Problem Set 4

Applied Stats II

Due: April 12, 2024//Wei Tang 23362496

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.

```
# load the data set
data(child)
View(child)
# use the Surv() to build the survival model
child_surv <- with(child, Surv(enter, exit, event))
# run Cox proportional hazard model
cox <- coxph(child_surv ~ sex + m.age, data = child)
# get the summary
summary(cox)
# perform a test to assess the quality of the model
drop1(cox, test = "Chisq")
# display the Cox model using stargazer
stargazer(cox, type = "text")
```

```
> stargazer(cox, type = "text")

================================================
                         Dependent variable:
                     ---------------------------
                              child_surv
------------------------------------------------
sexfemale                      -0.082***
                                (0.027)

m.age                          0.008***
                                (0.002)

------------------------------------------------
Observations                    26,574
R2                               0.001
Max. Possible R2                 0.986
Log Likelihood                -56,503.480
Wald Test                  22.520*** (df = 2)
LR Test                    22.518*** (df = 2)
Score (Logrank) Test       22.530*** (df = 2)
================================================
Note:                *p<0.1; **p<0.05; ***p<0.01
```

Interpretation of coefficients:
Holding m.age constant, relative to male infants, female infants have a decrease of 0.082 in the expected log hazard. Holding sex constant, for each one-year increase in maternal age, there is an increase of 0.34 in the expected log hazard.

And as we can see, the coefficients of two variables are both significant.

```r
# Exponentiated parameter estimates to obtain hazard ratios
exp(-0.082)
exp(0.008)
```

```
> exp(-0.082)
[1] 0.921272
> exp(0.008)
[1] 1.008032
```

Explanation: The hazard ratio for female infants compared to male infants is 0.92, indicating that female infants are 0.92 times as likely to die relative to male infants. The hazard ratio for female infants compared to male infants is 1.008, indicating that for each additional year of maternal age, the relative hazard increases by 1.008 times.