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 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.

Table 1:

	Dependent variable:
	child_surv
m.age	0.008^{***}
	(0.002)
sexfemale	-0.082^{***}
	(0.027)
Observations	26,574
R^2	0.001
Max. Possible R^2	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

The Cox model gives the hazard ratio for female infants relative to the male infants. The beta coefficient for sex = -0.082 indicates that females have a 8.2% lower risk of exit than males, in this data. The exponential coefficients also give the effect on hazard ratio of membership in the female group:

```
    > summary(coxm)
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 *exp(coef)* for *sexfemale* also shows being a female baby reduces the risk of exit or death by a factor of 0.9211, i.e. by around 8%.
The positive sign of the first coefficient (*m.age*) means that the hazard (risk of exit) is higher for children whose mother's are older. A one-unit increase in the age of a child's mother is associated with a 0.76% increase in the hazard ratio for that child.