Problem Set 4

Yana Konshyna

April 6, 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.

Uploading "child" dataset from the `eha` library, then creating a survival object for survival analysis `child_surv`.

Running a Cox Proportional Hazard regression `child_surv_CoxPH`.

```
1  # Loading data
2  data(child)
3  summary(child)
4
5  # Creating a survival object for survival analysis
6  child_surv <- with(child, Surv(enter, exit, event))
7  summary(child_surv)
8
9  # Estimating duration Cox Proportional Hazard regression model
10 child_surv_CoxPH <- coxph(child_surv ~ m.age + sex, data = child)
11 stargazer(child_surv_CoxPH)
```

Table 1:

	Dependent variable:
	child_surv
m.age	0.008***
	(0.002)
sexfemale	−0.082***
	(0.027)
Observations	26,574
R²	0.001
Max. Possible R²	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

Interpretation of the Estimated Coefficients:

- **m.age:** On average, one unit increase in mother's age is associated with an increase in the logged hazard rate for child by a multiplicative factor of $e^{0.008} \approx 1.01$, holding gender of child constant. The coeffcient has a statistically significant effect.

- **sexfemale:** On average, having a girl, comparing to having a boy, is associated with a decrease in the logged hazard rate by a multiplicative factor of $e^{-0.082} \approx 0.921$, holding age of mother constant. When hazard rate < 1 that means we have decreasing in hazard, and girls are about 8% more likely to survive than boys, holding age of mother constant. The coeffcient has a statistically significant effect.