APTS - Survival Analysis
Lab Session 2 - Solutions

Ingrid Van Keilegom

August 25, 2017

1. (a)
```
> library("survival")
> library("KMsurv")
> data(burn)

> fit1 = coxph(Surv(T3,D3)~Z1,data=burn)
> summary(fit1)
Call:
coxph(formula = Surv(T3, D3) ~ Z1, data = burn)

  n= 154, number of events= 48

      coef exp(coef) se(coef)      z Pr(>|z|)
Z1 -0.5614    0.5704   0.2934 -1.914   0.0557 .
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

   exp(coef) exp(-coef) lower .95 upper .95
Z1    0.5704      1.753     0.321     1.014

Concordance= 0.566  (se = 0.039 )
Rsquare= 0.024   (max possible= 0.942 )
Likelihood ratio test= 3.73  on 1 df,   p=0.05347
Wald test            = 3.66  on 1 df,   p=0.05567
Score (logrank) test = 3.76  on 1 df,   p=0.05256
```

   (b)
```
> fit2 = coxph(Surv(T3,D3)~Z1+Z4,data=burn)
> summary(fit2)
Call:
coxph(formula = Surv(T3, D3) ~ Z1 + Z4, data = burn)

  n= 154, number of events= 48

        coef exp(coef)  se(coef)      z Pr(>|z|)
Z1 -0.524764  0.591695  0.295769 -1.774    0.076 .
Z4  0.007248  1.007275  0.007145  1.015    0.310
```

```
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

   exp(coef) exp(-coef) lower .95 upper .95
Z1    0.5917     1.6901    0.3314     1.056
Z4    1.0073     0.9928    0.9933     1.021

Concordance= 0.585  (se = 0.046 )
Rsquare= 0.03   (max possible= 0.942 )
Likelihood ratio test= 4.7  on 2 df,   p=0.09551
Wald test            = 4.72  on 2 df,   p=0.09461
Score (logrank) test = 4.82  on 2 df,   p=0.08972

> fit3 = coxph(Surv(T3,D3)~Z4,data=burn)
> summary(fit3)
Call:
coxph(formula = Surv(T3, D3) ~ Z4, data = burn)

  n= 154, number of events= 48

       coef exp(coef) se(coef)    z Pr(>|z|)
Z4 0.008906  1.008946 0.007010 1.27    0.204

   exp(coef) exp(-coef) lower .95 upper .95
Z4     1.009     0.9911    0.9952     1.023

Concordance= 0.552  (se = 0.045 )
Rsquare= 0.01   (max possible= 0.942 )
Likelihood ratio test= 1.49  on 1 df,   p=0.2215
Wald test            = 1.61  on 1 df,   p=0.2039
Score (logrank) test = 1.62  on 1 df,   p=0.2025
```

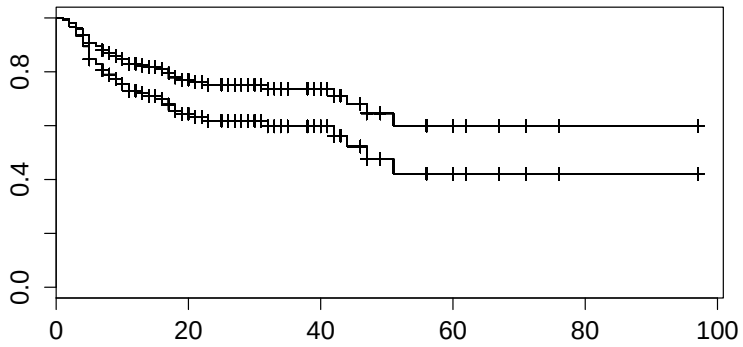
P-values :

- Wald test : 0.076
- Likelihood ratio test : $T_{LR} = 4.70 - 1.49 = 3.21 \Rightarrow P\text{-value} = 0.073$

```
> 1-pchisq(3.21,1)
[1] 0.07318948
```

(c)
```
> fit4 = survfit(fit2,newdata=data.frame(Z1=c(0,1),Z4=25),conf.type="log-log")
```

```
> plot(fit4)
```



(d)
```
> mat=cbind(fit4$time,fit4$surv,fit4$lower,fit4$upper)
> t20=(fit4$time==20)
> mat[t20,c(2,4,6)]
        1         1         1
0.6442330 0.5166641 0.7461969
> mat[t20,c(3,5,7)]
        2         2         2
0.7709242 0.6632128 0.8480441
```

2. (a) We proceed as follows : the variable `stage` is categorical with 4 levels. For the estimation, the levels 2, 3 and 4 will be recoded as binary variables. Level 1 is considered as the reference level.

```
> library(KMsurv)
> library(survival)
> data(larynx)
> Fit = survreg(Surv(time=time,event=delta,type="right") ~ as.factor(stage)
+               + age, data=larynx, dist="weibull")
> summary(Fit)

Call:
survreg(formula = Surv(time = time, event = delta, type = "right") ~
    as.factor(stage) + age, data = larynx, dist = "weibull")
                    Value Std. Error      z        p
(Intercept)        3.5288     0.9041  3.903 9.50e-05
as.factor(stage)2 -0.1477     0.4076 -0.362 7.17e-01
```

```
as.factor(stage)3 -0.5866     0.3199 -1.833 6.68e-02
as.factor(stage)4 -1.5441     0.3633 -4.251 2.13e-05
age               -0.0175     0.0128 -1.367 1.72e-01
Log(scale)        -0.1223     0.1225 -0.999 3.18e-01

Scale= 0.885

Weibull distribution
Loglik(model)= -141.4   Loglik(intercept only)= -151.1
Chisq= 19.37 on 4 degrees of freedom, p= 0.00066
Number of Newton-Raphson Iterations: 5
n= 90
```

(b) The acceleration factor is

$$AF = \exp\left(\hat{\beta}_1 \times 0 + \hat{\beta}_2 \times 0 + \hat{\beta}_3 \times 0 + \hat{\beta}_4 \times 50\right) = \exp(-0.0175 \times 50) = 0.418$$

(c) To answer this question, we estimate the parametric models given by the function `survreg` and we select the one that gives the smallest AIC value.

```
> logn = survreg(Surv(time,delta) ~ as.factor(stage) + age, data=larynx, dist="lognormal")
> weib = survreg(Surv(time,delta) ~ as.factor(stage) + age, data=larynx, dist="weibull")
> expon = survreg(Surv(time,delta) ~ as.factor(stage) + age, data=larynx, dist="exponential")
> loglogist = survreg(Surv(time,delta) ~ as.factor(stage) + age, data=larynx, dist="loglogistic")
> AIC = c(extractAIC(logn)[2], extractAIC(weib)[2], extractAIC(expon)[2], extractAIC(loglogist)[2])
> names(AIC) = c( "log(normal)", "weibull", "exponential", "log(logistic)")
> AIC
  log(normal)       weibull   exponential log(logistic)
     294.7833      294.8468      293.7955      295.1780
```

Therefore, the optimal model in the sense of the AIC criterion is the exponential model.