University of Edinburgh, School of Mathematics

Biostatistics (MATH11230), 2021/2022

Vanda Inácio

Here we reproduce the results for the larynx data.

```
require(KMsurv) #for the larynx data
require(survival)
data(larynx)
res_1 <- coxph(Surv(time, delta) ~ as.factor(stage) + age, data = larynx)
res_1
```

```
## Call:
## coxph(formula = Surv(time, delta) ~ as.factor(stage) + age, data = larynx)
##
##                     coef exp(coef) se(coef)     z        p
## as.factor(stage)2 0.14004   1.15032  0.46249 0.303   0.7620
## as.factor(stage)3 0.64238   1.90100  0.35611 1.804   0.0712
## as.factor(stage)4 1.70598   5.50678  0.42191 4.043 5.27e-05
## age               0.01903   1.01921  0.01426 1.335   0.1820
##
## Likelihood ratio test=18.31  on 4 df, p=0.001072
## n= 90, number of events= 50
```

In order to apply the likelihood ratio test, let us fit a proportional hazards model that does not include the stage of the cancer as risk factor. Our null hypothesis is then: $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$.

```
res_2 <- coxph(Surv(time, delta) ~ age, data = larynx)
res_2
```

```
## Call:
## coxph(formula = Surv(time, delta) ~ age, data = larynx)
##
##        coef exp(coef) se(coef)     z     p
## age 0.02328   1.02356  0.01449 1.607 0.108
##
## Likelihood ratio test=2.63  on 1 df, p=0.1048
## n= 90, number of events= 50
```

```
anova(res_1,res_2)
```

```
## Analysis of Deviance Table
##  Cox model: response is  Surv(time, delta)
##  Model 1: ~ as.factor(stage) + age
##  Model 2: ~ age
##    loglik  Chisq Df P(>|Chi|)
## 1 -187.71
## 2 -195.55 15.681  3  0.001318 **
```

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

We can also obtain this value manually as follows.

```
#the first is the log likelihood of a model that contains none of the risk factors,
#so we need the second one
res_1$loglik
```

```
## [1] -196.8635 -187.7074
```

```
test_stat <- 2*(res_1$loglik[2] - res_2$loglik[2])
pchisq(test_stat, df = 4 - 1, lower.tail = FALSE)
```

```
## [1] 0.001318283
```

If we want to compare non nested models, we can use the AIC or the BIC.

```
AIC(res_1)
```

```
## [1] 383.4147
```

```
-2*res_1$loglik[2] + 2*4
```

```
## [1] 383.4147
```

```
BIC(res_1)
```

```
## [1] 391.0628
```

```
-2*res_1$loglik[2] + log(sum(larynx$delta))*4
```

```
## [1] 391.0628
```