NEWCASTLE UNIVERSITY

SCHOOL OF MATHEMATICS & STATISTICS

SEMESTER 2 2007/2008

MAS3311

Biostatistics

Time allowed: 2 hours 15 minutes

Credit will be given for ALL answers to questions in Section A, and for the best THREE answers to questions in Section B. No credit will be given for other answers and students are strongly advised not to spend time producing answers for which they will receive no credit.

Marks for each question are indicated. However you are advised that marks indicate the relative weight of individual questions, they do not correspond directly to marks on the University scale.

There are FOUR questions in Section A and FOUR questions in Section B.

Calculators may be used. Statistical tables will be provided.

SECTION A

A1. In the context of a randomized controlled trial explain briefly:

(a) the advantage or advantages of Random Permuted Blocks over simple randomization

(b) the advantage or advantages of Analysis of Covariance over analysing change from baseline

[8 marks]

A2. (a) Let T denote the survival time of an individual, $f(t)$ be the probability density function and $S(t)$ the survivor function.

(i) Explain what is meant by the "hazard function."

(ii) Express the probability density function in terms of the survivor function.

(iii) Express the hazard function in terms of the probability density and survivor functions.

(iv) Show that

$$S(t) = \exp\{-H(t)\},$$

where

$$H(t) = \int_0^t h(u)\, du$$

is the cumulative hazard function.

(b) Using the above result, find the probability density function if T has hazard function

$$h(t) = \alpha\rho^{\alpha}t^{\alpha-1}.$$

[10 marks]

A3. Give the usual model for the outcome from an AB/BA crossover trial (you may assume that the outcome follows a Normal distribution). Suppose that the observed mean outcome in period $j = 1, 2$ of sequence AB is m_{1j}, with the corresponding quantities in sequence BA being m_{2j}. Show that both $m_{11} - m_{21}$ and $\frac{1}{2}[(m_{11} - m_{12}) - (m_{21} - m_{22})]$ are unbiassed estimators of the treatment effect. Which estimator would you use in practice? Justify your answer.

[12 marks]

A4. (a) Let $t_0, t_1, \ldots, t_n$ be increasing, equally spaced, times with $t_0 = 0$. Define intervals $I_0, \ldots, I_{n-1}$ where $I_j = [t_j, t_{j+1})$. For the actuarial estimator of the survivor function, if there is censoring, the adjusted number of individuals at risk N'_j for interval I_j is needed. Give the formula for N'_j, defining any terms you need and explaining the reasoning behind N'_j.

(b) The survival times in months are given below for 72 people with brain tumours. Censoring is indicated by an asterisk * next to the survival time.

1,	1,	1,	1,	1,	1,	2,	2,	2,	3,
3,	3,	3,	4,	4,	4,	4,	4,	4,	4,
5,	5,	5,	6,	6,	6,	6,	7,	7,	7,
7,	7,	7*,	7*,	8,	8,	8*,	8*,	9,	9,
9,	9,	9,	9,	9*,	10,	10,	10,	10*,	10*,
11*,	11*,	12*,	13*,	14*,	14,	16,	17*,	17,	21*,
23*,	24,	25*,	25*,	29*,	34*,	42*,	43*,	46*,	50*,
56,	58.								

Calculate the actuarial estimate of the survivor function, up to the end of the third year, using intervals of a year.

[10 marks]

SECTION B

B5. A trial to compare two treatments is planned to use groups of size m and n with Type I (two-sided) error α and Type II error β when the true difference is τ_M. If $1 - \xi = \Phi(z_\xi)$ where $\Phi(.)$ is the distribution function of a standard Normal variable, and the outcome variable in the trial has variance σ^2 then:

$$\frac{\tau_M}{\sigma\lambda} = z_{\frac{1}{2}\alpha} + z_\beta$$

where $\lambda = \sqrt{\frac{1}{m} + \frac{1}{n}}$.

(a) Show carefully that if this result is adapted to the case when the outcome is binary then

$$\frac{2(\arcsin\sqrt{p_1} - \arcsin\sqrt{p_2})}{\lambda} = z_{\frac{1}{2}\alpha} + z_\beta$$

where the trial is now designed to detect a change in success rates from p_2 to p_1, where $p_2 < p_1$

(b) By writing $p = \sin^2\theta$ show that:

$$\arcsin\sqrt{p} + \arcsin\sqrt{1-p} = \frac{1}{2}\pi$$

Hint: recall that $\cos\theta = \sin(\frac{1}{2}\pi - \theta)$

(c) How will the group sizes of a trial designed to detect a change in success rates from p_2 to p_1 relate to those for a trial designed to detect a change in failure rates from $1 - p_1$ to $1 - p_2$?

(d) By using the result in (b), reconcile your answer to part (c) with that to part (a).

[20 marks]

Question B6 continued on next page

B6. (a) Let $t_1, t_2, \ldots, t_k$ be observed failure times in increasing order, i.e. times at which at least one individual fails. Let $t_0 = 0$. Let d_j be the number who fail at time t_j. Let c_j be the number of censorings in the interval $[t_j, t_{j+1})$. Let N_j be the number at risk just before t_j. Define N_j in terms of N_{j-1}, d_{j-1} and c_{j-1}.

State the estimator of the conditional probability that an individual survives the interval $[t_j, t_{j+1})$, given that the individual survives to just before t_j, which is used in calculating the Kaplan-Meier estimates. Hence define the Kaplan-Meier estimator $\hat{S}(t)$ of the survivor function.

Show that, if there is no censoring,

$$\hat{S}(t) = \frac{N_{j+1}}{N_1}$$

for $t_j \leq t < t_{j+1}$. Hence show that, in this case, $\hat{S}(t)$ has variance approximately

$$\frac{\hat{S}(t)\{1 - \hat{S}(t)\}}{N}$$

where N is the number of individuals in the study.

(b) Experiments were performed as part of a research programme investigating motion sickness at sea. Subjects were randomaly allocated to three groups and the groups experienced motion at three different frequencies. The times, in minutes, until the subjects vomited were recorded. Some subjects requested an early stop to the experiment and no experiment lasted more than two hours. The data are given below. Censored observations, where the experiment was stopped at the request of the subject or the subject reached two hours without vomiting, are marked with an asterisk *.

Frequency										
Low	30,	50,	50*,	51,	66*,	82,	92,	120*,	120*,	120*.
Middle	11,	13,	24*,	30,	30,	36,	51,	66,	71,	120*.
High	5,	6,	11,	11,	13,	24,	51,	66,	68,	69.

(i) Find the Kaplan-Meier estimate of the probability $S(30)$ of a person in the high-frequency group lasting half an hour without

Question B6 continued on next page

vomiting and give an approximate 95% confidence interval for $S(30)$.

(ii) Find the Kaplan-Meier estimate of the probability $S(30)$ of a person in the middle-frequency group lasting half an hour without vomiting and give an approximate 95% confidence interval for $S(30)$.
You may use Greenwood's formula:

$$\mathsf{var}[\hat{S}(t)] \approx [\hat{S}(t)]^2 \sum_{j=1}^{k} \frac{d_j}{N_j(N_j - d_j)}$$

for $t_k \leq t < t_{k+1}$.

(iii) The log-rank test for two groups can be extended to compare the survivor functions for three groups. Give the first line of a suitable table and a formula for a suitable χ^2 statistic. (Use the simpler form of the log-rank test). The value of the test statistic is 9.189 . State clearly your conclusion.

[20 marks]

B7. A randomized trial of two groups each of 50 patients was performed to compare two bronchodilators, A and B, and the outcome variable of interest was peak expiratory flow rate (PEFR, in l/min), which can be assumed to follow a Normal distribution. The PEFR was measured at baseline and at outcome. Summaries of the results are shown in the table below (all values in l/min).

	Outcome Mean (SD)	Change from baseline Mean (SD)
A	295 (64.0)	2.9 (63.0)
B	323.5 (64.0)	29.4 (59.0)

(a) Using only the outcome data, estimate the mean treatment difference and its associated 95% confidence interval. [You should assume that the population outcome variances in the two treatment groups are equal]

(b) Using the change from baseline, estimate the mean treatment difference and its associated 95% confidence interval. [You should assume that the population variances of change in the two treatment groups are equal]

(c) Explain why the differences in parts (a) and (b) estimate the same quantity.

(d) Which analysis gives a more precise estimate of the treatment effect?

(e) Assume that the outcome variable has variance σ^2 and the baseline has variance σ_B^2 and that the correlation between outcome and baseline is ρ. Derive a condition on these parameters which would imply that a more precise estimate of treatment effect would be obtained by analysing the change from baseline.

Hint: you may wish to note that a random variable with a t-distribution on 98 degrees of freedom is less than 1.98 with probability 0.975

[20 marks]

B8. The times, in days, that 266 heroin addicts spent in either of two clinics were recorded. Some observations are right-censored. In addition there are three covariates, as follows.

Clinic	:	0 for Clinic 1, 1 for Clinic 2.
Prison	:	1 if the patient has a prison record, 0 otherwise.
Dose	:	Dose of methadone.

(a) A Weibull proportional hazards model was used to analyse these data. Models with each of the possible combinations of covariates were fitted by maximum likelihood. The resulting values of $-2\log L$, where L is the maximised likelihood, are as follows.

Model	$-2\log L$
null	2229.838
Dose	2195.360
Prison	2228.474
Clinic	2200.156
Dose, Prison	2193.868
Dose, Clinic	2172.502
Prison, Clinic	2197.500
Dose, Prison, Clinic	2168.954

Use suitable tests to compare these models and state your conclusions about the evidence for the effects of the covariates on the length of time spent in a clinic.

(b) A Weibull proportional hazards model is fitted with just the covariates Dose, x, and Clinic. The parameter estimates are as follows.

Variable	Estimated coefficient	Standard error
Clinic	1.538	0.106
Dose x	0.136	0.0081

Let $h_1(t)$ be the hazard function for a person at Clinic 1 with Dose $x = 20$, let $h_2(t)$ be the hazard function for a person at Clinic 1 with $x = 40$ and let $h_3(t)$ be the hazard function for a person at Clinic 2 with $x = 20$. Give 95% confidence intervals for each of $h_2(t)/h_1(t)$ and $h_3(t)/h_1(t)$.

Question B8 continued on next page

(c) Suppose that, as well as the lengths of stay, we had data on Prison and Clinic, but not Dose. Explain, with suitable formulae, how you would carry out a log-rank test, stratified by clinic, of the hypothesis that having a prison record has no effect on length of stay.

[20 marks]

THE END