Problem Set 4

Applied Stats II

Zhuo Zhang/23346227

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.

```r
1  # Load the eha library
2  library(eha)
3
4  # Load the child dataset
5  data(child)
6  child_surv <- with(child, Surv(enter, exit, event))
7
8  # Fit Cox Proportional Hazard model with mother's age and infant's gender as
      covariates
9  cox_model <- coxph(child_surv ~ m.age + sex, data = child)
10
11 # Display summary of the Cox Proportional Hazard model
12 summary(cox_model)
```

Result:
```
Call:
coxph(formula = child_surv ~ 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
```

The coefficient of m.age is 0.007617, which means that for every one year increase in maternal age, the risk of child death increases by approximately 0.76%.

The coefficient of sexfemale is -0.082215, which means that female infants have a lower risk of death compared to male infants, and the risk of death for female infants is about 7.89% lower than that for male infants.

In statistics, the coefficients of both maternal age and infant gender variables are significant (p.value <0.05), indicating that their impact on predicting child mortality is significant.

The Concordance statistic is 0.519, indicating the consistency of the model. The p-values of the log likelihood ratio test, Wald test, and score logrank test are all very small (p-value = 0.00001), indicating a good overall fit of the model.