Survival Analysis2

2023-11-12

Consider the Mayo Clinic Lung Cancer Data in R package survival : data(lung) or data(cancer): including the variables inst: Institution code time: Survival time in days status: censoring status 1=censored, 2=dead age: Age in years sex: Male=1 Female=2 ,etc

```
#import dataset
library(survival)
data(cancer)
```

Q1

Using a Cox proportional hazards model, estimate the hazard rate for Female relative to Male, without including “age” or other variables in the model

```
fitcox <- coxph(Surv(time,status)~sex,data=cancer)
summary(fitcox)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ sex, data = cancer)
##
##   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
```

```
exp(-0.5310*2)
```

```
## [1] 0.3457636
```

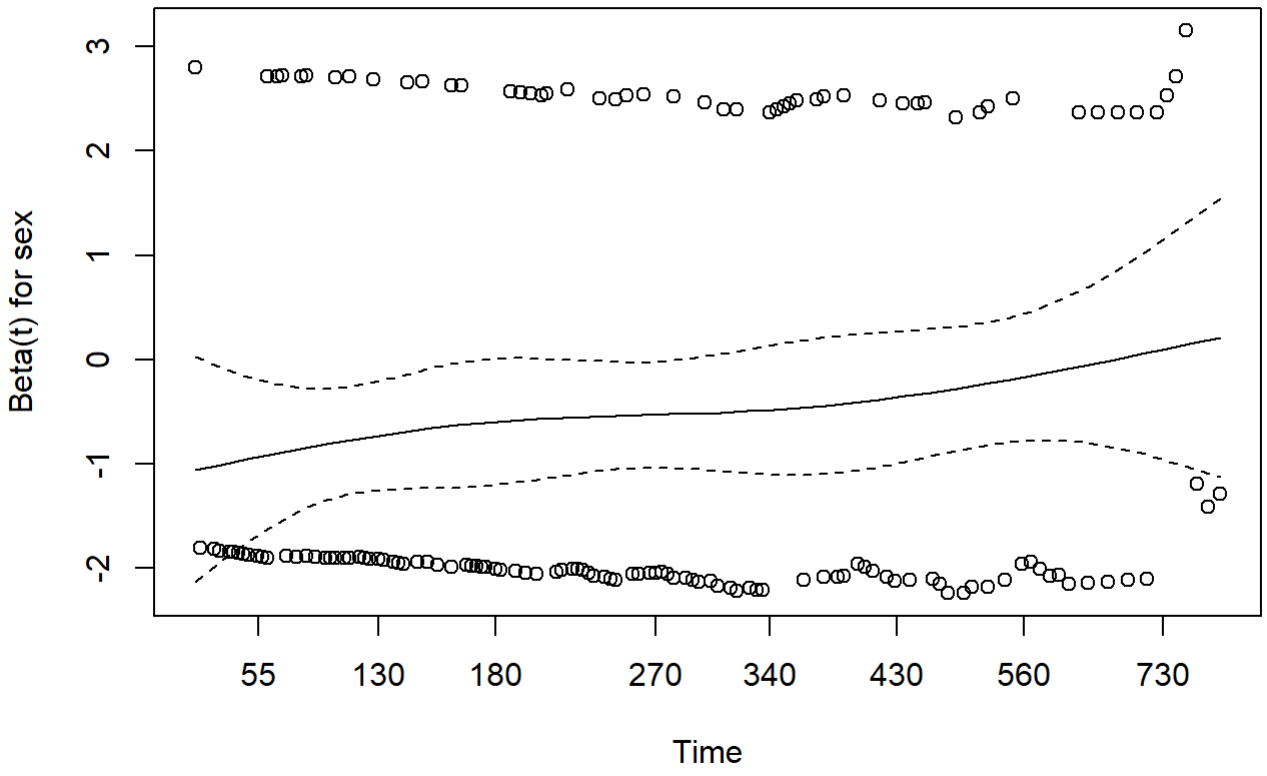
Answer:

The estimated coefficient beta is -0.5310. The hazard is exp(-0.5310)=0.5880. The hazard of experiencing the death of individual with sex=2 is 0.588 times the hazard of sex=1, suggesting a longer survival time of sex=2/female compared to sex=1/male. P-value as 0.00149 shows the effect of sex on the hazard is statistically significant.

Q2

Assess the validity of the proportional hazards assumption in (1)

```
# test proportional hazard assumption
prop <- cox.zph(fitcox)
plot(prop)
```




```
prop$table
```

```
##           chisq df          p
## sex    2.863235  1 0.09062506
## GLOBAL 2.863235  1 0.09062506
```

Answer:

Null Hypothesis: The slope of the beta is equal to zero, implying that the hazard (or log hazard) does not change with changes in the predictor variable. This indicates proportionality and is consistent with the proportional hazards assumption.

We plot the estimate of time-dependent coefficient beta. As p-value 0.09062506, the slope of beta is equal to zero. Thus, the assumption is satisfied.

Q3

Repeat 1, adjusting for “age”.

```
fitcox <- coxph(Surv(time, status) ~ sex + age, data = cancer)
summary(fitcox)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ sex + age, data = cancer)
##
##   n= 228, number of events= 165
##
##          coef exp(coef)  se(coef)      z Pr(>|z|)
## sex -0.513219  0.598566  0.167458 -3.065  0.00218 **
## age  0.017045  1.017191  0.009223  1.848  0.06459 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##     exp(coef) exp(-coef) lower .95 upper .95
## sex    0.5986     1.6707    0.4311    0.8311
## age    1.0172     0.9831    0.9990    1.0357
##
## Concordance= 0.603  (se = 0.025 )
## Likelihood ratio test= 14.12  on 2 df,   p=9e-04
## Wald test            = 13.47  on 2 df,   p=0.001
## Score (logrank) test = 13.72  on 2 df,   p=0.001
```

Answer:

The estimated coefficient beta is -0.513219 The hazard is exp(-0.513219)=0.598566. The hazard of experiencing the death of individual with sex=2 is 0.598566 times the hazard of sex=1, suggesting a longer survival time of sex=2/female compared to sex=1/male. P-value as 0.00218 shows the effect of sex on the hazard is statistically significant.

P-value as 0.06459 shows the effect of age on the hazard may not be statistically significant. The estimated coefficient beta is 0.017045 The hazard is exp(0.017045)=1.017191. For each one-unit increase in age, the hazard of death increases by approximately 1.7%. Overall, decreased survival time and increased risk/hazard are shown with increasing age.