UNIVERSITY OF CAMBRIDGE

## M. Phil. in STATISTICAL SCIENCE

Monday 7 June to Thursday 10 June 2004

# APPLIED STATISTICS

*Attempt* **THREE** *questions.*

*There are four questions in total.*

*The questions carry equal weight.*

*This is an 'Open Book' examination, involving the use of the Statistical Laboratory's network of workstations. Candidates will receive this paper at 9.00am on Monday 7 June, and must hand in their scripts to the Chairman of Examiners by 1.00pm on Thursday 10 June.*

*The data sets will be emailed to candidates on Monday 7 June.*

*(The Statistical Laboratory Computer Officer and an Examiner will normally be available for consultation if required between 9.00am and 4.30pm on these four days.)*

*Each candidate should submit his/her script with a signed statement that the work has been carried out without any collaboration with others.*

*The scripts may be handwritten. Candidates are requested to submit at most 25 pages in total. They are advised that the total work set should take between 4 and 6 hours.*

**1** The Independent, on February 20, 2003, gave the following data on Police Performance Monitoring, for the 43 police forces of England and Wales. The 6 columns of the table are, respectively

| | | |
|---|---|---|
| Burg | = | number of Burglaries for every 1000 households |
| Vehc | = | Vehicle crimes for every 1000 residents |
| Robb | = | Robberies for every 1000 residents |
| Offdet | = | percentage of offences detected |
| HiDis | = | percentage of residents perceiving disorder as high |
| GoodJob | = | percentage of residents thinking police do a good job |

(i) Summarise the data with appropriate graphs, tables and a paragraph of text.

(ii) How does HiDis depend on the first 4 variables Burg, . . . , Offdet? How does GoodJob depend on these first 4 variables? Illustrate the use of the stepAIC( ) function in your solution.

(iii) What is the sample correlation matrix for the 6 variables? What is the partial correlation of HiDis and GoodJob, conditional on the remaining 4 variables?

| | Burg | Vehc | Robb | OffDet | HiDis | GoodJob |
|---|---|---|---|---|---|---|
| AvonandSomerset | 25.1 | 27.0 | 3.2 | 14 | 19 | 46 |
| Bedfordshire | 15.7 | 22.5 | 1.7 | 21 | 26 | 47 |
| Cambridgeshire | 15.4 | 17.5 | 0.9 | 16 | 16 | 43 |
| Cheshire | 14.5 | 13.7 | 0.5 | 24 | 17 | 46 |
| CityofLondon | 13.0 | 140.0 | 7.2 | 32 | 32 | 49 |
| Cleveland | 35.8 | 25.6 | 2.3 | 20 | 18 | 38 |
| Cumbria | 10.1 | 9.1 | 0.3 | 27 | 14 | 40 |
| Derbyshire | 16.4 | 16.7 | 1.1 | 22 | 18 | 50 |
| Devon&Cornwall | 10.3 | 11.1 | 0.3 | 25 | 13 | 51 |
| Dorset | 11.0 | 14.0 | 0.5 | 23 | 14 | 51 |
| Durham | 15.7 | 12.6 | 0.5 | 29 | 28 | 48 |
| Dyfed-Powys | 3.6 | 4.2 | 0.1 | 51 | 12 | 53 |
| Essex | 8.2 | 12.8 | 0.6 | 20 | 16 | 47 |
| Gloucestershire | 14.2 | 14.4 | 0.9 | 26 | 14 | 46 |
| Gr‘Manchester | 36.2 | 28.9 | 4.3 | 16 | 27 | 44 |
| Gwent | 11.3 | 12.6 | 0.4 | 46 | 15 | 49 |
| Hampshire | 9.6 | 12.2 | 0.5 | 25 | 17 | 51 |
| Hertfordshire | 11.4 | 13.7 | 0.6 | 21 | 11 | 49 |
| Humberside | 29.8 | 24.2 | 1.3 | 17 | 18 | 43 |
| Kent | 11.4 | 12.9 | 0.6 | 24 | 20 | 45 |
| Lancashire | 20.4 | 14.7 | 1.1 | 25 | 17 | 46 |
| Leicestershire | 17.3 | 17.4 | 1.2 | 23 | 10 | 51 |
| Lincolnshire | 14.3 | 10.7 | 0.4 | 21 | 12 | 45 |
| Merseyside | 24.8 | 21.4 | 2.2 | 21 | 27 | 47 |
| Met.Police | 23.2 | 23.6 | 7.3 | 12 | 32 | 49 |
| Norfolk | 10.6 | 12.2 | 0.5 | 23 | 14 | 45 |
| Northamp’s | 15.0 | 18.1 | 1.4 | 24 | 25 | 45 |
| Northumbria | 18.5 | 14.4 | 1.0 | 29 | 24 | 56 |
| NorthWales | 8.9 | 11.7 | 0.3 | 24 | 22 | 50 |
| NorthYorkshire | 15.1 | 10.8 | 0.4 | 22 | 15 | 47 |
| Notts | 33.1 | 27.7 | 2.6 | 17 | 18 | 42 |
| SouthWales | 13.6 | 20.8 | 0.5 | 28 | 22 | 47 |
| SouthYorkshire | 29.5 | 22.1 | 1.5 | 23 | 27 | 45 |
| Staffordshire | 18.7 | 16.9 | 0.9 | 19 | 23 | 46 |
| Suffolk | 8.6 | 10.2 | 0.4 | 24 | 14 | 50 |
| Surrey | 8.2 | 8.7 | 0.5 | 19 | 12 | 53 |
| Sussex | 11.4 | 14.0 | 0.8 | 79 | 23 | 46 |
| ThamesValley | 15.5 | 19.6 | 1.4 | 20 | 17 | 45 |
| Warwickshire | 14.4 | 15.7 | 0.7 | 20 | 14 | 52 |
| WestMercia | 12.9 | 11.4 | 0.6 | 23 | 16 | 50 |
| WestMidlands | 29.5 | 24.3 | 5.1 | 24 | 18 | 46 |
| WestYorkshire | 39.1 | 30.8 | 2.7 | 18 | 20 | 46 |
| Wiltshire | 9.5 | 8.6 | 0.5 | 26 | 14 | 50 |

**2** In an experiment to assess the toxicity of pollutants in aquatic systems, females of species C. dubia were observed following exposure to a particular pollutant. Ten individuals were randomly allocated to each of five concentrations ($\mu g/l$) of the pollutant, and were observed over three subsequent breeding seasons. The number of young were recorded.

The data are given in the table below. Assess the effect the pollutant has on the numbers of young, both season to season and in total.

| Concentration | Season 1 | Season 2 | Season 3 | Total number | Concentration | Season 1 | Season 2 | Season 3 | Total number |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 3 | 14 | 10 | 27 | 160 | 6 | 13 | 12 | 31 |
| 0 | 5 | 12 | 15 | 32 | 160 | 6 | 12 | 12 | 30 |
| 0 | 6 | 11 | 17 | 34 | 160 | 5 | 10 | 11 | 26 |
| 0 | 6 | 12 | 15 | 33 | 160 | 6 | 13 | 10 | 29 |
| 0 | 6 | 15 | 15 | 36 | 160 | 6 | 12 | 11 | 29 |
| 0 | 5 | 14 | 15 | 34 | 235 | 4 | 13 | 6 | 23 |
| 0 | 6 | 12 | 15 | 33 | 235 | 6 | 10 | 5 | 21 |
| 0 | 5 | 13 | 12 | 30 | 235 | 2 | 5 | 0 | 7 |
| 0 | 3 | 10 | 11 | 24 | 235 | 6 | 0 | 6 | 12 |
| 0 | 6 | 11 | 14 | 31 | 235 | 6 | 13 | 8 | 27 |
| 80 | 6 | 11 | 16 | 33 | 235 | 6 | 0 | 10 | 16 |
| 80 | 5 | 12 | 16 | 33 | 235 | 7 | 0 | 6 | 13 |
| 80 | 6 | 11 | 18 | 35 | 235 | 4 | 2 | 9 | 15 |
| 80 | 5 | 12 | 16 | 33 | 235 | 6 | 8 | 7 | 21 |
| 80 | 8 | 13 | 15 | 36 | 235 | 7 | 0 | 10 | 17 |
| 80 | 3 | 9 | 14 | 26 | 310 | 6 | 0 | 0 | 6 |
| 80 | 5 | 9 | 13 | 27 | 310 | 6 | 0 | 0 | 6 |
| 80 | 7 | 12 | 12 | 31 | 310 | 7 | 0 | 0 | 7 |
| 80 | 5 | 13 | 14 | 32 | 310 | 0 | 0 | 0 | 0 |
| 80 | 3 | 12 | 14 | 29 | 310 | 5 | 10 | 0 | 15 |
| 160 | 6 | 12 | 11 | 29 | 310 | 5 | 0 | 0 | 5 |
| 160 | 6 | 12 | 11 | 29 | 310 | 6 | 0 | 0 | 6 |
| 160 | 2 | 8 | 13 | 23 | 310 | 4 | 0 | 0 | 4 |
| 160 | 6 | 10 | 11 | 27 | 310 | 6 | 0 | 0 | 6 |
| 160 | 6 | 11 | 13 | 30 | 310 | 5 | 0 | 0 | 5 |

**3** You see in the Table below an extract from the Metropolitan Police Statistics for offences in the category

"Violence against the Person"

for each of the 33 Metropolitan boroughs, for September 2003 and October 2003.

The Metropolitan Police Service, Offences by Borough
www.met.police.uk/crimestatistics/stat

Violence against the person

.....................................................

September 2003

| | Murder | GBH | ABH | ComAss | OffW | Har | OViol | VAP.Tot |
|---|---|---|---|---|---|---|---|---|
| Westminster | 0 | 12 | 159 | 316 | 82 | 117 | 47 | 733 |
| Camden | 1 | 21 | 147 | 196 | 33 | 88 | 29 | 515 |
| Islington | 1 | 10 | 124 | 227 | 40 | 116 | 40 | 558 |
| Hackney | 0 | 27 | 132 | 232 | 38 | 89 | 53 | 571 |
| Tower_Ham | 1 | 12 | 70 | 298 | 37 | 136 | 34 | 588 |
| Greenwich | 2 | 10 | 122 | 283 | 19 | 115 | 34 | 585 |
| Lewisham | 3 | 16 | 85 | 246 | 24 | 69 | 54 | 497 |
| Southwark | 2 | 26 | 165 | 322 | 32 | 146 | 67 | 760 |
| Lambeth | 2 | 31 | 190 | 325 | 55 | 155 | 51 | 809 |
| Wandsworth | 0 | 19 | 90 | 236 | 23 | 90 | 17 | 475 |
| Hamm&Fulham | 0 | 8 | 91 | 141 | 20 | 78 | 20 | 358 |
| Kens&Chelsea | 0 | 6 | 88 | 115 | 10 | 49 | 8 | 276 |
| Walt_Forest | 0 | 15 | 90 | 259 | 25 | 115 | 23 | 527 |
| Redbridge | 0 | 8 | 37 | 221 | 15 | 76 | 32 | 389 |
| Havering | 0 | 5 | 39 | 207 | 11 | 44 | 14 | 320 |
| Bark&Dagenham | 0 | 10 | 50 | 194 | 14 | 44 | 14 | 326 |
| Newham | 1 | 13 | 101 | 383 | 34 | 128 | 42 | 702 |
| Bexley | 0 | 2 | 64 | 175 | 6 | 72 | 31 | 350 |
| Bromley | 0 | 7 | 67 | 217 | 12 | 82 | 30 | 415 |
| Croydon | 0 | 21 | 146 | 333 | 21 | 133 | 40 | 694 |
| Sutton | 0 | 5 | 24 | 162 | 4 | 49 | 4 | 248 |
| Merton | 0 | 6 | 76 | 126 | 17 | 73 | 14 | 312 |
| Kingston_u_T | 1 | 5 | 41 | 180 | 3 | 36 | 6 | 272 |
| Richmond_u_T | 0 | 1 | 32 | 112 | 9 | 41 | 11 | 206 |
| Hounslow | 0 | 11 | 139 | 264 | 20 | 124 | 59 | 617 |
| Hillingdon | 0 | 10 | 92 | 178 | 16 | 64 | 25 | 385 |
| Ealing | 0 | 16 | 142 | 317 | 13 | 94 | 48 | 630 |
| Brent | 2 | 16 | 77 | 372 | 18 | 86 | 52 | 623 |
| Harrow | 1 | 7 | 68 | 126 | 9 | 31 | 17 | 259 |
| Barnet | 0 | 9 | 93 | 216 | 20 | 85 | 33 | 456 |
| Haringey | 1 | 14 | 162 | 133 | 31 | 52 | 39 | 432 |
| Enfield | 0 | 15 | 91 | 180 | 18 | 46 | 35 | 385 |
| H/R_Airport | 0 | 0 | 5 | 14 | 5 | 5 | 1 | 30 |
| Total | 18 | 394 | 3099 | 7306 | 734 | 2728 | 1024 | 15303 |

October 2003

| | Murder | GBH | ABH | ComAss | OffW | Har | OViol | VAP.Tot |
|---|---|---|---|---|---|---|---|---|
| Westminster | 0 | 21 | 180 | 307 | 87 | 148 | 46 | 789 |
| Camden | 0 | 29 | 130 | 225 | 35 | 110 | 37 | 566 |
| Islington | 1 | 18 | 106 | 223 | 38 | 124 | 42 | 552 |
| Hackney | 5 | 22 | 148 | 277 | 44 | 80 | 53 | 629 |
| Tower_Ham | 1 | 22 | 60 | 331 | 29 | 153 | 38 | 634 |
| Greenwich | 1 | 6 | 106 | 262 | 12 | 111 | 34 | 532 |
| Lewisham | 0 | 11 | 93 | 273 | 30 | 98 | 54 | 559 |
| Southwark | 1 | 21 | 158 | 313 | 55 | 116 | 64 | 728 |
| Lambeth | 1 | 25 | 209 | 295 | 52 | 123 | 72 | 777 |
| Wandsworth | 0 | 11 | 91 | 180 | 27 | 95 | 26 | 430 |
| Hamm&Fulham | 1 | 10 | 73 | 139 | 21 | 80 | 20 | 344 |
| Kens&Chelsea | 0 | 4 | 73 | 94 | 15 | 42 | 11 | 239 |
| Walt_Forest | 2 | 15 | 80 | 205 | 26 | 103 | 18 | 449 |
| Redbridge | 1 | 2 | 63 | 179 | 14 | 67 | 24 | 350 |
| Havering | 0 | 9 | 48 | 184 | 12 | 39 | 14 | 306 |
| Bark&Dagenham | 0 | 7 | 91 | 225 | 32 | 78 | 20 | 453 |
| Newham | 1 | 17 | 109 | 386 | 25 | 109 | 40 | 687 |
| Bexley | 0 | 7 | 69 | 145 | 11 | 70 | 23 | 325 |
| Bromley | 0 | 6 | 70 | 201 | 17 | 119 | 20 | 433 |
| Croydon | 1 | 24 | 134 | 270 | 29 | 113 | 30 | 601 |
| Sutton | 0 | 3 | 28 | 147 | 25 | 49 | 3 | 255 |
| Merton | 0 | 1 | 63 | 149 | 17 | 66 | 19 | 315 |
| Kingston_u_T | 0 | 5 | 64 | 156 | 11 | 51 | 15 | 302 |
| Richmond_u_T | 0 | 8 | 31 | 95 | 6 | 32 | 11 | 183 |
| Hounslow | 1 | 7 | 125 | 272 | 12 | 147 | 32 | 596 |
| Hillingdon | 0 | 8 | 82 | 205 | 15 | 78 | 29 | 417 |
| Ealing | 0 | 17 | 121 | 273 | 19 | 91 | 30 | 551 |
| Brent | 1 | 16 | 77 | 409 | 48 | 100 | 53 | 704 |
| Harrow | 0 | 5 | 49 | 115 | 17 | 21 | 13 | 220 |
| Barnet | 0 | 10 | 94 | 191 | 11 | 104 | 30 | 440 |
| Haringey | 1 | 20 | 110 | 137 | 46 | 73 | 35 | 422 |
| Enfield | 2 | 11 | 112 | 181 | 18 | 49 | 32 | 405 |
| H/R_Airport | 0 | 1 | 1 | 8 | 9 | 3 | 0 | 22 |
| Total | 21 | 399 | 3048 | 7052 | 865 | 2842 | 988 | 15215 |

Key: Murder = Murder, GBH = Grievous Bodily Harm, ABH = Aggravated Bodily Harm, ComAss = Common Assault, Har = Harassment, OViol= Other Violence, VAP.Tot = Violence against the person, total.

(i) Summarise the data with appropriate graphs, and a paragraph of text. This summary should include informal comparisons of the crime statistics for September with those of October.

(ii) Let ySept be the "VAP.Tot" figures for September 2003, and let yOct be the corresponding figures for October 2003. Use appropriate non-parametric methods to see whether

(a) ySept are different from yOct;

(b) ySept are related to yOct.

(iii) Now let $(y_{1j})$ be ySept, and let $(y_{2j})$ be yOct.

Let $(CA_{1j})$ be the number of Common Assaults for September, and let $(CA_{2j})$ be the corresponding number for October. (Here $j = 1, \ldots, 33$, corresponding to the 33 boroughs)

$$\text{Let } \mathbb{E}(CA_{ij}/y_{ij}) = \pi_{ij}, \quad \text{for } i = 1, 2 \text{ and } j = 1, \ldots, 33.$$

(a) Test the hypothesis $H_0 : \pi_{1j} = \pi_1$ for all $j$.

(b) Fit the model $g(\pi_{ij}) = \mu + \alpha_i + \beta_j$, for $i = 1, 2$ and $j = 1, \ldots, 33$.

where $g(\ )$ is the usual logit link. Can you identify one particular Borough whose removal makes this model fit quite well?

How would you interpret the model to a layman?

**4** Consider a data-set from a multi-centre, placebo-controlled randomised trial on 1000 patients with liver cirrhosis and no previous history of bleeding. Patients were randomised to receive either propranolol or a placebo. Eligible patients included patients in whom cirrhosis was histologically confirmed and where endoscopy had shown oesophageal varices of either Grade 2 or 3. The aim of the study was to evaluate the effect of propranolol versus placebo on the risk of a *first* bleed and on survival (either from having a first bleed or from never having one). Additional information on gender and the base-line Child-Pugh classification score (which is an indication of a patient's prognosis, and is graded A, B or C corresponding to having a good, an intermediate or a bad prognosis respectively) was recorded.

A subset of data is shown below (with the time variable suitable rounded to two decimal places). The codes for the headers are also represented.

| subject | time | trt | sex | CPclass | grade | state |
|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 1 | 3 | 0 | 1 |
| 1 | 1 | 0 | 1 | 3 | 0 | 1 |
| 1 | 2 | 0 | 1 | 3 | 0 | 1 |
| 1 | 2.77 | 0 | 1 | 3 | 0 | 3 |
| 2 | 0 | 0 | 0 | 2 | 0 | 1 |
| 2 | 1 | 0 | 0 | 2 | 0 | 1 |
| 2 | 2 | 0 | 0 | 2 | 0 | 1 |
| 2 | 3 | 0 | 0 | 2 | 0 | 2 |
| 2 | 4 | 0 | 0 | 2 | 0 | 2 |
| 2 | 5 | 0 | 0 | 2 | 0 | 2 |
| . | | | | | | |
| . | | | | | | |
| . | | | | | | |
| 999 | 0 | 0 | 0 | 3 | 0 | 1 |
| 999 | 1 | 0 | 0 | 3 | 0 | 2 |
| 999 | 1.04 | 0 | 0 | 3 | 0 | 3 |
| 1000 | 0 | 0 | 1 | 3 | 1 | 1 |
| 1000 | 0.49 | 0 | 1 | 3 | 1 | 3 |

subject = Patient identification number

time = The time in the study (in years)

state = The state the patient is in at a particular time point (1 corresponds to the no bleeding state; 2 corresponds to the bleeding state and 3 to the death state.)

trt = The treatment received (0 corresponds to placebo; 1 to propranolol)

sex = The gender of the patient (0 = female; 1 = male)

CPclass = The Child-Pugh classification (1 = A; 2 = B; 3 = C)

grade = Grade of varices (0 = Grade 2; 1 = Grade 3)

a) Construct a descriptive table of the patients' characteristics by each treatment group. Should you perform formal statistical tests to determine whether there are any differences in the characteristics between the two treatment groups? Give a reason for your answer.

b) Assuming a progressive disease model, draw the appropriate multi-state (transition) diagram that corresponds to the study's aim. What are the transition intensities and sojourn times of the multi-state model (assuming no covariates) that follow your transition diagram? Interpret them. What is the probability that a patient who is observed with a first bleed at a particular visit will be *alive* a year on from that visit?

c) If you ignore the effects of the other covariates, what are your estimates of the effects of treatment (propranolol versus placebo) on the transitions? (The model that you have fitted must be described and the R-code presented.)

d) Now investigate the simultaneous effects of the covariates on the transitions by fitting an appropriate multi-state model to the data, which takes into account the following assumptions:

(i) Propranolol will have no effect on the "no bleed to death" transition intensity.

(ii) Gender has no influence on the transition intensity from no bleed to first bleed, but has a *common* (i.e. the same) effect on the transition intensities that leads to death.

(iii) There is a *common* effect of having a Child-Pugh classification of B (compared to A) on the transition intensities that lead to death. There is also a *common* effect (but, in general, different from the above) of having a Child-Pugh classification of C (relative to A) on the transition intensities that lead to death.

(iv) There is a *common* effect of the grade of the oesophageal varices on the transitions from no bleed to first bleed and first bleed to death. However, there is no effect of the grade of the oesophageal varices on the transition intensity from no bleed to death.

Interpret carefully the results obtained. The model that you have fitted must be described and the R-code presented.