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.

I created a survival object using the `Surv()` function.

```
child_surv <- with(child, Surv(enter, exit, event))
```

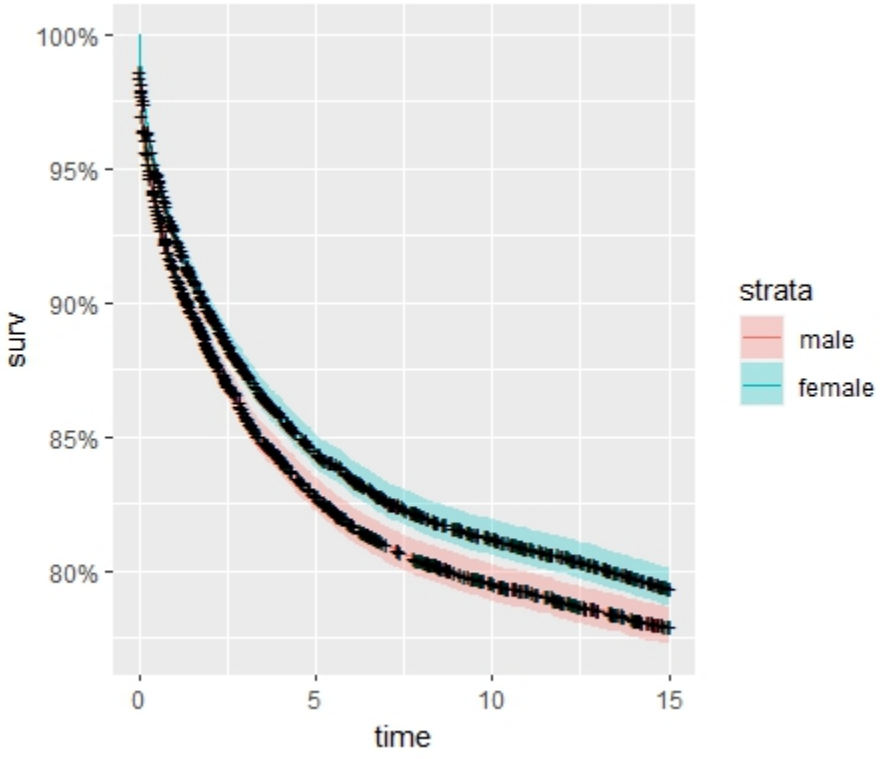
And then fitted a Cox Proportional Hazard Model.

```
cox <- coxph(child_surv ~ sex + m.age, data = child)
```

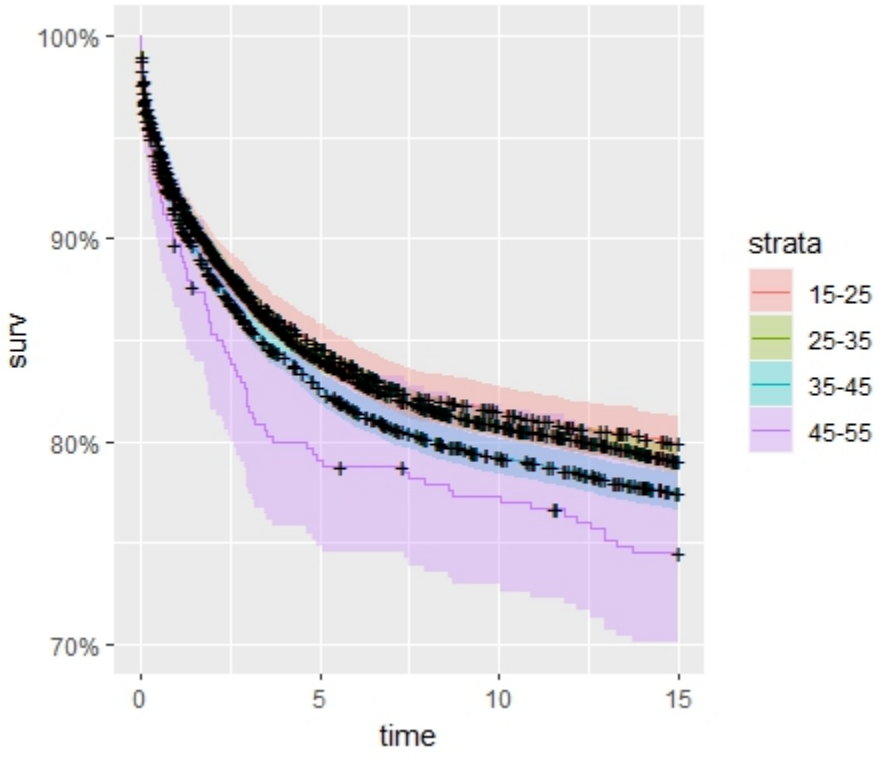
This provided the results seen in table 1. As we can see, both sex and the age of the mother showed statistically significant results. There is a 0.082 decrease in the expected log of the hazard for female babies compared to male, holding the age of the mother constant constant. There is a 0.008 increase in the expected log of the hazard for every 1 year increase in age, holding sex constant.

Table 1:

	Dependent variable:
	child_surv
sexfemale	-0.082^{***}
	(0.027)
m.age	0.008^{***}
	(0.002)
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



When plotting the results, we can see that male children have a slightly lower rate of survival compared to female children.



In order to plot the effect of the mother's age, i binned the column into the categories 15-25, 25-35, 35-45 and 45-55. This allows us to compare survival rates for the children of differently aged mothers.

```r
# Define custom breakpoints for age bins
breakpoints <- c(15, 25, 35, 45, 55)

# Create bins using cut() function
child$m.age_bin <- cut(child$m.age, breaks = breakpoints, labels = c("15-25",
    "25-35", "35-45", "45-55"), include.lowest = TRUE)
```

As the results in the table showed, the older the mother, the lower the rate of survival for children.