UNIVERSITY OF ABERDEEN **SESSION 2003–04**

Degree Examination

MX3527 Survival Statistics with Actuarial and Medical Applications

Tuesday 25 May 2004 (9am to 11am)

Only calculators approved by the Department of Mathematical Sciences may be used in this examination. Calculator memories must be clear at the start of the examination.

Marks may be deducted for answers that do not show clearly how the solution is reached.

Answer THREE questions. All questions carry equal weight.

1. **(a)** Give a brief description of the theory of graduation of life tables by means of a mathematical formula. You may assume that the function being graduated is $q_x(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ denotes an s-vector of parameters and x denotes age, and that data of the form $\{d_x, E_x\}$ is available for r consecutive integer ages x, where d_x, E_x denote the number of deaths and the "initial" exposed to risk respectively at age x.

(b) The mortality experience of a large group of non-smokers is being graduated by reference to a standard mortality table by means of the formula

$$q_x(\boldsymbol{\alpha}) = \alpha_1 + \alpha_2 q_x^A$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^\top$ is a vector of unknown parameters and q_x^A is according to the specified standard table. The ages involved are $x = 17, 18, \ldots, 88$, and you have obtained data of the form $\{d_x, E_x\}$.

(i) Give formulae for the likelihood function and the log-likelihood function for $\boldsymbol{\alpha}$.

(ii) Give two equations which must be solved (numerically) to find the maximum likelihood estimate $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \hat{\alpha}_2)^\top$ of $\boldsymbol{\alpha}$.

(iii) Define $\mathring{q}_x = q_x(\hat{\boldsymbol{\alpha}})$ for $x = 17, 18, \ldots, 88$, and let $\{\mathring{q}_x\}$ be referred to as the graduated rates of mortality. Define the standardised deviations, $\{z_x\}$.

(iv) You are given that

$$\sum_{x=17}^{88} (z_x)^2 = 139.07\,.$$

Test the hypothesis that graduation is successful, using a 0.1% significance level.

(c) In another graduation, the pattern of signs of $\{z_x\}$ as age increases from left to right was as follows:

$$- - - - - + + + + + + + + + + - - - - - - + +$$

Using a 1% significance level, test the suitability of the graduation in respect of the number of runs.

2. **(a)** Suppose that you are conducting a meta-analysis of the results of several clinical trials of treatment X against a control treatment. There are ℓ trials (labelled from $i = 1$ to $i = \ell$) and for trial i you find an unbiased estimator $\hat{\beta}_i$ of a certain parameter β (which is taken to be the same in all ℓ trials), together with the (approximate) variance of $\hat{\beta}_i$, which is denoted by v_i.

(i) Derive from first principles an expression for the estimator, β^*, of β which has the smallest variance among the family of unbiased estimators of β which are linear combinations of $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_\ell$.

(ii) Peto's method of estimating the log-odds ratio, β, leads to the unbiased estimator

$$P_i = \frac{O_i - E_i}{V_i}$$

in trial i, where O_i and E_i are the observed and expected deaths among "treatment" patients in trial i and V_i is a variance term. The variance of P_i is approximately equal to $(V_i)^{-1}$. Derive a simple formula for the minimum variance estimator β^* (defined in (i) above) in terms of O_i, E_i, V_i $(i = 1, 2, \ldots, \ell)$, and give an approximate formula for its variance.

(iii) Using the formulae of (ii), you have found the estimate $\beta^* = -0.2711$ of the log-odds ratio β, with approximate variance 0.003212. Find an approximate 95% confidence interval for the "true" odds ratio.

(b) A study of impaired assured lives in the period 1987 to 1998 gave the following facts for males:

impairment	actual deaths	actual/expected deaths
cerebrovascular disorders	79	1.72

The expected deaths were calculated using a standard table for male assured lives.

Find an approximate 95% confidence interval for the Standardised Mortality Ratio of male assured lives with this impairment relative to the standard table.

3. **(a)** Suppose that lives belong to exactly one of 3 types (A, B and C), and let p_1, p_2, p_3 denote the chances that a randomly-chosen life belongs to type A, B and C respectively. A random sample of n lives is chosen, and the numbers of people of types A, B and C are n_1, n_2 and n_3 respectively.

(i) You wish to test the hypothesis $H_0 : p_1 = p_2$. Assuming that n is large, derive formulae for McNemar's test of this hypothesis.

(ii) In a Finnish case-control study, each woman whose baby had a birth defect (a "case") was matched with another woman whose baby did not have a birth defect (a "control"), and the coffee consumption of the women during pregnancy was recorded. Let the variable X be defined as follows:

$$X = \begin{cases} 1 & \text{if "case" drank more coffee than "control"} \\ 2 & \text{if "control" drank more coffee than "case"} \\ 3 & \text{if consumption was the same for both women.} \end{cases}$$

The results were as follows:

$$\begin{array}{ll} X = 1 & \text{for 237 pairs of women} \\ X = 2 & \text{for 246 pairs of women} \\ X = 3 & \text{for 223 pairs of women.} \end{array}$$

Use McNemar's test to test the hypothesis that there is no association between coffee consumption and the incidence of birth defects.

(b) Let n lives be rated by two raters, A and B. Consider the opinions of raters A and B concerning life k ($k = 1, 2, \ldots, n$) as given by the scores x_k, y_k respectively, where x_k and y_k may take the values $0, 1, 2, \ldots, m-1$. Let n_{ij} be the number of lives for which $x_k = i$ and $y_k = j$, and let

$$n_{i\cdot} = \sum_{j=0}^{m-1} n_{ij} \qquad (i = 0, 1, \ldots, m-1)$$

$$n_{\cdot j} = \sum_{i=0}^{m-1} n_{ij} \qquad (j = 0, 1, \ldots, m-1)$$

(i) Using the weight function $(i-j)^2$, define weighted kappa, κ.

(ii) Show that

$$\kappa = \frac{\frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})(y_k - \bar{y})}{\frac{1}{2n}\left\{\sum_{k=1}^{n}(x_k - \bar{x})^2 + \sum_{k=1}^{n}(y_k - \bar{y})^2\right\} + \frac{1}{2}(\bar{x} - \bar{y})^2}.$$

(iii) State the circumstances under which κ is a good estimator of the (true) correlation coefficient between the opinions of the two raters. (You may assume that n is large.)

4. **(a)** You are given that Kolmogorov's Forward Equations for an n-state model are as follows (in the usual notation):

$$\frac{d}{dt}p_{ij}(x, x+t) = -\mu_j(x+t)p_{ij}(x, x+t) + \sum_{\nu \neq j} \mu_{\nu j}(x+t)p_{i\nu}(x, x+t) \quad (1 \leq i, j \leq n,\ t \geq 0)$$

where $\mu_j(y) = \sum_{\nu \neq j} \mu_{j\nu}(y)$ for $1 \leq j \leq n$.

Consider the life table model, with states 1 and 2 representing "alive" and "dead" respectively.

(i) Write down the Kolmogorov Forward Equation when $i = 1$ and $j = 1$.

(ii) Solve this equation to express $p_{11}(x, x+t)$ in terms of $\mu_{12}(x+r)$ $(0 \leq r \leq t)$.

(iii) Express the result of (ii) in traditional actuarial notation.

(b) It has been found that, at all ages, the force of mortality of those suffering from a certain condition is twice that of English Life Table No.12 - Males. You are given that $q_{60} = 0.02287$ according to English Life Table No.12 - Males. Find the chance that a life aged exactly 60 with the above condition will die within a year.

(c) The time to failure, T hours, of a certain type of component has distribution function

$$F_T(t) = 1 - \exp\left[-\frac{\alpha t^2}{2}\right] \qquad (t > 0,\ 0 \text{ otherwise})$$

where $\alpha = 2 \times 10^{-6}$.

(i) Find the hazard rate, $h(t)$ $(t > 0)$.

(ii) Find the modal lifetime of a component of this type.