UNIVERSITY OF CAMBRIDGE

MATHEMATICAL TRIPOS Part III

Monday, 13 June, 2011 9:00 am to 11:00 am

PAPER 37

BIOSTATISTICS

*Attempt no more than **THREE** questions, with at most **TWO** questions from **Survival Data***

.

*There are **FIVE** questions in total.*

The questions carry equal weight.

STATIONERY REQUIREMENTS
Cover sheet
Treasury Tag
Script paper

SPECIAL REQUIREMENTS
None

You may not start to read the questions printed on the subsequent pages until instructed to do so by the Invigilator.

1 Statistical In Medical Practice

Male prisoners sentenced to custody for up to a year and due for release from Prison P will be randomized in the ratio 1:1 to: "support as usual for making the transition from prison to community" or "transitional support services from Provider K". The extra cost of Provider K's service will be justified if there is a sufficient reduction in the "1-year reconviction rate" (defined as the proportion of randomized offenders who are reconvicted within 18 months for an offence committed in the year following their release).

(a) Advise the Ministry of Justice on how many eligible prisoners need to be randomized to have at least 80% power to detect a decrease in the "1-year reconviction rate" from 65% (support as usual) to 60% (Provider K).

Economists point out that the cost-effectiveness of Provider K's service depends on reducing the number of reconvictions in the 3 years following release from a mean of 3 (sd 8) for 'support as usual' to a mean of 2.5.

(b) Re-estimate how many eligible prisoners need to be randomized to have at least 80% power to detect the economists' target difference in the mean number of reconvictions within 3 years.

(c) Explain why a participant randomized in September 2010, who is due for re-release from Prison P in July 2011 after a 6-month sentence, should not be eligible for re-randomization.

Provider K thinks it would be unfair for half the eligible male releases from Prison P to be denied their transitional support service, and would prefer that a control group of eligible released male prisoners were selected from other prisons.

(d) Give two reasons that the Ministry of Justice should reject Provider K's preference on controls.

Transitional support services from an External Provider (J or K) cost £1,000 per eligible release on average. The Ministry of Justice decides on a randomized controlled trial involving 12 prisons to compare four options in respect of "1-year reconviction rate":

(i) transitional support services from External Provider K,

(ii) transitional support services from External Provider J,

(iii) prison-designed extra support at £500 per eligible release, versus

(iv) support as usual.

Prisons will be randomized to External Provider (J or K) and, within each prison, eligible releases are to be randomized in the ratio 2:1:1 to External Provider: prison-designed extra support: 'support as usual'.

(e) What advantage is there in randomizing prisons to Provider (J or K)?

(f) What advantage is there in comparing between equal-cost External Providers?

2 Statistical In Medical Practice

Table 1

Counts of events

Physicians	A	B	C	D	E	F	G	H	I	J	K	L
Pre-Training	9	1	7	2	10	5	6	1	7	8	7	16
Post-Training	2	11	21	7	8	9	0	9	17	8	10	8

Table 1 presents study data on the number of positive communication events initiated by physicians during filmed patient consultations. Data are presented from 12 physicians, one observation being taken prior to communication training and one after training for each physician.

Denote the counts for physicians i as y_{ij}, $i = 1, \ldots, 12$, $j = 1, 2$, where i indexes the physicians and $j = 1$ corresponds to the pre–intervention counts and $j = 2$ corresponds to the post–intervention count. Assume that x_{ij} is defined to be a binary covariate taking the value 0 for pre-intervention observations and 1 otherwise. Assume that y_{ij} is a Poisson variate with mean $\lambda_i \exp(\beta x_{ij})$.

1. Derive a conditional likelihood for the estimation of β that will eliminate the parameters $\lambda_i, i = 1, \ldots 12$. Express the conditional likelihood as a function of $p = \exp(\beta)/(1 + \exp(\beta))$.
2. The maximum conditional likelihood estimate, $\hat{\beta}_c$ of β is 0.33 with an estimated asymptotic standard errors of 0.15. Provide a 95% confidence interval for logarithm of the ratio of the mean number of positive communication events for a physician after training to that before training. Does this provide evidence for an effect of training?
3. Would the conditional maximum likelihood estimate of β derived in part (1) provide a sensible estimate if the y_{ij} were in fact negative binomially distributed?
4. Explain how the estimated asymptotic standard error of 0.15 from the conditional Poisson is derived and why it might be sensible also to consider an estimated asymptotic standard error based on a "information-sandwich estimator".
5. The information sandwich estimator of the asymptotic standard error of $\hat{\beta}_c = 0.33$ is 0.25. Given this, and the additional estimation results above, provide an assessment of the evidence for the effectiveness of the communication training which could be communicated to the study investigators.

You may find one or more of the following quantiles of the standard Normal distribution useful: $\Phi^{-1}(0.9) = 1.28, \Phi^{-1}(0.95) = 1.64, \Phi^{-1}(0.975) = 1.96$.

Note also (if useful): $\exp(0.15) = 1.16, \exp(0.25) = 1.28, \exp(0.33) = 1.39$

3 Analysis of Survival Data

(a) What is meant by a hazard function? How can the hazard function be derived from the survivor function?

Show that if two hazard functions $h_1(t)$ and $h_2(t)$ are related by $h_2(t) = \lambda h_1(t)$, with $\lambda > 0$, then the corresponding survivor functions are related by $F_2(t) = [F_1(t)]^{\lambda}$.

Two survivor functions are related by $F_4(t) = F_3(\kappa t)$, with $\kappa > 0$: find a relationship between the corresponding hazard functions.

(b) The time-to-event T_{ij} of the ith individual from the jth stratum is given by

$$\log T_{ij} = a_j + b \log U_{ij}$$

where the a_j and b are constants ($b > 0$) and the U_{ij} are independent random variables with an `exponential(1)` distribution. Show that the survivor distribution $F_j(t) = \mathcal{P}(T_{ij} > t)$ is given by:

$$F_j(t) = \exp\left\{ - \left[\frac{t}{\alpha_j} \right]^{\beta} \right\}$$

with $\alpha_j = \exp(a_j)$ and $\beta = 1/b$.

Consider two individuals from *different* strata. Show that their survival distributions belong both (i) to the same proportional hazards family and (ii) to the same accelerated life family.

4 Analysis of Survival Data

What is meant by an *empirical likelihood* estimate of the survivor function $F(t) = \mathcal{P}(T > t)$?

What is the contribution of the following individuals to the empirical likelihood function for F:

(a) an individual with an event at $T = 5$;

(b) an individual with an event known to have taken place before or at $T = 3$;

(c) an individual with an event known to have taken place in the interval $6 < T \leqslant 8$;

(d) an individual with an event known to have taken place strictly after $T = 10$?

A survival dataset includes just these four individuals. Write down the empirical likelihood function for F.

Justify simplifying the empirical likelihood function by replacing $F(10)$ by $F(8)$, $F(6)$ by $F(5)$ and $\lim_{\delta \downarrow 0} F(5 - \delta)$ by $F(3)$.

Show that the empirical likelihood can now be written as

$$[1 - F(3)]\,[F(3) - F(5)]\,[F(5) - F(8)]\,[F(8)]$$

and find the maximum empirical likelihood estimators $\hat{F}(3)$, $\hat{F}(5)$ and $\hat{F}(8)$.

What can be said, if anything, about $\hat{F}(4)$, $\hat{F}(7)$ and $\hat{F}(9)$?

5 Analysis of Survival Data

A survival dataset comprises n individuals with survival times x_i and censoring indicators v_i, with $i = 1, \ldots, n$, where $v_i = 0$ when x_i corresponds to a right-censored observation and $v_i = 1$ when x_i corresponds to an observed event. The integrated hazard $H_i(t)$ of the ith individual is given by:

$$H_i\,(t) = p_i t + \theta t^q$$

where the p_i and q are known constants ($p_i \geqslant 0$ for all i and $q > 0$) and θ is a parameter to.be estimated.

Describe how to set up the likelihood function for a survival distribution. Obtain the log-likelihood for θ.

(a) in the case $p_i = 0$ (for all i): obtain the maximum likelihood estimate $\hat{\theta}$ of θ. Give an expression for the second derivative of the log-likelihood, evaluated at $\hat{\theta}$. Indicate how you would obtain an interval estimate for θ.

(b) in the case $q = 1$: obtain an approximate expression for the maximum likelihood estimate of θ. (You may assume that p_i is small relative to θ, for all i.)

END OF PAPER