Analysis of Survival Data: Task Sheet 4

Notes & Solutions

1) *The table below gives details of a proportional hazards model fitted to some data obtained from patients being treated for kidney failure where 'survival time' is in terms of time to relapse.*

Variable	Coefficient	Standard Error	χ^2 statistic (using L.R.T)	p-value
Treatment 0 =Treatment A 1 =Treatment B	-1.63	0.75	4.71	< 0.05
Age (years)	-0.003	0.024	0.01	>>0.10
Sex 0 = female 1 = male	0.67	0.32	3.91	< 0.05
Obesity 0 = no 1 = yes	0.0092	0.0045	4.44	< 0.05
Duration of symptoms prior to treatment (months)	-0.003	0.075	0.01	>>0.10

Describe the effects of treatment and additional covariates on time to relapse.

It is clear that there is little evidence that either the subject's age or the duration of symptoms affect the relapse time. There is good evidence that (a) Treatment B gives a longer time to relapse, (b) females have a longer relapse time than males and (c) obese subjects have shorter relapse times than non-obese ones.

The ratio of hazards for subjects on treatment B to treatment A (with otherwise common values of covariates) is $e^{-1.63} = 0.196$ with 95% CI (0.044, 0.878). Similarly for males relative to females the corresponding figures are 1.954 with 95%CI (1.03, 3.706) and for obese to non-obese 1.009 and 95%CI (1.0002, 1.0184) (i.e. actually very little effect). For an increase of one year in age, the 95%CI for the proportional change in hazard is (0.950, 1.04) and for an increase of one month in duration of symptoms it is (0.985, 1.012).

In summary, the most important effects on relapse time are the treatment, treatment B reducing the hazard of relapse to about a fifth of that on treatment A, and the sex of the subject, with males having a hazard of about twice that of comparable females.

Analysis of Survival Data: Task Sheet 5

Notes & Solutions

1) *Now that the course is complete this is a good time to consolidate the material by reading the introduction and summary sections of all the chapters, perhaps going back to particular sections if there are items in the summaries which you feel you did not understand fully at the time.*

   Trust that you have done this.

2) *The data given below represent survival times for lymphoma patients according to the stage of tumour (where * denotes a censored value):*

Stage 3	*6*	*20*	*42*	*43**	*169**	*207*	*253*	*255**		
Stage 4	*4*	*10*	*20*	*21**	*30*	*33**	*43**	*46*	*110*	*235**

   i) *Compute the Kaplan-Meier product limit estimates of the survivor functions for stage 3 and stage 4 separately.*

   ii) *Provide estimates of the two cumulative hazard functions and comment on any differences.*

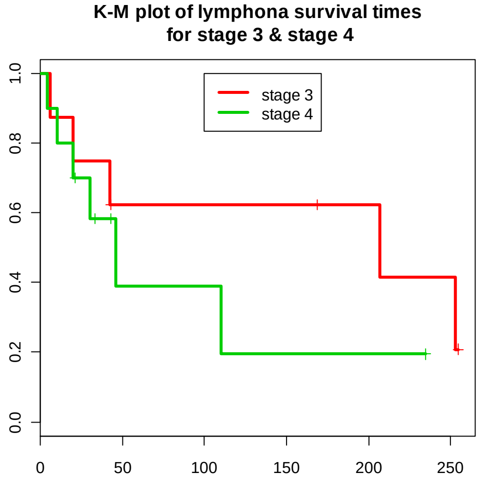
By using the log-rank test, compare the survival distributions for the two stages.

First the S-Plus version, first using the menu in `Statistics>Survival>Nonparametric Survival…`). Note the production of the Survival plot on a logarithmic vertical scale by clicking the appropriate box on the dialogue box of the plot menu.

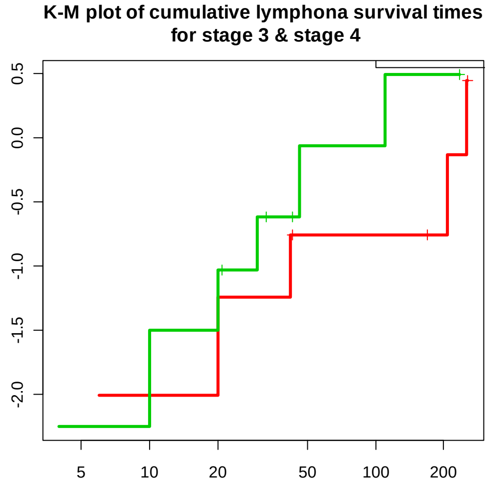
```
> library(survival)
Loading required package: splines
> attach(lymphoma)
> lymph.sv<-Surv(time,censor)
> lymphsurv<-survfit(lymph.sv~stage)
>
> summary(lymphsurv)
Call: survfit(formula = lymph.sv ~ stage)

                stage=3
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    6      8       1    0.875   0.117       0.6734            1
   20      7       1    0.750   0.153       0.5027            1
   42      6       1    0.625   0.171       0.3654            1
  207      3       1    0.417   0.205       0.1590            1
  253      2       1    0.208   0.179       0.0385            1
```

```
                stage=4
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    4     10       1    0.900  0.0949       0.7320            1
   10      9       1    0.800  0.1265       0.5868            1
   20      8       1    0.700  0.1449       0.4665            1
   30      6       1    0.583  0.1610       0.3396            1
   46      3       1    0.389  0.1916       0.1480            1
  110      2       1    0.194  0.1676       0.0359            1
> plot(lymphsurv,lwd=3,col=c(2,3),
+ main="K-M plot of lymphona survival times\n for stage 3 & stage 4")
> legend(100,1,c("stage 3", "stage 4"), lwd=3, col=c(2,3))
>
```




```
> plot(lymphsurv,lwd=3,col=c(2,3),fun="cloglog",
+ main="K-M plot of cumulative lymphona survival times\n for stage 3
& stage 4")
> legend(100,1,c("stage 3", "stage 4"), lwd=3, col=c(2,3))
>
```




```
>
```

For the log rank test has to be obtained from the command line:

```
> survdiff(lymph.sv~stage)
Call:
survdiff(formula = lymph.sv ~ stage)

         N Observed Expected (O-E)^2/E (O-E)^2/V
stage=3  8        5     6.37     0.296     0.804
stage=4 10        6     4.63     0.408     0.804

 Chisq= 0.8  on 1 degrees of freedom, p= 0.37
>
```

3) *In an accelerated-life survival model the survivor function for an individual with covariate x satisfies*

$$S(t{:}x) = S_0(te^{\beta x}),$$

where $S_0(t)$ is some baseline survivor function.

i) *Show that the corresponding hazard function satisfies*

$$h(t{:}x) = e^{\beta x}h_0(te^{\beta x})$$

where $h_0(t)$ is the baseline hazard function for $S_0(t)$.

$h(t;x) = –S′(t;x)/S(t;x) = –e^{\beta x}S_0′(te^{\beta x})/ S_0(te^{\beta x}) = e^{\beta x}\{S_0′(t;x)/ S_0(t;x) = e^{\beta x}h_0(te^{\beta x})$.

ii) *In a trial where n independent patients, with covariate values $x_1,x_2,\ldots,x_n$ enter at the same time, suppose that all death times are observed and that $S_0(t) = e^{-\lambda t}$ (t>0). Show that the survival time T has an exponential form, and is of proportional hazard form.*

$S_0(t) = e^{-\lambda t}$ so $h_0(t) = -(-\lambda e^{-\lambda t})/e^{-\lambda t} = \lambda$. So T_i has survivor function $\exp\{-\lambda t\exp(\beta x_i)\}$ and hazard function $\lambda\exp(\beta x_i)$ so is of proportional hazards form *iii)* ***** *Show that the distribution of the time to the first death in the trial is exponential with mean*

$$[\lambda\Sigma\exp(\beta x_i)]^{-1}$$

Let $M=\min\{T_1,T_2,\ldots,T_n\}$ then $S_M(t) = P[T_1>t, T_2>t,\ldots,T_n>t]$

$$=\prod_{i=1}^{n}P[T_i>t]=\prod_{i=1}^{n}S(t;x_i)=\prod_{i=1}^{n}\exp\{-\lambda te^{\beta x_i}\}=\exp\{-\lambda t\Sigma e^{\beta x_i}\}$$

and so M is exponential with mean $[\lambda\Sigma e^{\beta x_i}]^{-1}$.

iv) ***** *Show that the probability that the j^{th} patient is the first to die is given by $\exp(\beta x_j)/\Sigma\exp(\beta x_i)$*

j^{th} is the first if $T_1>T_j$, $T_2>T_j,\ldots ,_{\text{not } j,},\ldots,T_n>T_j$ so this has probability $P[T_1>T_j, T_2>T_j,\ldots,_{\text{not } j,},\ldots,T_n>T_j]$ `

$$=\int_{t_j=0}^{\infty}P[T_1>T_j,\ldots,_{\text{not } j},\ldots T_n>T_j \mid T_j=t_j]f(t_j;x_j)dt_j$$

$$=\int_{t_j=0}^{\infty}P[T_1>T_j,\ldots,_{\text{not } j},\ldots T_n>T_j]f(t_j;x_j)dt_j$$

$$=\int_{t_j=0}^{\infty}\prod_{\substack{i=1\\i\neq j}}^{n}\exp\{-\lambda t_j e^{\beta x_i}\}\lambda e^{\beta x_j}\exp\{-\lambda t_j e^{\beta x_j}\}dt_j=\int_{t_j=0}^{\infty}\lambda e^{\beta x_j}\exp\{-\lambda t_j\sum_{i=1}^{n}e^{\beta x_i}\}dt_j$$

$$=\frac{e^{\beta x_j}}{\sum_{i=1}^{n}e^{\beta x_i}}$$