Notes on Proportional Hazards Models using R

Chris Sutton, October 2023

In the R package 'survival', the command coxph() fits a Cox proportional hazards model to the supplied data:

```
library(survival)
coxph(<formula>, ties="breslow")
```

The argument formula needs to be a formula object with a survival object response. For example:

```
coxph(Surv(data$time,data$status)~data$smoker,
ties="breslow")
```

where the object data includes variables recording the time of each event, whether it was a censored event or not, and the covariate value of interest.

Note that the argument ties is a character string specifying the method for handling ties. If there are no tied death times all the options for this argument are equivalent. Nearly all Cox regression programs in R use the Breslow method by default, but not this one. The Breslow method is consistent with the theory presented above.

Example

Data is collected on 100 patients who have been admitted to hospital having been bitten by a poisonous snake. The time in hours before they are discharged from hospital is measured allowing for right censoring (those patients who leave hospital before they recover). Two covariates are also measured – the distance to the nearest mile that the patient had to travel to reach hospital and the number of times previously the patient had been bitten by a snake. A Cox's Proportional Hazard model is fitted to the data to estimate "survival" time in hours before hospital discharge.

The data set is found in a .csv file called CoxPHdataset.csv

The R code needed to fit the Cox's Proportional Hazard model is:

```
> library(survival)
> CoxPHdataset <- read.csv("CoxPHdataset.csv")
```

This reads the data set from the .csv file in the R workspace and assigns it to a data frame called CoxPHdataset. We can then use the head() command to view the first 6 entries.

```
> head(CoxPHdataset)
    time censored distance previous
1 5.021        0        0        0
2 2.157        0        5        0
3 5.685        0        4        0
4 1.860        0        2        0
5 0.151        0        1        0
6 3.028        0        5        0
```

First let us consider the simple model that just uses one of the covariates – distance. We will call this model CoxD.

```
> CoxD <-
coxph(Surv(CoxPHdataset$time,CoxPHdataset$censored)~C
oxPHdataset$distance, ties="breslow")
> CoxD
Call:
coxph(formula = Surv(CoxPHdataset$time,
CoxPHdataset$censored) ~
    CoxPHdataset$distance, ties = "breslow")


                      coef exp(coef) se(coef)
CoxPHdataset$distance 0.2006    1.2221   0.2184
                         z     p
CoxPHdataset$distance 0.918 0.358


Likelihood ratio test=0.9  on 1 df, p=0.343
n= 100, number of events= 7
```

From the R output we can see the beta coefficient in the one covariate Cox model (relating to distance) is 0.2006 and exp(beta) is 1.221. The R output also

gives a standard error for the coefficient and p and z statistics for the model fitted.

We can do similar for the previous covariate and have another one coefficient model CoxP

```
> CoxP <-
coxph(Surv(CoxPHdataset$time,CoxPHdataset$censored)~C
oxPHdataset$previous, ties="breslow")
> CoxP
Call:
coxph(formula = Surv(CoxPHdataset$time,
CoxPHdataset$censored) ~
    CoxPHdataset$previous, ties = "breslow")


                        coef exp(coef) se(coef)
CoxPHdataset$previous -0.2188    0.8035   0.8565
                          z     p
CoxPHdataset$previous -0.255 0.798


Likelihood ratio test=0.07  on 1 df, p=0.7882
n= 100, number of events= 7
```

This time the coefficient beta is negative -0.2188 so the hazard (of being discharged from hospital) reduces as the number of previous snake bites increases.

Finally we can find the Cox model with both covariates and two coefficients. We do this by using `~ CoxPHdataset$distance + CoxPHdataset$previous` *as the formula object in coxph().*

We will call this model Cox2

```
> Cox2 <-
coxph(Surv(CoxPHdataset$time,CoxPHdataset$censored)~C
oxPHdataset$distance + CoxPHdataset$previous,
ties="breslow")
> Cox2
```

```
Call:
coxph(formula = Surv(CoxPHdataset$time,
CoxPHdataset$censored) ~
    CoxPHdataset$distance + CoxPHdataset$previous,
ties = "breslow")


                         coef exp(coef) se(coef)
CoxPHdataset$distance  0.1972    1.2180   0.2196
CoxPHdataset$previous -0.1469    0.8634   0.8659
                         z     p
CoxPHdataset$distance  0.898 0.369
CoxPHdataset$previous -0.170 0.865


Likelihood ratio test=0.93  on 2 df, p=0.6282
n= 100, number of events= 7
```