Problem Set 4

Applied Stats II

Due: April 12, 2024

Instructions

- Please show your work! You may lose points by simply writing in the answer. If the problem requires you to execute commands in R, please include the code you used to get your answers. Please also include the .R file that contains your code. If you are not sure if work needs to be shown for a particular problem, please ask.

- Your homework should be submitted electronically on GitHub in .pdf form.

- This problem set is due before 23:59 on Friday April 12, 2024. No late assignments will be accepted.

Question 1

We're interested in modeling the historical causes of child mortality. We have data from 26855 children born in Skellefteå, Sweden from 1850 to 1884. Using the "child" dataset in the eha library, fit a Cox Proportional Hazard model using mother's age and infant's gender as covariates. Present and interpret the output.

```
1  # load libaries
2  library(eha)
3  library(survival)
4
5  data(child)
6
7  # Fit the Cox Proportional Hazards model
8  cox_model <- coxph(Surv(enter, exit, event) ~ m.age + sex, data = child)
9
10 plot_cox<-coxreg(Surv(enter, exit, event) ~ m.age + sex, data = child)
11
12 plot(plot_cox)
13
14 # Display the model summary
15 summary(cox_model)
```

```
  n= 26574, number of events= 5616

               coef exp(coef)  se(coef)       z Pr(>|z|)
m.age      0.007617  1.007646  0.002128  3.580 0.000344 ***
sexfemale -0.082215  0.921074  0.026743 -3.074 0.002110 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

          exp(coef) exp(-coef) lower .95 upper .95
m.age        1.0076     0.9924     1.003    1.0119
sexfemale    0.9211     1.0857     0.874    0.9706

Concordance= 0.519  (se = 0.004 )
Likelihood ratio test= 22.52  on 2 df,    p=1e-05
Wald test            = 22.52  on 2 df,    p=1e-05
Score (logrank) test = 22.53  on 2 df,    p=1e-05
```
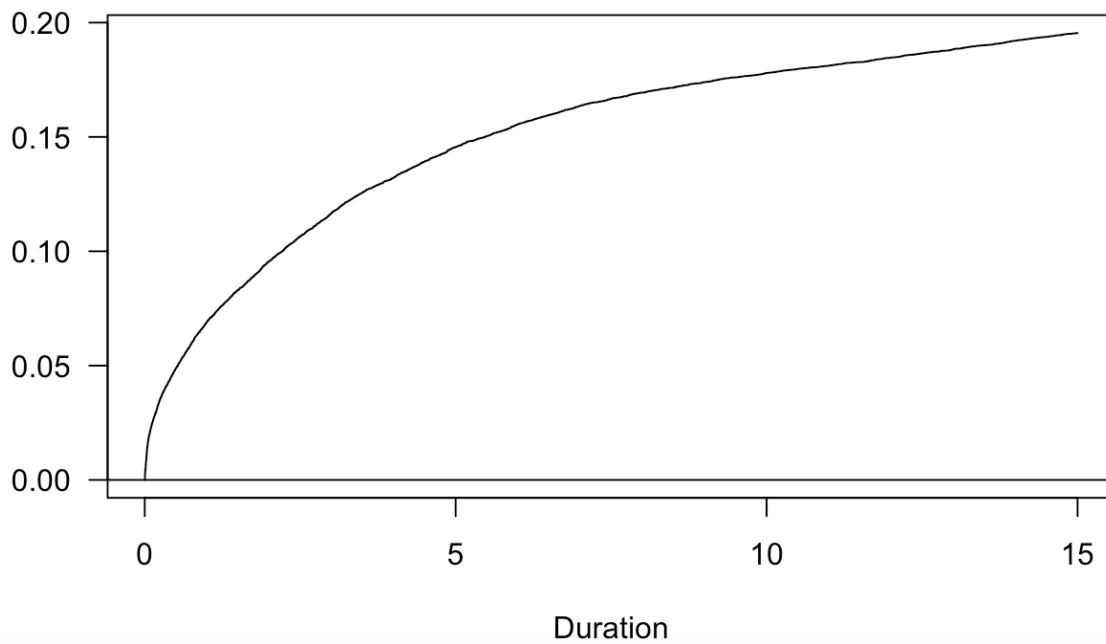

Figure 1: plotting cumulative hazard function

- The coefficient for the mother's age is 0.007617. A positive value indicates an increased hazard (risk) for the event of an infant's death when the covariate increases by one unit. The 95% confidence interval of the hazard ratio ranges from 1.003 to 1.0119. When holding other variables constant, with each additional year of the mother's age, the hazard of the event happening increases by a factor of 1.007646 (hazard ratio) and the effect is statistically significant. This hazard ratio suggests that for each additional year of the mother's age, the risk of the mortality event happening increases by 0.7646%.

- exp(-coef) for the mother's age is 0.9924, which means that for each one-year decrease in the mother's age, the risk of child mortality is multiplied by 0.9924, or decreases by approximately 0.76% (since 1 - 0.9924 = 0.0076 or 0.76%). It's the reciprocal effect of a one-unit increase in the mother's age.

- The coefficient for being female (sex) is -0.082215, the negative sign refers to being female is associated with a decrease in the hazard (or risk) of the event happening, compared to the man (baseline category). The 95% confidence interval of the hazard ratio is between 0.874 and 0.9706, the hazard ratio is 0.921074 indicating when holding other variables constant, female infants have a 7.8926% (1 - 0.921074) lower hazard of the mortality event occurring at any given time compared to male infants.

- exp(-coef) for the female is 1.0857, the opposite direction of the original coefficient: males have an 8.57% higher hazard of the event occurring compared to females, assuming all other factors are constant.

- The concordance index is 0.519, indicating the model's predictive ability is a bit better than by chance. (1 suggests perfect, 0.5 equal by random chance)