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.

The R script for cox proportional hazard model is:

```
1
2 # load libraries
3 install.packages("eha")
4 install.packages("survival")
5 library(eha)
6 library(survival)
7 library(stargazer)
8
9 # Load child dataset
10 child_data = data("child")
11
12 # Estimation of Cox proportional hazard model
13 child_mortality <- coxph(Surv(enter, exit, event) ~ m.age + sex, data = child)
14 summary(child_mortality)
```

As evident from table 1, the significant p-values of both the predictors that are mother's age and sex of the infant suggests that both are statistically strong predictors of infant mortality. The 0.008 coefficient value of mother's age shows that with a one year increase in mother's age, the infant mortality also increases by the exponent of this coefficient. Thus the logged hazard ratio is higher with an increase of mother's age. Similarly, the sex female coefficient of - 0.082 implies that there is a lower logged hazard ratio for female infants and thus a reduced risk of female infant mortality by exponent of sexfemale coefficient as compared to male infants.

Table 1:

	Dependent variable:
	enter
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