

## Medical Statistics

### Exercises 3

**(Released 28 November 2011 work submitted by ~Tuesday 13 December 2011 will be marked and returned. Work submitted after solutions are made available will not be marked)**

- 1) The table below gives some details of fitting a proportional hazards regression model to times to recurrence of a certain disease. The data were obtained during a randomised clinical trial of a new treatment. The factors investigated were treatment (coded by  $x_1 = 0$  for placebo,  $x_1 = 1$  for treatment), stage of disease (coded by  $x_2 = 0$  for stage I,  $x_2 = 1$  for stage II,  $x_2 = 2$  for stage III) and the interaction between treatment and stage of disease (coded by  $x_3$  where  $x_3 = x_1 \times x_2$ )

	variable	coefficient	standard error
Treatment	$x_1$	-0.18	0.10
Stage	$x_2$	+0.32	0.21
Interaction	$x_3$	-0.66	0.11

- i) Specify the form of the proportional hazards model used for this analysis in terms of the baseline hazard function  $h_0(t)$  and the covariates.
- ii) Describe in detail the effects of these factors on the time to recurrence of the disease.
- iii) Show diagrammatically the form of the relationship between the survivor functions and the stage of the disease for the two different treatment groups.



2) The data in the file `AHprostate.Rdata` have been adapted from Andrews and Herzberg (1985) and give results of a trial on treatments for prostate cancer. Various covariates were recorded. The variables in the data file are given in the table below.

Variable	Description	Levels
Stage	Stage of Disease	3=No evidence of distant metastasis, 4=evidence of distant metastasis
RX	Treatment Groups	1=Control Arm, 2=Experimental Arm
Dtime	Complete months to follow up	
Status	Survival Status	
AgeYrs	Age of Patient	89 denotes 89 or more
Wt	Weight Index	Weight (Kg) – Height (cm) + 200
PF	Performance Rating	
HX	History of cardiovascular disease	0=no, 1=yes
SBP	Systolic blood pressure	
DBP	Diastolic blood pressure	
EKG	Electrocardiogram	
HG	Serum Haemoglobin,g/100ml	
SZ	Size of primary tumour	cm <sup>2</sup>
SG	Combined index of tumour	stage and histologic grade
AP	Serum prostatic acid phosphatase	
BM	Bone metastases	0=no, 1=yes

Investigate whether there is evidence that the treatment improves survival time to follow up, making due allowance for any other prognostic factors on the subjects.



- 3) The data file Prostatic given in S-Plus, SPSS and Minitab formats contains data on a double blind randomised controlled clinical trial to compare treatments for prostatic cancer. The data are extracted from Collett (2003) who gives the original reference. The data file contains records for each patient of the treatment received (coded as 0 or 1 for placebo and 1.0 mg of diethylstilbestrol respectively, treatments being administered daily by mouth), survival time from entry to trial, with a status variable indicating whether or not the observation was censored (value 0) or complete (value 1), age at entry to the trial, serum haemoglobin level in gm/100ml, size of primary tumour in  $\text{cm}^2$  and the value of a combined index of tumour stage and grade (the Gleason Index), larger values indicating a more advanced stage of tumour.
- i) Construct Kaplan-Meier plots of the survival times for the two treatment groups.
  - ii) Making allowance for the values of the various covariates, assess whether the data provide evidence that the two treatment groups experience different survival prospects.
  - iii) Construct a log–log plot for treatment, averaging over other covariates.
  - iv) ★ Choosing any parametric regression (see Survival tasks 4) model which does **not** have the proportional hazards property, fit the model and assess whether this alters your conclusions reached in part ii).
  - v) ★ Choosing a parametric AFT model, estimate the parameters and compare your conclusions with those from parts ii) and iv).
- [credit will not be lost if parts iii) – v) are not submitted, they are for ‘interest’ and as an aid to those continuing to MAS6062]

