Medical Statistics: Exercises 3

Notes & Solutions

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.*

The form of the model is $h(t)=h_0(t)\exp\{\beta_1x_1+\beta_2x_2+\beta_3x_3\}$ where $h_0(t)$ is the baseline hazard function and x_i (i=1,2,3) are as defined above. For subjects on placebo this becomes

$h(t)=h_0(t)$ for stage I

$h(t)=h_0(t)\exp\{\beta_2\}$ for stage II

$h(t)=h_0(t)\exp\{2\beta_2\}$ for stage III

and for those receiving treatment it is

$h(t)=h_0(t)\exp\{\beta_1\}$ for stage I

$h(t)=h_0(t)\exp\{\beta_1+\beta_2+\beta_3\}$ for stage II

$h(t)=h_0(t)\exp\{\beta_1+2\beta_2+2\beta_3\}$ for stage III *ii) Describe in detail the effects of these factors on the time to recurrence of the disease.*

estimated values of $h(t)/h_0(t)$ (i.e. hazard ratio relative to those on placebo at stage I), with approximate 95%CIs, are

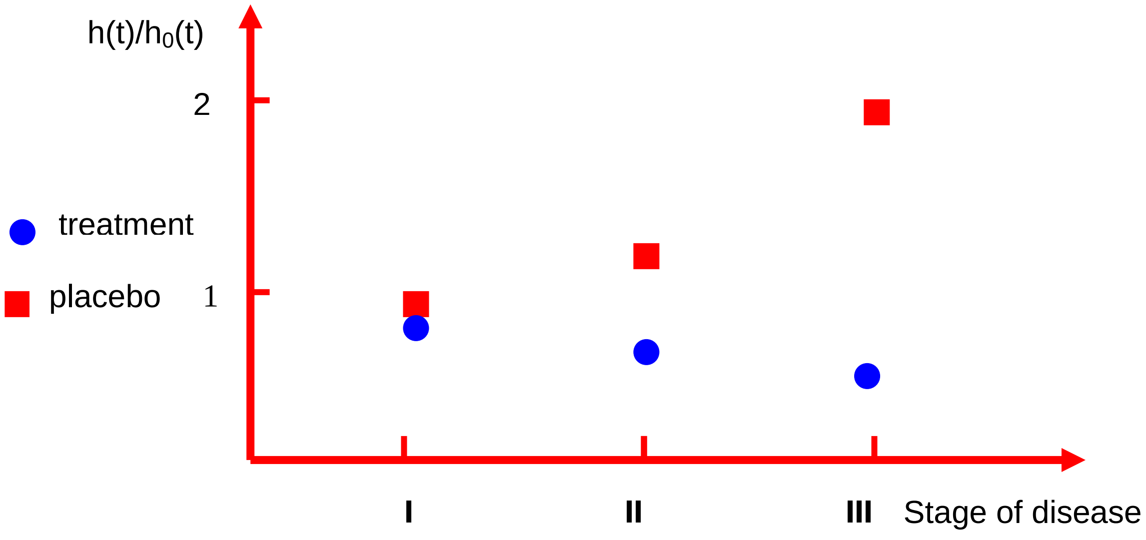
placebo, stage II:	1.38, (0.90, 2.10)
placebo:stage III:	1.90, (0.82, 4.39)
treatment, stage I:	0.84, (0.68, 1.22)
treatment, stage II:	0.59, (0.36, 0.99)
treatment, stage III:	0.42, (0.16, 1.11)

(assuming estimates of the β_i are independent). Note, CIs calculated as estimate+/– 2×s.e. and e.g. s.e.$(\beta_1+\beta_2)=(0.10^2+0.21^2)^{½}$ etc. and to get CI of e.g. $\exp\{\beta_1\}$ take exp{CI for β_1}.

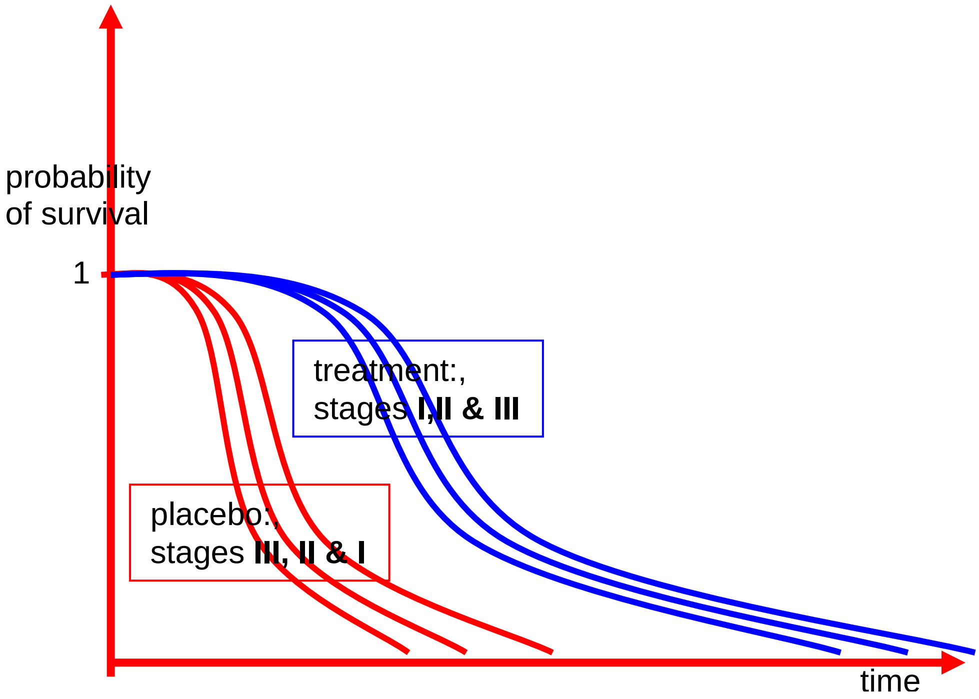
Thus, on untreated patients the hazards increase with stage of disease and so their survival prospects decrease with stage of disease. For patients on the treatment the effect of stage of disease is negated and indeed perhaps slightly reversed, though looking att he CIs the evidence for an actual improvement in survival with stage of disease is weak.

iii) Show diagrammatically the form of the relationship between the survivor functions and the stage of the disease for the two different treatment groups.

First, a diagram of the hazard ratios:



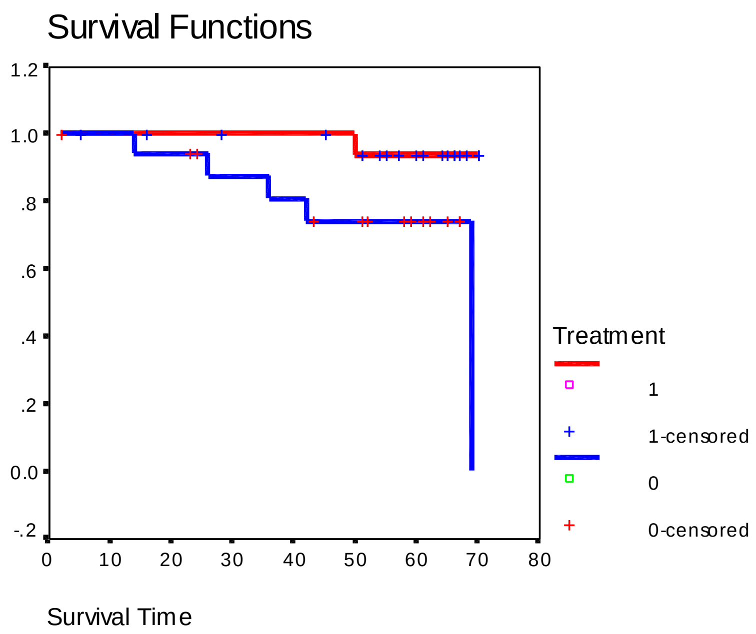
This allows a sketch of the relative positions of the survival times



2) *The data file Prostatic given in SPSS, S-PLUS 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 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.

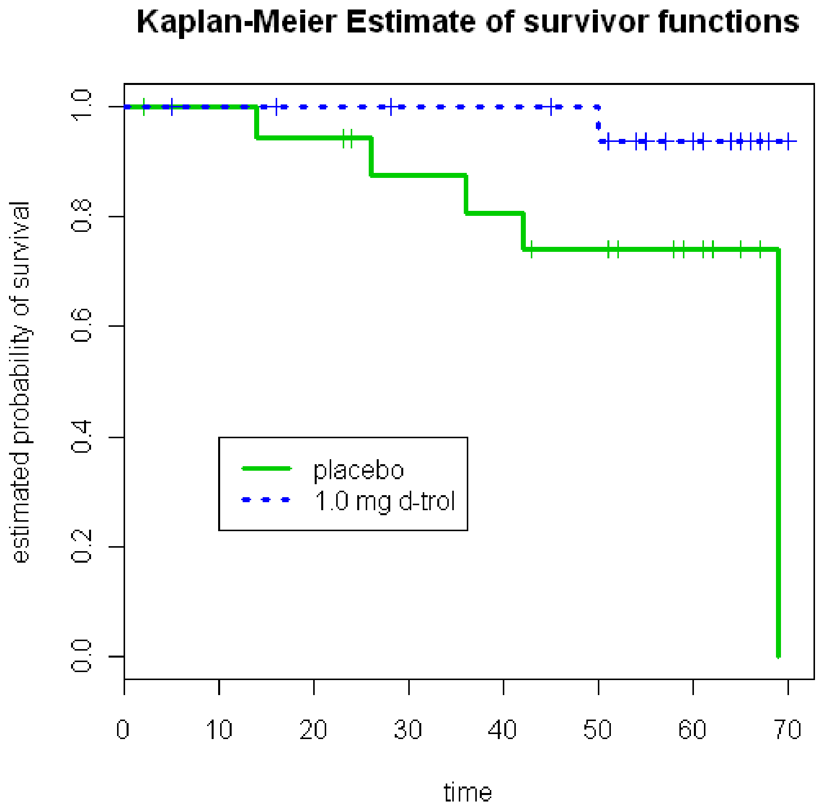
Plot produced by SPSS, Analyze>Survival>Kaplan-Meier, (and then edited to make the lines thicker and more distinct colours — double click on picture in SPSS to call up chart editor to do this).



Note the large number of censored cases with only 1 death on treatment and that treatment has higher survival prospects than placebo.

In **R** we have

```
> attach(prostatic)
> prostatic[1:5,]
  Treatment Survival.Time Status Age Serum.Haem. Tumour.Size Gleason.Index
1         0            65      0  67        13.4          34             8
2         1            61      0  60        14.6           4            10
3         1            60      0  77        15.6           3             8
4         0            58      0  64        16.2           6             9
5         1            51      0  65        14.1          21             9
>
> library(survival)
Loading required package: splines

> prostatic.sv<-Surv(Survival.Time,Status)
> prostaticfit <- survfit(Surv(Survival.Time,Status)~ Treatment)
>
> plot(prostaticfit, lty=c(1,3), lwd=3, col=3:4,
+ main="Kaplan-Meier Estimate of survivor functions",xlab="time",
+ ylab="estimated probability of survival")
> legtext<-c("placebo", "1.0 mg d-trol")
> legend(10,0.4,legtext,lty=c(1,3), lwd=3, col=3:4)
>
```



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.

Performing a Cox Regression in SPSS with Analyze>Survival>Cox Regression gives the following. Here Treatment has been declared as categorical but then the reference category has been changed from 'last' to 'first' (& then click on 'change') to ensure that it keeps the effective coding of Treatment as 0 for placebo and 1 for treatment instead of swapping them around, so coefficient < 0 indicates enhanced survival. In S-PLUS this is not necessary and apart from this the parameter estimates and standard errors etc are essentially identical. Investigation of treating 'Gleason' as categorical reveals this is not sensible since there is evidence of gross over-fitting (large estimates with enormous standard errors).

Variables	B	SE	Sig.	Exp(B)
TREATMEN	-1.182	1.210	.329	.307
AGE	.044	.072	.541	1.045
SERUM_HA	-.022	.453	.961	.978
TUMOUR_S	.094	.052	.071	1.099
GLEASON	.723	.350	.039	2.061

The conclusion to be drawn is that although the hazard ratio of those on treatment to placebo is estimated as about 0.3 there is little evidence that this is not due to the differing values of the covariates in the two treatment groups, notably Gleason index (treated on a linear scale) and tumour size.

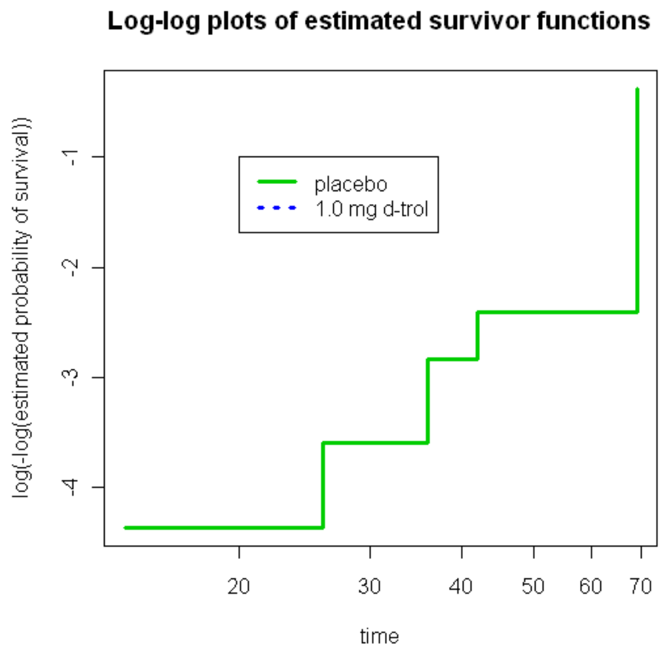
In **R** we have

```
> prostatic.ph<-coxph(prostatic.sv ~ Treatment + Age + Serum.Haem. +
Tumour.Size + Gleason.Index)
> options(digits=2)
> summary(prostatic.ph)
Call:
coxph(formula = prostatic.sv ~ Treatment + Age + Serum.Haem. +
    Tumour.Size + Gleason.Index)

  n= 38

                 coef exp(coef) se(coef)     z Pr(>|z|)
Treatment     -1.1821    0.3066   1.2103 -0.98    0.329
Age            0.0440    1.0450   0.0720  0.61    0.541
Serum.Haem.   -0.0221    0.9781   0.4527 -0.05    0.961
Tumour.Size    0.0940    1.0985   0.0521  1.80    0.071 .
Gleason.Index  0.7234    2.0615   0.3500  2.07    0.039 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
Treatment         0.307      3.261    0.0286      3.29
Age               1.045      0.957    0.9074      1.20
Serum.Haem.       0.978      1.022    0.4027      2.38
Tumour.Size       1.099      0.910    0.9919      1.22
Gleason.Index     2.061      0.485    1.0382      4.09

Rsquare= 0.311   (max possible= 0.616 )
Likelihood ratio test= 14.2  on 5 df,   p=0.0145
Wald test            = 10.1  on 5 df,   p=0.0735
Score (logrank) test = 15  on 5 df,   p=0.0104
```

iii) *Construct a log–log plot for treatment, averaging over other covariates.*

In **R**

```
> prostatic.ph2<-coxph(prostatic.sv ~ strata(Treatment) + Age +
Serum.Haem. + Tumour.Size + Gleason.Index)
> plot(survfit(prostatic.ph2),fun="cloglog", lty=c(1,3), lwd=3,
col=3:4,
+ main="Log-log plots of estimated survivor functions",xlab="time",
+ ylab="log(-log(estimated probability of survival))")
> legtext<-c("placebo", "1.0 mg d-trol")
> legend(20,-1,legtext,lty=c(1,3), lwd=3, col=3:4)
>
```



Note that the estimated survivor function for those on medication takes only two values (since there is just one event) and since log(–log(1.0)) = log(0) = ∞ the plot of that function is suppressed.

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).*

Choosing the lognormal distribution (which does not have the proportional hazards property and using `survreg()` gives

```
> prostatic.ln<-survreg(prostatic.sv ~ Treatment + Age +
+Serum.Haem. + Tumour.Size + Gleason.Index, dist="lognormal")
> summary(prostatic.ln)
                Value Std. Error      z      p
(Intercept)   10.1405     4.2999  2.358 0.0184
Treatment      0.7338     0.4593  1.598 0.1101
Age           -0.0226     0.0349 -0.646 0.5186
Serum.Haem.   -0.0449     0.1426 -0.315 0.7531
Tumour.Size   -0.0288     0.0203 -1.422 0.1552
Gleason.Index -0.3285     0.1753 -1.875 0.0609
Log(scale)    -0.4799     0.3123 -1.536 0.1244

Scale= 0.619
```

(omitting some details).

At first sight these results may appear to be substantially different from those using the Cox model (e.g. signs of coefficients are reversed) but this is because parametric models consider survival times rather than hazard rates.

v) ✭ Choosing a parametric AFT model, estimate the parameters and compare your conclusions with those from parts ii) and iv).

Accelerated failure time models are available in the library `eha` which must be downloaded and opened and regression function `aftreg()`.

```
> library(eha)
> prostatic.gp<-aftreg(prostatic.sv ~ Treatment + Age +
Serum.Haem. +
+ Tumour.Size + Gleason.Index, dist="gompertz")
>
> summary(prostatic.gp)
Call:
aftreg(formula = prostatic.sv ~ Treatment + Age + Serum.Haem.
+
    Tumour.Size + Gleason.Index, dist = "gompertz")

Covariate           W.mean      Coef Exp(Coef)  se(Coef)    Wald p
Treatment            0.566    -0.956     0.384     1.127     0.396
Age                 68.043    -0.005     0.995     0.043     0.903
Serum.Haem.         14.086     0.132     1.141     0.369     0.720
Tumour.Size         10.210     0.064     1.066     0.034     0.061
Gleason.Index        9.037     0.486     1.625     0.240     0.043

log(scale)                    12.732 338331.237     6.359     0.045

 Shape is fixed at  1

Events                    6
Total time at risk          1890
Max. log. likelihood      -33.237
LR test statistic         14.2
Degrees of freedom        5
Overall p-value           0.0144738
>
```

[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]