Problem Set 4

Applied Stats II

Due: April 4, 2022

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 class on Monday April 4, 2022. No late assignments will be accepted.
- Total available points for this homework is 80.

Question 1

We're interested in modeling the historical causes of infant mortality. We have data from 5641 first-born in seven Swedish parishes 1820-1895. Using the "infants" 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.

Firstly, install the libraries and load in the dataset (see Table 1; pg. 2). The cox proportional hazard model can be used from the survival package and the dataset here is the 'infants' data from the eha library. Then call the help file on this dataset to make assumptions about the outocome and predictor variables that will be used in this analysis. Importantly, the 'sex' predictor is a dummy categorical variable and the 'age' predictor is a numeric variable.

```
## Install Libraries
library(eha)
library(survival)
library(survminer)

# Load dataset
data <- eha::infants
```

Table 1:

stratum	enter	exit	event	mother	age	sex	parish	civst	ses	year
1	55	365	0	dead	26	boy	Nedertornea	married	farmer	1,877
1	55	365	0	alive	26	boy	Nedertornea	married	farmer	1,870
1	55	365	0	alive	26	boy	Nedertornea	married	farmer	1,882
2	13	76	1	dead	23	girl	Nedertornea	married	other	1,847

```

# Call help file on dataset and analyse outocme and predictors
?eha::infants
# Predictors: sex is a binary predictor with options boy/girl; Mothers age at
    infants birth a numerical
# predictor
## Outcome variable: enter, exit and event; charts whether an infant died (i.e
     the event) and over what time (enter and exit).
```

Now, create the survival rate variable using the enter, exit, event variables. Then, fit the cox proportional hazards model using the coxph() function. Summarize the output using the summary() function. Finally, interpret this output

```
#### COX PROPORTIONAL MODEL ###
### Fit Cox Proportional Hazard Model
inf_surv <- with(data, Surv(enter, exit, event))
cox_mod <- coxph(inf_surv ~ age + sex, data = data)
summary(cox_mod)
```

```
THE COX-PROPORTIONAL HAZARDS SUMMARY OUTPUT
Call:
coxph(formula = inf_surv ~ age + sex, data = data)

n= 105, number of events= 21

coef  exp(coef)  se(coef)      z  Pr(>|z|)
age    -0.04044   0.96037  0.04507 -0.897    0.370
sexboy -0.48518   0.61559  0.44224 -1.097    0.273

exp(coef) exp(-coef) lower .95  upper .95
age       0.9604      1.041     0.8792      1.049
sexboy    0.6156     1.624     0.2587      1.465

Concordance= 0.586  (se = 0.058 )
Likelihood ratio test= 1.99  on 2 df,   p=0.4
```

```
Wald test            = 2  on 2 df,   p=0.4
Score (logrank) test = 2.03  on 2 df,   p=0.4
```

INTERPRETING THE OUTPUT

The sample size of infants is 105. The no. of deaths is 21

Coefficients:

Coefficient for sexboy is -0.04. The expected log of the hazard decreases by -0.04 when an infant is a boy compared to a girl, holding age constant.

Coefficient for age is -0.49. The expected log of the hazard decreases by -0.49 if a mother is one year older, holding the sex of the infant constant.

Exponate the coefficients for hazard ratios. This information is already given in the summary output.

Both hazard ratios are below 1 (i.e. they are associated with increased survivablity). Hazard ratio for sexboy (i.e. exp(coef)) = 0.62. Therefore, the hazard ratio for boys as opposed to girls is 62 percent; in other words, out of 100 infants, 62 that die are most likley to be male as to apposed to 58 females.

Hazard ratio for age = 0.96.

P-values: None of the coefficients are statistically significant

Run a chi-square test to test the null hypothesis that the hazards are proportional.

```r
### Run Chisq test
drop1(cox_mod, test = "Chisq")
## Both p-values are < 0.05. There is therefore evidence that we can fail to
    reject the null hypothesis
## that the hazards are proportional.
```