Problem Set 4

Yuanyuan Liu

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.

```
#Load library
library(eha)
library(survival)
library(stargazer)
library(ggfortify)

#Question 1
#Load child dataset
library(eha)
data(child)

#creat a variable of child_surv
child_surv <- with(child, Surv(enter, exit, event))

#fit a Cox Proportional Hazardmodel using mother's age and infants gender as
    covariates
```

```
cox_model <- coxph(child_surv~ m.age+sex, data = child)
summary(cox_model)
#test the coefficient
drop1(cox_model,test = "Chisq")
#summary this model
stargazer(cox_model,type = "text")

# Calculate and interpret the hazard ratios (exponential of coefficients)
exp(coef(cox_model))

#plot this model
cox_fit <- survfit(cox_model)
autoplot(cox_fit)
```

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

One unit increase in mother's age associate with a 0.008 increase in the expected to log of the hazard , holding infant's gender constant. There is a 0.08 decrease in the expected log of hazard for female babies compared to male,holding mother's age constant.

Table 2:

m.age	sexfemale
1.008	0.921

The hazard ratio of female babies is 0.92 that of male babies, female babies are less likely to die(92 female babies die for every 100 male babies; female deaths are 8One unit increase of mother's age associate with hazard by 1.0076.

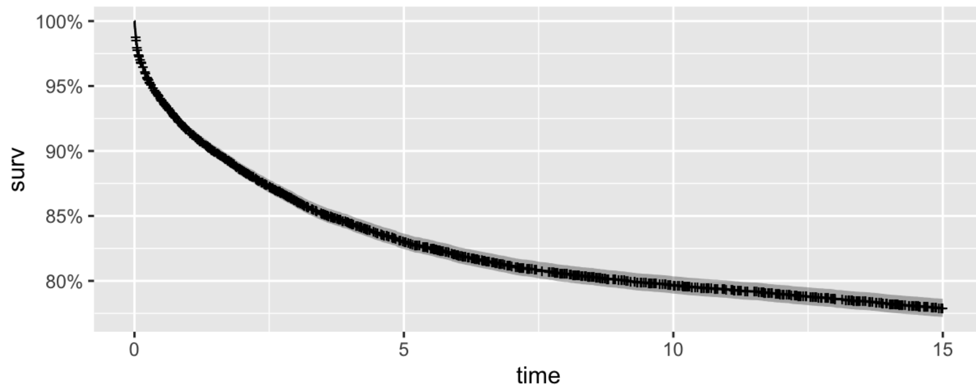

Figure 1: plot

The Kaplan-Meier curve starts at 100% and decreases over time, indicating that the individual's survival rate decreases over time. The small black line in the figure represents the uncertainty in the estimate of survival time, or the number of survival state changes at a certain point in time, which is often called censoring. When censoring occurs, estimates of the survival function can only be based on those individuals who are still alive (or in study) up to that point. It can be seen that as time goes by, the survival probability drops to about 75%.