Problem Set 4

Yajie Dong

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.

Step 1: Get the Data, let's do that by R:

```
1 # Install the 'eha' package
2 install.packages("eha")
3 # Install the 'survival' package
4 install.packages("survival")
5 # Load the 'eha' library to access the 'child' dataset, which contains
      information on child mortality influenced by the mother's background and
      the child's sex
6 library(eha)
7 # Load the 'survival' package
8 library(survival)
9 # Load the 'child' dataset from the 'eha' library
10 data("child", package = "eha")
```

Step 2: Fitting and Summarizing the Cox Proportional Hazards Model, let's do that by R

```
1 # Apply the Cox Proportional Hazards model to assess survival, considering
    factors such as the mother's age and the child's sex
2 infantMortality <- coxph(Surv(enter, exit, event) ~ m.age + sex, data = child)
3 # To get a summary of the model's results:
4 summary(infantMortality)
```

Output from R:

```
Call:
coxph(formula = Surv(enter, exit, event) ~ m.age + sex, data = child)

  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
```

Note: The recent update to the 'eha' package, which provides the 'child' dataset, may have resulted in a revised number of observations, now totaling 26,574, likely due to data corrections or quality improvements.

From the result: Upon analyzing the data, it's evident that the coefficient related to the infant's gender can be interpreted as the logarithmic hazard ratio of mortality for male infants compared to female ones, with the condition that the mother's age remains the same. When we apply the exponential function to the gender coefficient (exp(coef) = 0.921074), it reveals that female infants have a roughly 7.89% lower mortality risk compared to male infants, assuming the mother's age is constant (a hazard ratio (HR) less than 1 indicates a decrease in hazard). This points to a survival advantage for female infants under these historical conditions.

When we look at the mother's age coefficient, it shows that an increase in the mother's age correlates with a slight rise in mortality risk for the child, assuming the infant's gender is constant. Specifically, for each year increase in the mother's age, there's a 0.76% increase in the child's mortality risk (indicated by a hazard ratio of 1.007646 for the mother's age). This implies that an older maternal age is a detrimental factor that increases the risk of mortality, suggesting a higher vulnerability associated with advanced maternal age.