Problem Set 4

Student: Shekhar Kedia (23351315)
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.

Solution:
Firstly using `R`, we load the library package and the required dataset by running the following codes:

```
1  # Loading the dataset in the environment
2  install.packages("eha")
3  library(eha)
4  data(child)
5  View(child)
```

Then, we run the Cox Proportional Hazard model using mother's age and infant's gender as covariates and create the summary output using the following codes:

```
1 # Estimating Cox P–H model
2 install.packages("survival")
3 library(survival)
4
5 child_mortality <- coxph(Surv(enter, exit, event) ~ m.age +
6                          sex, data = child)
7 summary(child_mortality)
8
9 library(stargazer)
10 stargazer(child_mortality) #Producing the LaTeX output
```

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

We can make the following interpretations from the results:

- **Mother's age (`m.age`):** Every additional increase in year of mother's age is associated with an increase of 0.008 log hazard on average for child mortality keeping gender of the child constant.

- **Gender of the child (`sexfemale`):** With increase in gender value i.e. moving from male to female child, the log hazard for child mortality decreases by 0.082 on average keeping mother's age constant.

Overall, this output suggests that both mother's age and gender of the child are significant predictors of the event, with older mothers having a slightly higher hazard and female child having a lower hazard compared to males.