Cox Regression

Mona Maher

September 21, 2019

```
library(survival)
```

```
## Warning: package 'survival' was built under R version 3.5.3
```

```
attach(lung)
View(lung)
obj1<-with(lung, Surv(time, status))
model1<-coxph(obj1~sex, data = lung)
summary(model1)
```

```
## Call:
## coxph(formula = obj1 ~ sex, data = lung)
##
##   n= 228, number of events= 165
##
##        coef exp(coef) se(coef)      z Pr(>|z|)
## sex -0.5310    0.5880   0.1672 -3.176  0.00149 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##     exp(coef) exp(-coef) lower .95 upper .95
## sex     0.588      1.701    0.4237     0.816
##
## Concordance= 0.579  (se = 0.021 )
## Likelihood ratio test= 10.63  on 1 df,   p=0.001
## Wald test            = 10.09  on 1 df,   p=0.001
## Score (logrank) test = 10.33  on 1 df,   p=0.001
```

- there is a significant difference between male and female on hazard of death
- as female is a referance , being a female reduce the hazard of death by 42%

```
model2<-coxph(obj1~sex+ age+ ph.ecog+ ph.karno+pat.karno, data = lung)
summary(model2)
```

```
## Call:
## coxph(formula = obj1 ~ sex + age + ph.ecog + ph.karno + pat.karno,
##     data = lung)
##
##   n= 223, number of events= 160
##    (5 observations deleted due to missingness)
##
##                coef exp(coef)  se(coef)      z Pr(>|z|)
## sex       -0.561464  0.570373  0.170689 -3.289  0.00100 **
## age        0.011383  1.011448  0.009510  1.197  0.23134
## ph.ecog    0.565533  1.760386  0.186716  3.029  0.00245 **
## ph.karno   0.015853  1.015979  0.009853  1.609  0.10762
## pat.karno -0.010111  0.989940  0.006881 -1.470  0.14169
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
##           exp(coef) exp(-coef) lower .95 upper .95
## sex          0.5704     1.7532    0.4082     0.797
## age          1.0114     0.9887    0.9928     1.030
## ph.ecog      1.7604     0.5681    1.2209     2.538
## ph.karno     1.0160     0.9843    0.9965     1.036
## pat.karno    0.9899     1.0102    0.9767     1.003
##
## Concordance= 0.647  (se = 0.025 )
## Likelihood ratio test= 32.9  on 5 df,   p=4e-06
## Wald test            = 33  on 5 df,   p=4e-06
## Score (logrank) test = 33.79  on 5 df,   p=3e-06
```

- sex, pat.karno are lower the risk of hazard of death but only sex is segnificantly lower the hazard
- ph.ecog is significantly higher the risk of death
- the global p-value from likelihood ratio test is less than 0.05 so the model is significantly differ from the exponential distribution of the hazard

```
library(survminer)
```

```
## Warning: package 'survminer' was built under R version 3.5.3
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

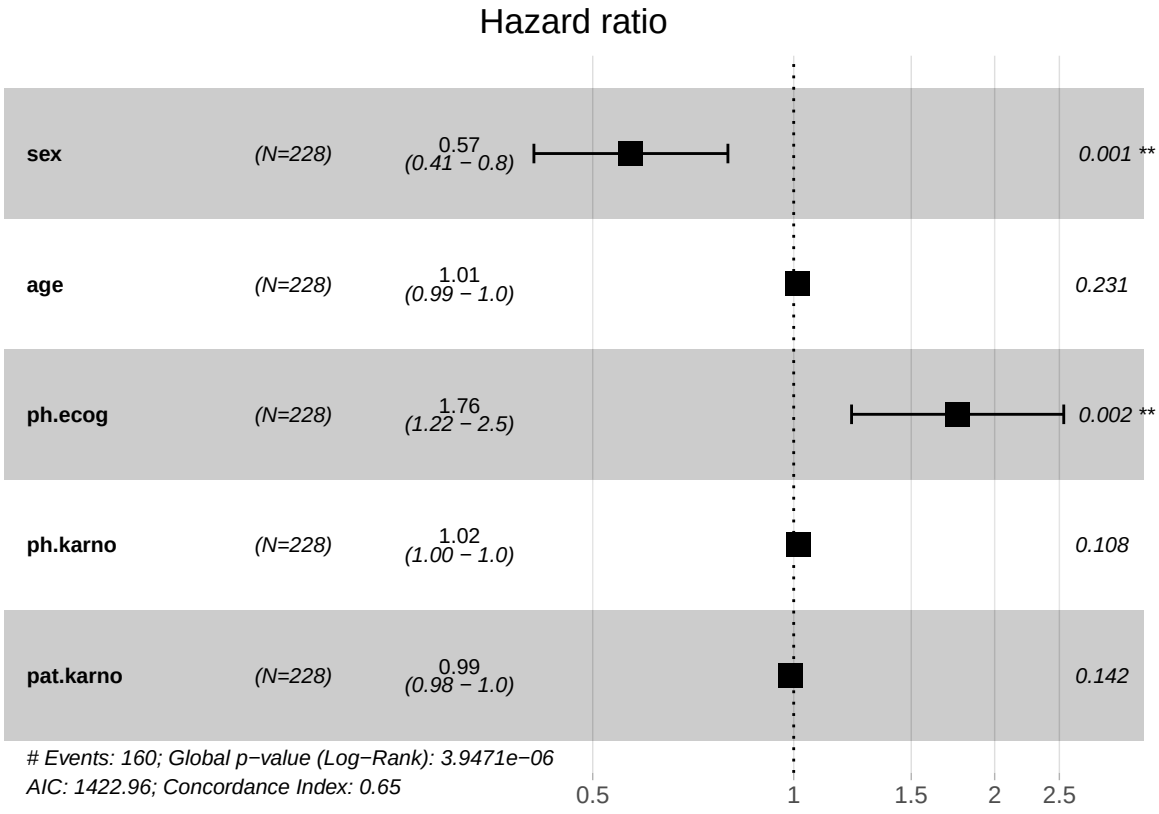
```
## Loading required package: ggpubr
```

```
## Warning: package 'ggpubr' was built under R version 3.5.3
```

```
## Loading required package: magrittr
```

```
## Warning: package 'magrittr' was built under R version 3.5.3
```

```
ggforest(model2, data = lung)
```




```
zph1<-cox.zph(model2)
zph1
```

```
##              rho chisq      p
## sex       0.0890 1.265 0.2607
## age       0.0388 0.260 0.6098
## ph.ecog   0.0690 0.746 0.3876
## ph.karno  0.1606 3.212 0.0731
## pat.karno 0.0874 1.373 0.2413
## GLOBAL        NA 9.274 0.0986
```

- the residuals are weak correlation with time, they are time independent
- p-value of residuals is greater than 0.05 So there is no linear relationship between resiual and time

Extended Cox model violation residual assumption of cox model

```
model3<-coxph(obj1~ ph.karno, data = lung)

summary(model3)
```

```
## Call:
## coxph(formula = obj1 ~ ph.karno, data = lung)
##
##   n= 227, number of events= 164
##    (1 observation deleted due to missingness)
##
##                coef exp(coef)  se(coef)     z Pr(>|z|)
## ph.karno -0.016448  0.983686  0.005854 -2.81  0.00496 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## ph.karno    0.9837      1.017    0.9725     0.995
##
## Concordance= 0.598  (se = 0.026 )
## Likelihood ratio test= 7.56  on 1 df,   p=0.006
## Wald test            = 7.89  on 1 df,   p=0.005
## Score (logrank) test = 7.95  on 1 df,   p=0.005
```

```
zph<-cox.zph(model3)
zph
```

```
##            rho chisq      p
## ph.karno 0.232  7.95 0.0048
```

- ph.karno is time dependent as p-value is less than 0.05
- Try to use interaction

```
model4<-coxph(obj1~ph.karno+ph.karno:time, data = lung)
summary(model4)
```

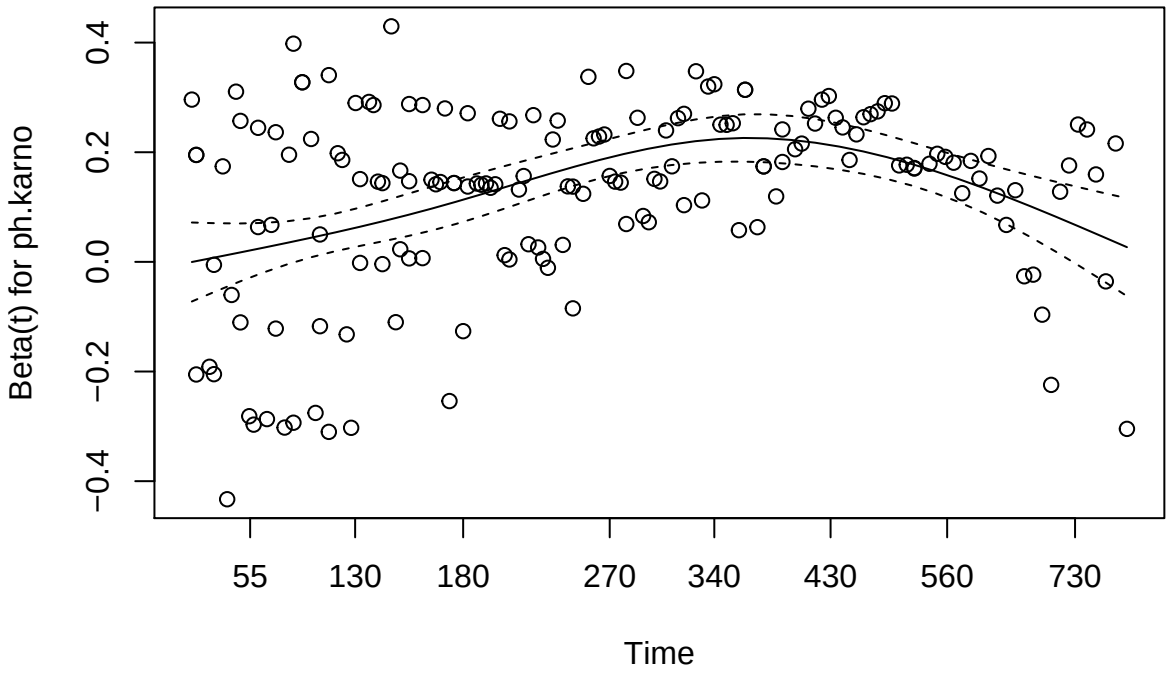
```
## Call:
## coxph(formula = obj1 ~ ph.karno + ph.karno:time, data = lung)
##
##   n= 227, number of events= 164
##    (1 observation deleted due to missingness)
##
##                      coef  exp(coef)   se(coef)      z Pr(>|z|)
## ph.karno       1.250e-01  1.133e+00  1.117e-02  11.19   <2e-16 ***
## ph.karno:time -4.174e-04  9.996e-01  2.986e-05 -13.98   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## ph.karno         1.1332     0.8825    1.1086    1.1582
## ph.karno:time    0.9996     1.0004    0.9995    0.9996
##
## Concordance= 0.965  (se = 0.004 )
## Likelihood ratio test= 547.8  on 2 df,   p=<2e-16
## Wald test            = 198.3  on 2 df,   p=<2e-16
## Score (logrank) test = 211.4  on 2 df,   p=<2e-16
```

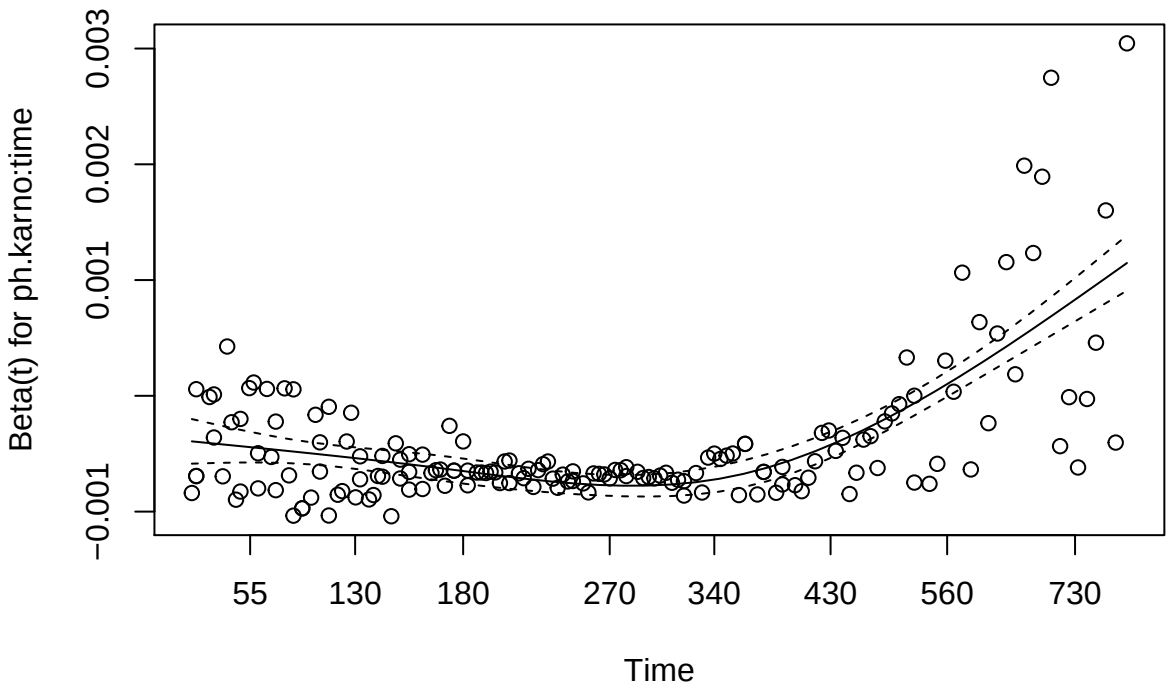
```
zph2<-cox.zph(model4)
zph2
```

```
##                 rho chisq        p
## ph.karno      0.242  13.7 2.18e-04
## ph.karno:time 0.432  81.2 2.07e-19
## GLOBAL             NA 291.9 4.22e-64
```

- residuals are still time dependent
- so try Step function

```
plot(zph2)
```


Beta(t) for ph.karno
0.4
0.2
0.0
−0.2
−0.4
55
130
180
270
340
430
560
730
Time




```
lung5<-survSplit(obj1~ph.karno, data = lung, cut = c(130,430), episode = "tgroup", id = "id")
model5<-coxph(obj1~ph.karno:strata(tgroup), data = lung5)
summary(model5)
```

```
## Call:
## coxph(formula = obj1 ~ ph.karno:strata(tgroup), data = lung5)
##
##   n= 467, number of events= 164
##    (1 observation deleted due to missingness)
##
##                                       coef  exp(coef)   se(coef)      z
## ph.karno:strata(tgroup)tgroup=1 -0.0381122  0.9626049  0.0121668 -3.132
## ph.karno:strata(tgroup)tgroup=2 -0.0157080  0.9844147  0.0083367 -1.884
## ph.karno:strata(tgroup)tgroup=3 -0.0003356  0.9996644  0.0117037 -0.029
##                                 Pr(>|z|)
## ph.karno:strata(tgroup)tgroup=1  0.00173 **
## ph.karno:strata(tgroup)tgroup=2  0.05954 .
## ph.karno:strata(tgroup)tgroup=3  0.97712
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                                 exp(coef) exp(-coef) lower .95 upper .95
## ph.karno:strata(tgroup)tgroup=1    0.9626      1.039    0.9399    0.9858
## ph.karno:strata(tgroup)tgroup=2    0.9844      1.016    0.9685    1.0006
## ph.karno:strata(tgroup)tgroup=3    0.9997      1.000    0.9770    1.0229
##
```

```
## Concordance= 0.598  (se = 0.025 )
## Likelihood ratio test= 12.66  on 3 df,   p=0.005
## Wald test            = 13.36  on 3 df,   p=0.004
## Score (logrank) test = 13.67  on 3 df,   p=0.003
```

```
zph3<-cox.zph(model5)
zph3
```

```
##                                     rho  chisq     p
## ph.karno:strata(tgroup)tgroup=1 -0.0301 0.1819 0.670
## ph.karno:strata(tgroup)tgroup=2  0.0985 1.2282 0.268
## ph.karno:strata(tgroup)tgroup=3  0.0185 0.0445 0.833
## GLOBAL                             NA 1.4546 0.693
```

- residual in groups ph.karno are time independent