Extensions to the Cox Model: Stratification

David M. Rocke

May 30, 2017

Anderson Data

Remission survival times on 42 leukemia patients, half on new treatment, half on standard treatment. This is the same data as the drug6mp data from KMsurv, but with two other variables and without the pairing.

Name	Description
treat	"standard", "new"
sex	"female", "male"
lwbc	log of white blood count
time	time to relapse or censoring
status	0 = censored, 1 = relapsed

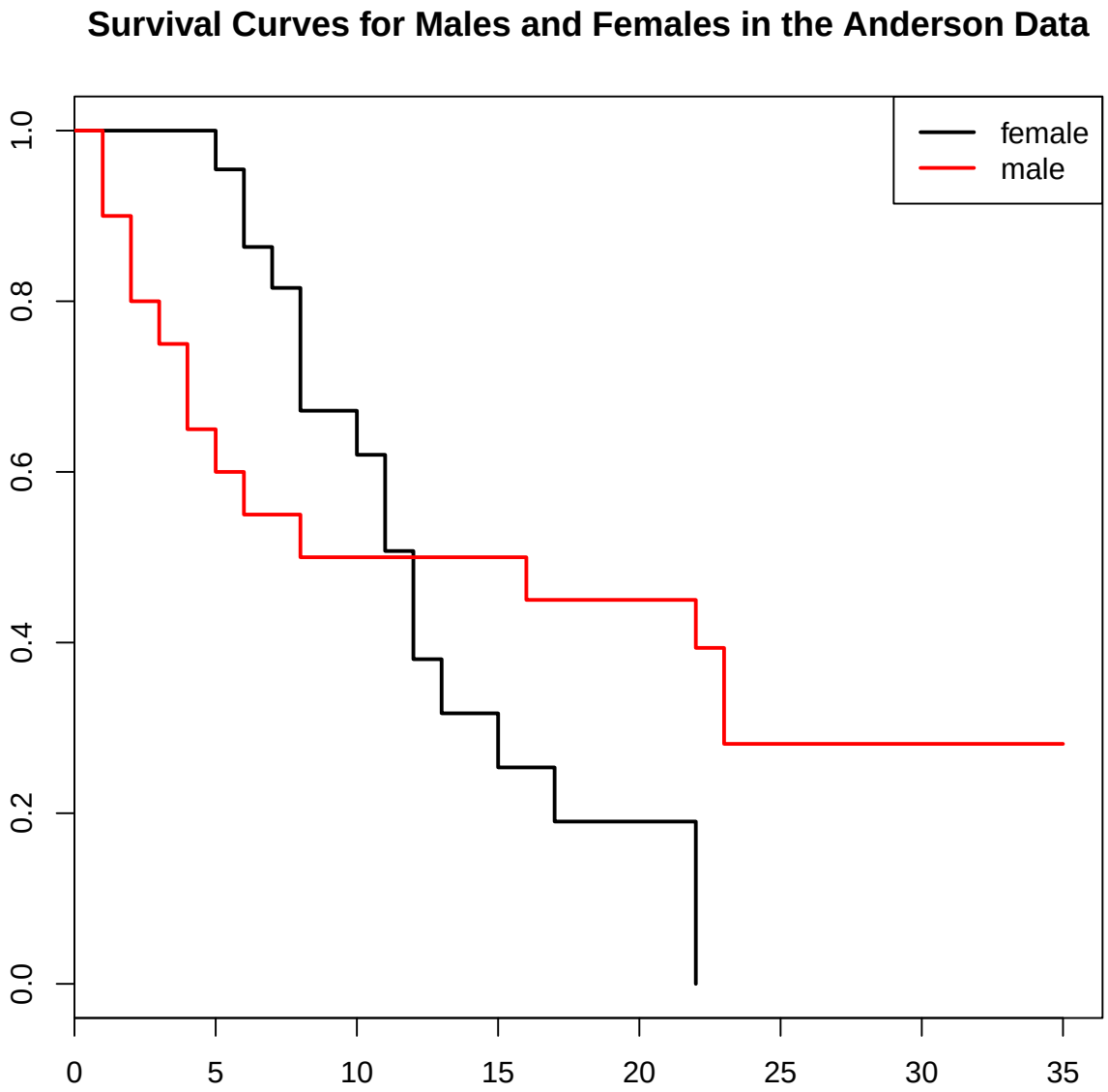
```
require(survival)
vars <- c("time","status","sex","lwbc","treat")
anderson <- read.table("anderson.dat",header=F,col.names=vars)
anderson$treat <- factor(anderson$treat,labels=c("new","standard"))
anderson$sex <- factor(anderson$sex,labels=c("female","male"))

anderson.surv <- with(anderson,Surv(time,status))
anderson.cox1 <- coxph(anderson.surv~treat+sex+lwbc,data=anderson)

> anderson.cox1

               coef exp(coef) se(coef)    z       p
treatstandard 1.504     4.498    0.462 3.26  0.0011
sexmale       0.315     1.370    0.455 0.69  0.4887
lwbc          1.682     5.376    0.337 5.00 5.8e-07

Likelihood ratio test=47.2  on 3 df, p=3.17e-10
n= 42, number of events= 30
> cox.zph(anderson.cox1)
                 rho chisq      p
treatstandard -0.1017 0.344 0.5578
sexmale       -0.3684 4.076 0.0435
lwbc           0.0595 0.161 0.6883
GLOBAL             NA 4.232 0.2374
```



The survival curves cross, which indicates a problem in the proportionality assumption by sex. This can be fixed by using strata or possibly by other model alterations.

The Stratified Cox Model

- In a stratified Cox model, each stratum, defined by one or more factors, has its own base survival function $h_0(t)$.
- But the coefficients for each variable not used in the strata definitions are assumed to be the same across strata.
- To check if this assumption is reasonable one can include interactions with strata and see if they are significant (this may generate a warning and NA lines but these can be ignored).
- Since the sex variable shows possible non-proportionality, we try stratifying on sex.

Stratified Model

```
> summary(coxph(anderson.surv~treat+lwbc+strata(sex),data=anderson))

  n= 42, number of events= 30

                coef exp(coef) se(coef)     z Pr(>|z|)
treatstandard 0.9981    2.7131   0.4736 2.108   0.0351 *
lwbc          1.4537    4.2787   0.3441 4.225 2.39e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

              exp(coef) exp(-coef) lower .95 upper .95
treatstandard     2.713     0.3686     1.072     6.864
lwbc              4.279     0.2337     2.180     8.398

Concordance= 0.812  (se = 0.093 )
Rsquare= 0.534   (max possible= 0.967 )
Likelihood ratio test= 32.06  on 2 df,   p=1.092e-07
Wald test            = 22.75  on 2 df,   p=1.15e-05
Score (logrank) test = 30.8  on 2 df,   p=2.052e-07
```

Separate Models

```
> summary(coxph(anderson.surv~treat+lwbc,data=anderson,sub=(sex=="male")))

  n= 20, number of events= 14

               coef exp(coef) se(coef)     z Pr(>|z|)
treatstandard 1.9779    7.2275   0.7392 2.676  0.00746 **
lwbc          1.7428    5.7132   0.5358 3.253  0.00114 **

> summary(coxph(anderson.surv~treat+lwbc,data=anderson,sub=(sex=="female")))

  n= 22, number of events= 16

               coef exp(coef) se(coef)     z Pr(>|z|)
treatstandard 0.3113    1.3652   0.5636 0.552   0.5807
lwbc          1.2061    3.3406   0.5035 2.396   0.0166 *
```

The coefficients of treatment look different. Are they statistically different?

Interaction Model

```
> summary(coxph(anderson.surv~(treat+lwbc)*strata(sex),data=anderson))
  n= 42, number of events= 30

                                coef exp(coef) se(coef)     z Pr(>|z|)
treatstandard                 0.3113    1.3652   0.5636 0.552   0.5807
lwbc                          1.2061    3.3406   0.5035 2.396   0.0166 *
treatstandard:strata(sex)male 1.6666    5.2942   0.9295 1.793   0.0730 .
lwbc:strata(sex)male          0.5366    1.7102   0.7352 0.730   0.4655
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

> anova(coxph(anderson.surv~treat+lwbc+strata(sex),data=anderson),
        coxph(anderson.surv~(treat+lwbc)*strata(sex),data=anderson),test="Chisq")
Analysis of Deviance Table
 Cox model: response is  anderson.surv
 Model 1: ~ treat + lwbc + strata(sex)
 Model 2: ~ (treat + lwbc) * strata(sex)
   loglik  Chisq Df P(>|Chi|)
1 -55.735
2 -53.852 3.7659  2    0.1521
```

Stratified Model for Anderson Data

- We chose to use a stratified model because of the apparent non-proportionality of the hazard for the sex variable.
- When we fit interactions with the strata variable, we did not get an improved model (via the likelihood ratio test).
- So we use the stratifed model with coefficients that are the same across strata.

Another Modeling Approach

- We used an additive model without interactions and saw that we might need to stratify by sex.
- Instead, we could try to improve the model—maybe the interaction of treatment and sex is real, and after fitting that we might not need separate hazard functions.
- Either approach may work.

```
> coxph(anderson.surv~treat+lwbc+sex+lwbc:sex+treat:sex,data=anderson)

                          coef exp(coef) se(coef)     z     p
treatstandard          0.37481   1.45471  0.55452  0.68 0.499
lwbc                   1.06370   2.89707  0.47261  2.25 0.024
sexmale               -4.98338   0.00685  2.11360 -2.36 0.018
lwbc:sexmale           1.23031   3.42230  0.63008  1.95 0.051
treatstandard:sexmale  2.17816   8.83008  0.91095  2.39 0.017

Likelihood ratio test=57  on 5 df, p=5.18e-11
n= 42, number of events= 30

> cox.zph(coxph(anderson.surv~treat+lwbc+sex+lwbc:sex+treat:sex,data=anderson))
                           rho    chisq     p
treatstandard          0.01970 0.010324 0.919
lwbc                  -0.00317 0.000459 0.983
sexmale               -0.19401 1.500183 0.221
lwbc:sexmale           0.19097 1.595360 0.207
treatstandard:sexmale -0.10997 0.307155 0.579
GLOBAL                     NA 3.901703 0.564
```

Homework

- Fit the model for the addicts data using clinic, prison, and methadone. Then perform all the model checking procedures.
- There is a problem with proportionality of hazards for the clinics. Fit the stratified model using clinics to define the strata.
- Is there an interaction with the strata?

Homework

- One of the model checks of the original model is to look for possible transformations of the methadone variable. What is your conclusion?
- If you considered a reformulation of the methadone variable, does this affect the decision to stratify.
- Interpret your final model and repeat the model checks. Plot the clinic hazards adjusted for methadone and prison record.