ST745, Spring 2009

Final Examination, 1:00-4:00PM, Tuesday, April 28th, 2009

1. Let Z be a time-independent covariate. Assume that conditioning on the covariate, the survival function is given by

$$S(t|Z) = 1/\{1 + \exp(\beta Z)t^{1/2}\},\ t > 0.$$

   (a) Calculate the corresponding hazard function. Show that it is not from the proportional hazards model. (5 pt)

   (b) We collect a data set $(x_i, \delta_i, z_i), i = 1, ..., n$, where x_i is the censoring/death time, δ_i is the censoring indicator taking three values (0=right censoring, 1=left censoring, 2=death). Write down the likelihood function of the data in terms of β. (6 pt)

2. The following table gives data from patients on a study to examine the possible benefit of the kidney transplant:

ID	x (year)	δ	`wait` (year)
1	1.1	1	.
2	3.0	1	0.8
3	4.2	1	1.2
4	2.5	0	.
5	3.5	0	2.5
6	5.0	0	3.2

where x is the observed survival or censoring time measured from the beginning of the study, δ is the death indicator (1 for death and 0 for censoring) and `wait` is the time in years that a patient waited to get a kidney transplant (missing means a patient did not get a transplant). Assume the following model for evaluating the possible benefit of the kidney transplant:

$$\lambda(t|\text{transplant history}) = \lambda_0(t)\exp(\beta z_i(t)), \tag{1}$$

where

$$z_i(t) = \begin{cases} 1 & \text{if patient } i \text{ already received a kidney transplant at time } t \\ 0 & \text{otherwise} \end{cases}$$

and t is measured as time from the entry into study. Based on the above model, answer the following questions:

(a) Construct the partial likelihood function for the proportional hazards model (1) using the data given in the above table. (8 pts)

(b) Conduct the score test for the null hypothesis $H_0 : \beta = 0$ (Here we choose the α-level as 0.05). (8 pt)

(c) Write down the SAS code for fitting the model (1) with time-dependent covariate. (8 pts)

3. The following SAS program was fitted to data obtained from a clinical trial comparing treatment 1 (trt=1) to treatment 0 (trt=0) when adjusting for the age effect:

```
proc lifereg;
  model months*cens(0) = trt age / dist = llogistic;
run;
```

where `months` is the (censored) survival time in months, `cens` is the failure indicator (1=survival time, 0=censoring time). Suppose we got the following output:

```
                      Analysis of Parameter Estimates

                             Standard   95% Confidence         Chi-
Parameter      DF Estimate     Error        Limits            Square   Pr > ChiSq

Intercept       1   2.5018    0.0712    2.3622    2.6415     1234.65       <.0001
trt             1   0.4201    0.0922    0.2394    0.6008       20.76       <.0001
age             1  -0.1253    0.1000   -0.3253    0.0747        1.57        0.790
Scale           1   0.5463    0.0397    0.4669    0.6257
```

(a) Write down the statistical model the above SAS program fitted. (5 pts)

(b) For patients with the same age, compute the percentage increase/decrease in the average survival time between two treatments. (5 pts)

(c) Compute the odds ratio of two patients: one from treatment 1 with age of 60 and the other from treatment 0 with age of 50. (5 pts)

4. In a clinical trial to compare a new treatment (`trt=1`) to a standard treatment (`trt=0`), censored survival times were obtained and the following SAS program was run

```
proc phreg;
  model years*cens(0) = trt trtyr;
  if years < 3 then do;
    trtyr = trt;
  end;
  else do;
    trtyr = 0;
  end;
run;
```

where `years` is censored survival times in years and `cens` is the failure indicator (1=survival time, 0=censoring time). The following is part of the output from the above program:

```
                Analysis of Maximum Likelihood Estimates

                        Parameter      Standard
Variable     DF          Estimate         Error    Chi-Square    Pr > ChiSq

trt           1            0.5000        0.0500        100.00        <.0001
trtyr         1           -1.0000        0.1000        100.00        <.0001
```

(a) Write down the statistical model fitted by the above `SAS` program. (5 pts)

(b) Based on the above `SAS` output, describe in detail the treatment effect. For example, is the treatment always beneficial? If it is, what is the effect? If the treatment is not always beneficial, when it is beneficial and when it is not? etc. You can write whatever you feel comfortable. (5 pts)

5. The following data gives number of patients who died and were censored in each interval (measured in years):

interval	# of patients	# of death	# of censoring
$[0, 1)$	40	2	4
$[1, 2)$	30	4	2
$[2, 3)$	20	4	2

Assuming censoring could occur **only at the end** of each interval, do the following:

(a) Estimate the conditional probability that a patient dies in interval [1, 2) given that he or she is still alive at the beginning of year 1. Find a 95% CI for this probability. (8 pts)

(b) Estimate the probability that a patient will survive beyond year 3. Find a 95% CI for this probability. (8 pts)

6. Suppose an investigator wants to see the effect of smoking on life expectancy. The investigator thinks that gender might be a confounder so she collected a data set like the following

x (year)	δ	`smoke`	`sex`
70	0	0	1
60	1	1	1
65	1	1	1
60	1	1	0
70	1	0	0
80	0	0	0

where x is the observed survival or censoring time, δ is the death indicator (1 = death, 0 = censoring), `smoke` is the smoking indicator (1 = smoker, 0 = non-smoker) and `sex` is the gender indicator (1 = male, 0 = female). After some exploratory data analysis, she decided to fit the following proportional hazards model

$$\lambda(t|Z_1, Z_2) = \lambda_0(t)\exp(\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_1 * Z_2), \tag{2}$$

where Z_1 denotes the smoking indicator and Z_2 denotes the gender indicator.

(a) Construct the partial likelihood function of the data based on the model (2). (Note: using the Breslow's approximation for handling ties) (8 pts)

(b) Interpret the meaning of the parameter β_3. Write down the SAS code for conducting the score test for testing $\beta_3 = 0$ in the above model. (8 pts)

(c) Another investigator questioned the assumption of model (2). He thought that the proportional hazards assumption for gender may not be right. He suggested to fit a stratified proportional hazards model for the smoking status while stratifying on the gender indictor. Write down the stratified proportional hazards model and the corresponding partial likelihood function of the data given in the table. (8 pts)