Problem Set 4

Applied Stats II

Due: April 16, 2023

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 Sunday April 16, 2023. 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.

Results:

```
    child$time <- as.numeric(child$time)

surv_obj <- Surv(child$time, child$event)

model <- coxph(surv_obj ~ m.age + sex, data = child)

model

Call:
coxph(formula = surv_obj ~ m.age + sex, data = child)

               coef  exp(coef)   se(coef)       z      p
```

```
m.age     -4.350e+01  1.282e-19  5.824e-01 -74.690 <2e-16
sexfemale -6.629e-02  9.359e-01  2.711e-02  -2.445 0.0145

Likelihood ratio test=66111  on 2 df, p=< 2.2e-16
n= 26574, number of events= 5616
```

```
mod1 <- stargazer(model, title = "Survival Analysis Results",
                  dep.var.caption = "Dependent variable",
                  covariate.labels = c("Age", "Sex"),
                  column.labels = c("Coefficient", "Standard Error"),
                  type = "text")
mod1 <- stargazer(model)
mod1
```

```
Survival Analysis Results
================================================
                         Dependent variable
                     ---------------------------
                              surv_obj
                            Coefficient
------------------------------------------------
Age                          -43.500***
                               (0.582)

Sex                           -0.066**
                               (0.027)

------------------------------------------------
Observations                    26,574
R2                               0.917
Max. Possible R2                 0.979
Log Likelihood                -18,280.960
Wald Test                5,584.510*** (df = 2)
LR Test                 66,110.830*** (df = 2)
Score (Logrank) Test    11,680.740*** (df = 2)
================================================
Note:                *p<0.1; **p<0.05; ***p<0.01
```

Interpert:
Both of the coefficients are significant predictors of survival.

Mothers age has a p-value of less than 0.01 and Sex of the child has a p-value less than 0.05.
The coeffiecient for sex is negative which indicates that being female is associated with a lower hazard of experiencing the event occuring.

The coefficient estimates indicate that older age is associated with a much lower hazard of experienceing the event over time with a coeffient of -43.50 and a p-value of less than 0.01.

The results show that age is a strong predictor of survival for the children represented in the dataset whereas sex has a smaller effect.