GR5291 Advanced Data Analysis Problem Set Survival Analysis 2

Francis Zhang

December 6, 2024

Question

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.
```

1.Using a Cox proportional hazards model, estimate the hazrad rate for Female relative to Male, without including "age" or other variables in the model

2.Assess the validility of the proportional hazards assumption in (1)

3.Repeat 1, adjusting for "age"

Solution

Question 1

Cox Proportional Hazards Model Without Adjusting for Age

We will estimate the hazard rate for Females relative to Males, using only the sex variable in the model.

```
# Load required package and data
library(survival)
data(lung)
```

```
## Warning in data(lung): data set 'lung' not found
```

```
# Fit the Cox Proportional Hazards Model
cox_model_sex <- coxph(Surv(time, status) ~ sex, data = lung)

# Summarize the model
summary(cox_model_sex)
```

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

The Cox Proportional Hazards model shows that the hazard rate for females (sex = 2) is significantly lower than that for males (sex = 1), with a hazard ratio of 0.588 (95% CI: 0.4237–0.816, p = 0.00149). This indicates that females have approximately 41.2% lower risk of death compared to males. The model's fit is supported by significant results from the likelihood ratio test (p = 0.001), Wald test (p = 0.001), and log-rank test (p = 0.001), confirming the importance of sex in explaining survival differences.

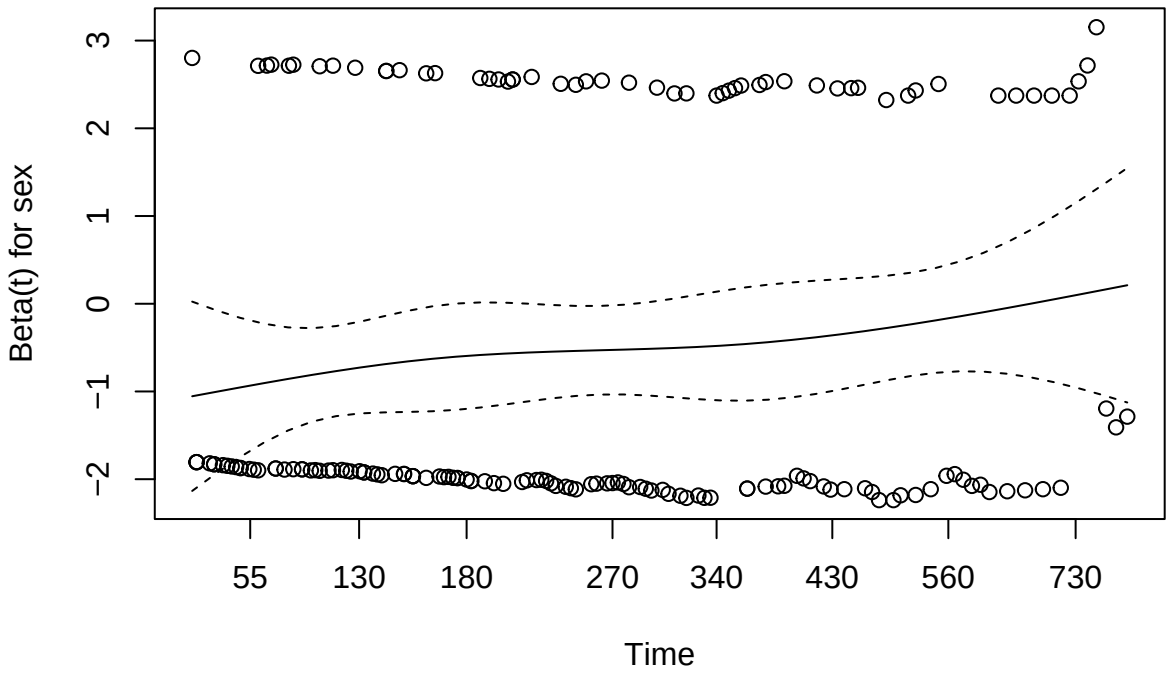
Question 2

Assess the Proportional Hazards Assumption

The proportional hazards assumption can be checked using scaled Schoenfeld residuals.

```r
# Test proportional hazards assumption
ph_assumption <- cox.zph(cox_model_sex)

# Plot Schoenfeld residuals for sex
plot(ph_assumption, var = "sex")
```



The Schoenfeld residuals plot for sex shows that the beta coefficient remains mostly constant over time, with the solid line staying within the confidence bands. However, slight deviations from horizontality, particularly at later times, suggest a potential time-varying effect. The proportional hazards assumption appears approximately valid but may require further investigation for late-time deviations.

Question 3

Cox Proportional Hazards Model Adjusting for Age

```
# Fit the Cox Proportional Hazards Model adjusting for age
cox_model_sex_age <- coxph(Surv(time, status) ~ sex + age, data = lung)

# Summarize the model
summary(cox_model_sex_age)
```

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

After adjusting for age, the hazard rate for females remains significantly lower compared to males, with a hazard ratio of 0.5986. The effect of age on survival is positive but not statistically significant ($p = 0.06459$). These results suggest that sex is an independent and significant predictor of survival, even after adjusting for age, with females showing a substantially lower risk of death compared to males.