
UNIVERSITI SAINS MALAYSIA

First Semester Examination
Academic Session 2007/2008

October/November 2007

MST 567 – Categorical Data Analysis
[Analisis Data Berkategori]

Duration : 3 hours
[Masa : 3 jam]

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

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

Instructions: Answer **all five** [5] questions.

[Arahan: Jawab **semua lima** [5] soalan.]

...2/-

1. (a) For testing the null hypothesis $H_0 : \pi_j = \pi_{j0}$ against $H_1 : \pi_j \neq \pi_{j0}$ for a specified multinomial distribution show the test statistic. Also show that the likelihood ratio test can be expressed as

$$-2 \log \Lambda = 2 \sum_{j=1}^c n_j \log \left(\frac{n_j}{n \pi_{j0}} \right).$$

[40 marks]

- (b) In a trial comparing a new drug to a standard, π denotes the probability that the new one is judged better. It is desired to estimate π and test $H_0 : \pi = 0.5$ against $H_1 : \pi \neq 0.5$. In 20 independent observations, the new drug is better 12 times.

- (i) Conduct a Wald test and construct a 95% Wald confidence interval for π .
- (ii) Conduct a score test and construct a 95% score confidence interval for π .
- (iii) Conduct a likelihood ratio based test and comment.
- (iv) Compare (i) and (ii) with adjusted intervals based on $n^* = n + z_{\alpha/2}^2$.

[60 marks]

2. (a) Find the joint probability mass functions for potential outcomes $\{n_{ij}\}$ under Poisson and multinomial sampling models for cell counts $\{Y_{ij}\}$. Find the connection between the Poisson and the Multinomial distribution.

[50 marks]

- (b) Consider the following table representing type of cerebral tumour and site of tumour:

| Site | Type of tumour | | |
|------|----------------|----|----|
| | A | B | C |
| I | 23 | 9 | 6 |
| II | 21 | 4 | 3 |
| III | 34 | 24 | 17 |

- (i) Test for independence between type of tumour and site.
- (ii) Comment on the results.
- (iii) Show the partitioning of chi square, and comment on the results.

[50 marks]

...3/-

1. (a) Tunjukkan statistik ujian bagi menguji hipotesis nol $H_0 : \pi_j = \pi_{j0}$ melawan $H_0 : \pi_j \neq \pi_{j0}$ bagi suatu taburan multinomial tertentu. Juga tunjukkan bahawa ujian nisbah kebolehdian boleh diungkap seperti

$$-2 \log \Lambda = 2 \sum_{j=1}^c n_j \log \left(\frac{n_j}{n \pi_{j0}} \right).$$

[40 markah]

- (b) Dalam suatu ujian membanding ubat baru terhadap ubat yang biasa diguna, π menandakan kebarangkalian bahawa ubat baru itu dinilai lebih baik. Di sini kita ingin menganggar π dan menguji $H_0 : \pi = 0.5$ against $H_1 : \pi \neq 0.5$. Di dalam 20 cerapan tak bersandar, terdapat 12 kali ubat baru itu adalah lebih baik.

- (i) Jalankan ujian Wald dan bina satu selang keyakinan Wald 95% bagi π .
- (ii) Jalankan satu ujian skor dan bina satu selang keyakinan skor 95% bagi π .
- (iii) Jalankan satu ujian berdasarkan nisbah kebolehdian dan berikan komen.
- (iv) Banding bahagian (i) dan bahagian (ii) dengan selang terlaras berdasarkan $n^* = n + z_{\alpha/2}^2$.

[60 markah]

2. (a) Dapatkan fungsi jisim kebarangkalian tercantum bagi kesudahan $\{n_{ij}\}$ di bawah model pensampelan Poisson dan model pensampelan multinomial bagi bilangan sel $\{Y_{ij}\}$. Dapatkan perkaitan diantara taburan dan taburan Multinomial.

[50 markah]

- (b) Pertimbangkan jadual berikut yang mewakili jenis ketumbuhan selebral dan tempat ketumbuhan tersebut:

| Tempat | Jenis Ketumbuhan | | |
|--------|------------------|----|----|
| | A | B | C |
| I | 23 | 9 | 6 |
| II | 21 | 4 | 3 |
| III | 34 | 24 | 17 |

- (i) Jalankan ujian bagi ketaksandaran diantara jenis ketumbuhan dan tempat.
- (ii) Berikan komen terhadap keputusan.
- (iii) Tunjukkan pembahagian khi kuasadua dan komen keputusannya.

[50 markah]

...4/-

3. (a) Define the homogeneous association in a $2 \times 2 \times K$ table. How can you extend this for an $I \times J$ table? [30 marks]
- (b) Consider the following contingency table representing admissions decisions by gender of applicant in a university. Three variables are: A= whether admitted, G=gender, and D=department. Find the sample AG conditional odds ratios and the marginal odds ratio. Interpret, and explain why they give such different indications of the AG association.

| Department | Whether Admitted | | | |
|------------|------------------|-----|--------|-----|
| | Male | | Female | |
| | Yes | No | Yes | No |
| A | 512 | 313 | 89 | 19 |
| B | 353 | 207 | 17 | 8 |
| C | 120 | 205 | 202 | 391 |
| D | 138 | 279 | 131 | 244 |
| E | 53 | 138 | 94 | 299 |
| F | 22 | 351 | 24 | 317 |

[70 marks]

- 4 (a) Show that the binomial distribution is a special case of the probability density/mass function

$$f(y, \theta, \phi) = e^{\{(y\theta - b(\theta)) / a(\phi) + c(y, \phi)\}}$$

Hence, show the logit of π , where $\theta = \text{logit}(\pi)$. Then find the binomial generalized linear model for 2×2 contingency table and show that $\text{logit}[\pi(0)] = \beta_0$, $\text{logit}[\pi(1)] = \beta_0 + \beta_1$.

[40 marks]

- (b) The following table represents data on birthweight and maternal age:

| Maternal Age | Birthweight | |
|---------------------|---------------------|------------------|
| | ≤ 2500 gms (0) | > 2500 gms (1) |
| ≤ 20 years (0) | 10 | 40 |
| > 20 years (1) | 15 | 135 |

...5/-

3. (a) Takrifkan kaitan homogen dalam satu jadual $2 \times 2 \times K$. Bagaimana anda boleh kembangkan takrifan ini bagi satu jadual $I \times J$? [30 markah]
- (b) Pertimbangkan jadual kontingensi berikut yang mewakili keputusan kemasukan ke sebuah universiti mengikut jantina pemohon. Tiga pembolehubah adalah: A = sama ada diterima, G = jantina, dan D = jabatan. Dapatkan nisbah odds bersyarat AG sampel dan nisbah odds sut AG sampel. Tafsirkan dan terangkan kenapa kedua-dua nisbah odds ini memberi petanda kaitan AG yang berbeza.

| Jabatan | Menerima Masuk | | | |
|---------|----------------|-------|-----------|-------|
| | Lelaki | | Perempuan | |
| | Ya | Tidak | Ya | Tidak |
| A | 512 | 313 | 89 | 19 |
| B | 353 | 207 | 17 | 8 |
| C | 120 | 205 | 202 | 391 |
| D | 138 | 279 | 131 | 244 |
| E | 53 | 138 | 94 | 299 |
| F | 22 | 351 | 24 | 317 |

[70 markah]

4. (a) Tunjukkan bahawa taburan binomial adalah satu kes khas bagi fungsi ketumpatan/ jisim kebarangkalian

$$f(y, \theta, \phi) = e^{\{(y\theta - b(\theta)) / a(\phi) + c(y, \phi)\}}$$

Oleh itu, tunjukkan logit bagi π , dengan $\theta = \text{logit}(\pi)$. Seterusnya dapatkan model linear binomial teritlak bagi jadual kontingensi 2×2 dan tunjukkan bahawa $\text{logit}[\pi(0)] = \beta_0$, $\text{logit}[\pi(1)] = \beta_0 + \beta_1$.

[40 markah]

- (b) Jadual berikut mewakili data berat kelahiran dan umur ibu:

| Umur Ibu | Berat Kelahiran | |
|---------------------|---------------------|------------------|
| | ≤ 2500 gms (0) | > 2500 gms (1) |
| ≤ 20 tahun (0) | 10 | 40 |
| > 20 tahun (1) | 15 | 135 |

...6/-

- (i) Obtain the estimates of the odds ratio and the relative risk. Obtain their standard errors. What is the estimate of the risk difference?
- (ii) Use different test procedures for testing the independence between exposure categories and birthweight. Comment on the results.
- (iii) Construct 95% confidence intervals for the log odds ratio and log relative risk as well as for the odds ratio and relative risk.
- (iv) Fit a logistic regression model and find the odds ratio and compare with (i).

[60 marks]

5. (a) Let the dependent variable can take three nominal scale values, 0,1 and 2. Then find the conditional probabilities of $Y=0$, $Y=1$ and $Y=2$ for given X and show the likelihood and log-likelihood functions.

[30 marks]

- (b) If the entries in contingency tables are considered as counts, then show that the generalized linear model for Poisson probability mass function can be expressed by the following log-linear model:

$$Q(\mu_i) = \log \mu_i = \sum_{j=1}^p \beta_j x_{ij}$$

Also show that Poisson GLM of independence in $I \times J$ contingency table can be expressed as $\log \mu_{ij} = \lambda + \alpha_i^* + \beta_j^*$.

[30 marks]

- (c) Consider the following table describing health opinion by gender and information opinion. Fit a log-linear model using the Poisson GLM and comment on the results:

| Gender | Information Opinion | Health Opinion | |
|--------|---------------------|----------------|--------|
| | | Support | Oppose |
| Male | Support | 76 | 160 |
| | Oppose | 6 | 25 |
| Female | Support | 114 | 181 |
| | Oppose | 11 | 48 |

[40 marks]

...7/-

- (i) Dapatkan anggaran nisbah odds dan risiko relatif. Dapatkan ralat piawai masing-masing. Apakah anggaran risiko perbezaan?
- (ii) Gunakan prosedur ujian lain untuk menguji ketaksandaran antara kategori pendedahan dan berat kelahiran. Berikan komen terhadap keputusannya.
- (iii) Bina selang-selang keyakinan 95% bagi log nisbah odds dan log risiko relatif serta nisbah odds dan risiko relative.
- (iv) Suaikan satu model regresi logistik dan dapatkan nisbah odds dan bandingkan dengan bahagian (i).

[60 markah]

5. (a) Andaikan pembolehubah bersandar boleh mengambil tiga nilai skala nominal, 0, 1 dan 2. Seterusnya dapatkan kebarangkalian bersyarat $Y=0$, $Y=1$ dan $Y=2$ diberikan X dan tunjukkan fungsi kebolehdajian dan fungsi log-kebolehdajian.

[30 markah]

- (b) Jika kemasukan dalam jadual kontingensi dipertimbangkan sebagai bilangan, tunjukkan bahawa model linear teritlak bagi fungsi jisim kebarangkalian Poisson boleh diungkap oleh model log-linear berikut:

$$Q(\mu_i) = \log \mu_i = \sum_{j=1}^p \beta_j x_{ij}$$

Tunjukkan juga bahawa Poisson GLM ketaksandaran dalam jadual kontingensi $I \times J$ boleh diungkap sebagai $\log \mu_{ij} = \lambda + \alpha_i^* + \beta_j^*$.

[30 markah]

- (c) Pertimbangkan jadual berikut yang menerangkan tentang pendapat kesihatan mengikut jantina dan pendapat maklumat. Suaikan satu model log-linear dengan menggunakan Poisson GLM dan berikan komen terhadap keputusannya:

| Jantina | Pendapat Informasi | Pendapat Kesihatan | |
|-----------|--------------------|--------------------|-----------|
| | | Menyokong | Menentang |
| Lelaki | Menyokong | 76 | 160 |
| | Menentang | 6 | 25 |
| Perempuan | Menyokong | 114 | 181 |
| | Menentang | 11 | 48 |

[40 markah]