Michael_Ghattas_WP9

Michael Ghattas

2024-11-10

Exercise:

1. Fit a Cox proportional hazard model to the data set with the predictors sex, and wt.loss (as main effects).

2. Interpret the regression coefficients for the model.

3. Predict the survival of a female who has wt.loss = 20.

4. Assess the model assumptions

```
attach(lung)
lung <- lung |>
  filter(!is.na(ph.ecog)) |>
  mutate(
    ph.ecog = factor(
      ph.ecog,
      levels = c(0, 1, 2, 3, 4),
      labels = c("Asymptomatic", "Ambulatory", "In bed <50%", "In bed >50%", "Bedridden")),
    ph.ecog = fct_drop(ph.ecog),
    ph.ecog = fct_lump_n(ph.ecog, n = 2, other_level = "Bedridden"),
    sex = factor(sex, levels = c(1, 2), labels = c("Male", "Female")),
    patient_id = row_number()
  )

head(lung)
```

```
##   inst time status age  sex      ph.ecog ph.karno pat.karno meal.cal wt.loss
## 1    3  306      2  74 Male   Ambulatory       90       100     1175      NA
## 2    3  455      2  68 Male Asymptomatic       90        90     1225      15
## 3    3 1010      1  56 Male Asymptomatic       90        90       NA      15
## 4    5  210      2  57 Male   Ambulatory       90        60     1150      11
## 5    1  883      2  60 Male Asymptomatic      100        90       NA       0
## 6   12 1022      1  74 Male   Ambulatory       50        80      513       0
##   patient_id
## 1          1
## 2          2
## 3          3
## 4          4
## 5          5
## 6          6
```

1. Fit a Cox proportional hazard model to the data set with the predictors sex, and wt.loss (as main effects).

```
cox_lung <- coxph(Surv(time, status) ~ sex + wt.loss, data = lung)
summary(cox_lung)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ sex + wt.loss, data = lung)
##
##   n= 213, number of events= 151
##    (14 observations deleted due to missingness)
##
##                 coef  exp(coef)   se(coef)      z Pr(>|z|)
## sexFemale -0.5336010  0.5864892  0.1743363 -3.061  0.00221 **
## wt.loss    0.0001296  1.0001296  0.0061051  0.021  0.98306
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## sexFemale    0.5865     1.7051    0.4167    0.8254
## wt.loss      1.0001     0.9999    0.9882    1.0122
##
## Concordance= 0.587  (se = 0.026 )
## Likelihood ratio test= 9.87  on 2 df,   p=0.007
## Wald test            = 9.38  on 2 df,   p=0.009
## Score (logrank) test = 9.6  on 2 df,   p=0.008
```

2. Interpret the regression coefficients for the model.

Interpretation:

- The negative coefficient indicates that being female is associated with a lower hazard (risk of death) compared to being male.
- The hazard ratio of 0.5865 means that females have approximately 41.35% lower hazard than males, holding wt.loss constant.
- Since the p-value is less than 0.01, this effect is statistically significant.
- The coefficient is close to zero, and the hazard ratio is approximately 1, suggesting no significant effect of wt.loss on the hazard.
- The p-value is very high, indicating that variations in wt.loss are not significantly associated with the hazard rate in this model.

3. Predict the survival of a female who has wt.loss = 20.

```
# Define the new data point
new_data <- data.frame(sex = "Female", wt.loss = 20)

# Obtain the survival curve
surv_pred <- survfit(cox_lung, newdata = new_data)

# Display the survival probabilities at specific times
summary(surv_pred, times = c(100, 200, 300, 400, 500))
```

```
## Call: survfit(formula = cox_lung, newdata = new_data)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   100    186      26    0.912  0.0202        0.873        0.952
##   200    140      35    0.783  0.0358        0.716        0.857
##   300     89      29    0.651  0.0482        0.563        0.753
##   400     57      22    0.521  0.0569        0.421        0.645
##   500     41      12    0.434  0.0603        0.331        0.570
```

Interpretation:

- Survival Probability at Time 100 Days: There's a 91.2% chance that the patient will survive beyond 100 days.
- At 200 Days: The survival probability decreases to 78.3%.
- At 300 Days: It further decreases to 65.1%.
- At 400 Days: The probability is 52.1%.
- At 500 Days: There's a 43.4% chance of surviving beyond 500 days.
- The standard error and 95% confidence intervals provide information about the precision of these estimates.

4. Assess the model assumptions

```r
# Test proportional hazards assumption
ph_test <- cox.zph(cox_lung)
print(ph_test)
```

```
##          chisq df    p
## sex     2.6849  1 0.10
## wt.loss 0.0305  1 0.86
## GLOBAL  2.7233  2 0.26
```

Interpretation:

- Since the p-value is greater than 0.05, we fail to reject the null hypothesis that the effect of sex is proportional over time. Thus, the PH assumption holds for sex.
- With a high p-value, there's no evidence against the PH assumption for wt.loss.
- The global test also suggests that the PH assumption holds for the model overall.
- The test results indicate that the PH assumption is not violated for either of the predictors or the model as a whole.

END.