 40 – 69 tahun
(1 unit lebih kurang 1 botol 33 cl beer dengan kekuatan 4% isipadunya atau yang setara)
Hasilnya dalam jadual kekontingenan empat-hala adalah seperti di bawah.

Pengambilan Alkohol	Kumpulan Sosial	Status Perkahwinan	Umur 20 –39	Umur 40 -- 69
1 unit satu hari	Atas	Berkahwin	43	38
		Bujang	11	6
	Pertengahan	Berkahwin	112	158
		Bujang	35	33
	Rendah	Berkahwin	195	169
		Bujang	76	44
	Bawahan	Berkahwin	135	216
		Bujang	72	61
2 unit satu hari	Atas	Berkahwin	149	134
		Bujang	37	28
	Pertengahan	Berkahwin	200	225
		Bujang	80	45
	Rendah	Berkahwin	296	238
		Bujang	195	50
	Bawahan	Berkahwin	131	212
		Bujang	130	49

(i) *Perihalkan proses mendapat model akhir menggunakan penyingkiran kebelakang dan nyatakan model akhir berdasarkan kepada output komputer dalam **Lampiran B 5 (a).***

(ii) *Tafsirkan anggaran parameter interaksi tertinggi dalam model akhir.*

[50 markah]

(b) *Untuk jadual 2 x 2, suatu ukuran sekutuan yang serupa dengan suatu pekali korelasi ialah Yule's Q yang ditakrifkan oleh*

$$Q = \frac{p_{11}\, p_{22} - p_{12} p_{21}}{p_{11} p_{22} + p_{12} p_{21}} \quad .$$

Carilah Q dalam sebutan nisbah ods. Tunjukkan bahawa Q mesti berada antara –1 dan 1 .

[25 markah]

(c) *Pertimbangkan suatu jadual r x 2 yang dibentuk bila mana sampel-sampel tak bersandar dengan masing-masing saiz* $n_{.1}$ *dan* $n_{.2}$ *diambil dari populasi yang mempunyai sama r kelas pengkelasan. Biar* N_{ij} *menandai pemboleh ubah rawak kekerapan individu berada dalam sel (i, j) , i = 1 , 2 , j = 1 , 2 . Tunjukkan bahawa untuk kes saiz sampel sama dengan* $n_{.1} = n_{.2}$ *, statistik* X^2 *mengambil bentuk*

$$X^2 = \sum_{i=1}^{r} \frac{(N_{i1} - N_{i2})^2}{N_{i1} + N_{i2}}$$

[25 markah]

Appendix A 1(b) *lampiran A 1(b)*

GENERAL LOGLINEAR ANALYSIS

Data Information

```
  8 cases are accepted.
  0 cases are rejected because of missing data.
200 weighted cases will be used in the analysis.
  8 cells are defined.
  0 structural zeros are imposed by design.
  0 sampling zeros are encountered.
```

Variable Information

Factor	Levels	Value
SEX	2	male
		0 female
		1 male
T_MENT	2	type of treatment
		0 control
		1 newt_ment
SUB_COND	2	subject condition
		0 not improved
		1 improved

Model and Design Information

Model: Multinomial
Design: Constant + SEX + SUB_COND + T_MENT + SEX*T_MENT + T_MENT*SUB_COND

Correspondence Between Parameters and Terms of the Design

Parameter	Aliased	Term
1		Constant
2		[SEX = 0]
3	x	[SEX = 1]
4		[SUB_COND = 0]
5	x	[SUB_COND = 1]
6		[T_MENT = 0]
7	x	[T_MENT = 1]
8		[SEX = 0]*[T_MENT = 0]
9	x	[SEX = 0]*[T_MENT = 1]
10	x	[SEX = 1]*[T_MENT = 0]
11	x	[SEX = 1]*[T_MENT = 1]
12		[T_MENT = 0]*[SUB_COND = 0]
13	x	[T_MENT = 0]*[SUB_COND = 1]
14	x	[T_MENT = 1]*[SUB_COND = 0]
15	x	[T_MENT = 1]*[SUB_COND = 1]

Note: 'x' indicates an aliased (or a redundant) parameter.
These parameters are set to zero.

Convergence Information

Maximum number of iterations: 20
Relative difference tolerance: .001
Final relative difference: 9.70104E-06

Maximum likelihood estimation converged at iteration 3.

Table Information

Factor	Value	Observed Count	%	Expected Count	%
SEX	female				
T_MENT	control				
SUB_COND	not improved	27.00	(13.50)	24.00	(12.00)
SUB_COND	improved	23.00	(11.50)	26.00	(13.00)
T_MENT	newt_ment				
SUB_COND	not improved	22.00	(11.00)	19.00	(9.50)
SUB_COND	improved	28.00	(14.00)	31.00	(15.50)
SEX	male				
T_MENT	control				
SUB_COND	not improved	21.00	(10.50)	24.00	(12.00)
SUB_COND	improved	29.00	(14.50)	26.00	(13.00)
T_MENT	newt_ment				
SUB_COND	not improved	16.00	(8.00)	19.00	(9.50)
SUB_COND	improved	34.00	(17.00)	31.00	(15.50)

...16/-

GENERAL LOGLINEAR ANALYSIS

Table Information

Factor	Value	Resid.	Adj. Resid.	Dev. Resid.
SEX	female			
T_MENT	control			
SUB_COND	not improved	3.00	1.20	2.52
SUB_COND	improved	-3.00	-1.20	-2.37
T_MENT	newt_ment			
SUB_COND	not improved	3.00	1.24	2.54
SUB_COND	improved	-3.00	-1.24	-2.39
SEX	male			
T_MENT	control			
SUB_COND	not improved	-3.00	-1.20	-2.37
SUB_COND	improved	3.00	1.20	2.52
T_MENT	newt_ment			
SUB_COND	not improved	-3.00	-1.24	-2.35
SUB_COND	improved	3.00	1.24	2.51

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	2.9787	2	.2255
Pearson	2.9703	2	.2265

GENERAL LOGLINEAR ANALYSIS

Parameter Estimates

Constant	Estimate
1	3.4340

Note: Constant is not a parameter under multinomial assumption.
Therefore, standard errors are not calculated.

Parameter	Estimate	SE	Z-value	Asymptotic 95% CI Lower	Upper
2	-1.061E-12	.2000	-5.306E-12	-.39	.39
3	.0000	.	.	.	.
4	-.4895	.2060	-2.38	-.89	-.09
5	.0000	.	.	.	.
6	-.1759	.2353	-.75	-.64	.29
7	.0000	.	.	.	.
8	9.038E-13	.2828	3.195E-12	-.55	.55
9	.0000	.	.	.	.
10	.0000	.	.	.	.
11	.0000	.	.	.	.
12	.4095	.2872	1.43	-.15	.97
13	.0000	.	.	.	.
14	.0000	.	.	.	.
15	.0000	.	.	.	.

Covariance Matrix of Parameter Estimates

Parameter	2	4	6	8	12
2	.0400				
4	2.478E-19	.0424			
6	.0200	.0161	.0554		
8	-.0400	-2.33E-18	-.0400	.0800	
12	4.389E-19	-.0424	-.0354	.0000	.0825

Aliased parameters and constant are not shown.

Appendix B 5(a) ***Lampiran B 5 (a)***

```
* * * * * * * *  H I E R A R C H I C A L   L O G   L I N E A R  * * * * * * * *

DATA   Information

         32 unweighted cases accepted.
          0 cases rejected because of out-of-range factor values.
          0 cases rejected because of missing data.
       3603 weighted cases will be used in the analysis.


FACTOR Information

   Factor  Level
   A_COHOL     2
   SOC_GRP     4
   M_STATUS    2
   AGE         2
```

```
* * * * * * * *  H I E R A R C H I C A L   L O G   L I N E A R  * * * * * * * *

Backward Elimination (p = .050) for DESIGN 1 with generating class

  A_COHOL*SOC_GRP*M_STATUS*AGE

 Likelihood ratio chi square =      .00000    DF = 0  P = 1.000

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -


If Deleted Simple Effect is                  DF   L.R. Chisq Change    Prob  Iter

 A_COHOL*SOC_GRP*M_STATUS*AGE                 3               3.853   .2777     4

Step 1

  The best model has generating class

      A_COHOL*SOC_GRP*M_STATUS
      A_COHOL*SOC_GRP*AGE
      A_COHOL*M_STATUS*AGE
      SOC_GRP*M_STATUS*AGE

  Likelihood ratio chi square =     3.85333    DF = 3  P =  .278

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -


If Deleted Simple Effect is                  DF   L.R. Chisq Change    Prob  Iter

 A_COHOL*SOC_GRP*M_STATUS                     3                .478   .9236     4
 A_COHOL*SOC_GRP*AGE                          3               1.670   .6436     4
 A_COHOL*M_STATUS*AGE                         1              12.174   .0005     4
 SOC_GRP*M_STATUS*AGE                         3              13.526   .0036     4

Step 2

  The best model has generating class

      A_COHOL*SOC_GRP*AGE
      A_COHOL*M_STATUS*AGE
      SOC_GRP*M_STATUS*AGE

  Likelihood ratio chi square =     4.33170    DF = 6  P =  .632

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

* * * * * * * * H I E R A R C H I C A L   L O G   L I N E A R * * * * * * * *

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
A_COHOL*SOC_GRP*AGE	3	1.595	.6605	4
A_COHOL*M_STATUS*AGE	1	12.552	.0004	4
SOC_GRP*M_STATUS*AGE	3	14.096	.0028	4

Step 3

The best model has generating class

A_COHOL*M_STATUS*AGE
SOC_GRP*M_STATUS*AGE
A_COHOL*SOC_GRP

Likelihood ratio chi square =   5.92666   DF = 9   P = .747

- -

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob	Iter
A_COHOL*M_STATUS*AGE	1	12.880	.0003	4
SOC_GRP*M_STATUS*AGE	3	14.212	.0026	4
A_COHOL*SOC_GRP	3	98.291	.0000	2

Step 4

The best model has generating class

A_COHOL*M_STATUS*AGE
SOC_GRP*M_STATUS*AGE
A_COHOL*SOC_GRP

Likelihood ratio chi square =   5.92666   DF = 9   P = .747

- -

...21/-

```
* * * * * * * *  H I E R A R C H I C A L   L O G   L I N E A R  * * * * * * * *

The final model has generating class

    A_COHOL*M_STATUS*AGE
    SOC_GRP*M_STATUS*AGE
    A_COHOL*SOC_GRP

The Iterative Proportional Fit algorithm converged at iteration 0.
The maximum difference between observed and fitted marginal totals is     .068
and the convergence criterion is     .296


 Goodness-of-fit test statistics

   Likelihood ratio chi square =     5.92666    DF = 9  P =  .747
            Pearson chi square =     6.10507    DF = 9  P =  .729

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

- ooo O ooo -