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.

Answer 1

The Cox Proportional Hazard model is a widely used survival regression model that estimates the relationship between covariates and the hazard function in survival analysis. It assumes that the hazard function is a product of a baseline hazard function and an exponential function of the covariates. The baseline hazard function represents the hazard rate for an individual with all covariates set to zero, while the exponential function captures the effect of the covariates on the hazard .

The general form of the Cox Proportional Hazard model is given by:

h(t—X) = h(t) * exp(X)

where h(t—X) is the hazard function at time t for an individual with covariate values X, h(t) is the baseline hazard function, X is the vector of covariate values, and is the vector of

regression coefficients. The exponential term exp(X) represents the relative hazard function, which quantifies the effect of the covariates on the hazard.

*$10_{D}uration_{p}rintable_{l}ecturenotes$

```
lapply(c("survival", "eha", "tidyverse", "ggfortify", "stargazer"),  pkgTest)

data(child)

summary(child)

# Creating the survival object with the necessary variables, exit and event
surv_obj <- Surv(time = child$exit, event = child$event)

#Cox  model with mother's age and infant's gender
cox_model <- coxph(surv_obj ~ m.age + sex, data = child)

#output
summary(cox_model)

> summary(cox_model)
Call:
coxph(formula = surv_obj ~ 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 p-values indicate the statistical significance of the coefficients . The Concordance value is 0.519, indicating a moderate level of concordance between predicted and observed survival times. Overall, the Cox Proportional Hazard model suggests that mother's age and infant's gender are significant predictors of the hazard rate in the "child" dataset

Burak Ozuer - 18302382