Applied Stats II: Problem Set 4

Luna Goldstein

Due: April 4, 2022

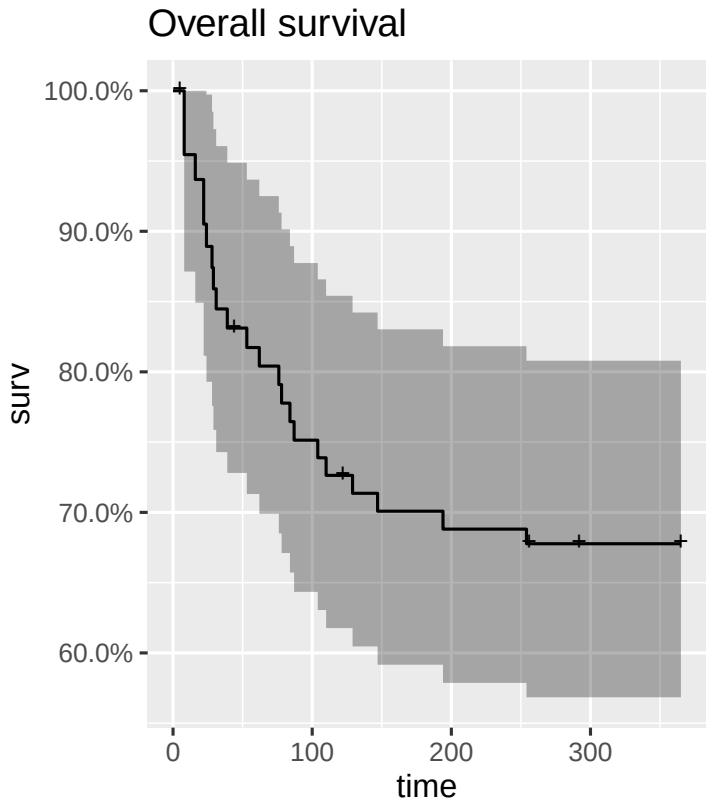
Instructions

- 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 class on Monday April 4, 2022. No late assignments will be accepted.
- Total available points for this homework is 80.

Question 1

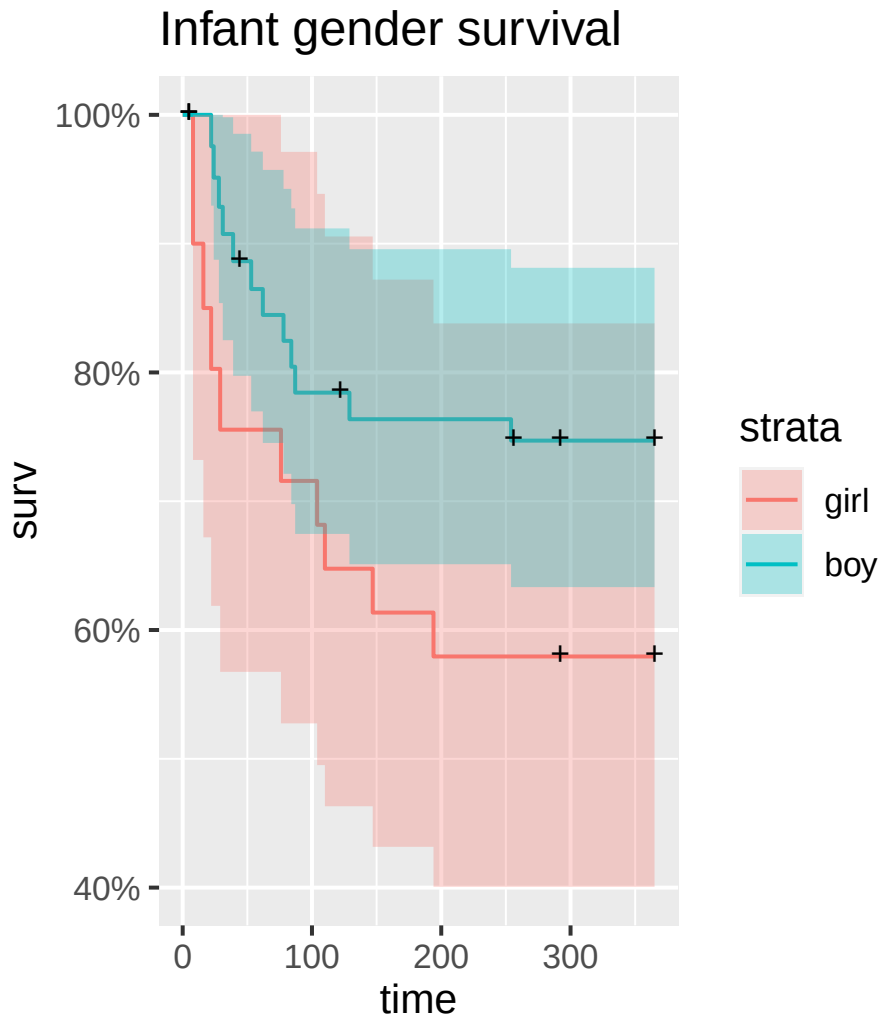
We're interested in modeling the historical causes of infant mortality. We have data from 5641 first-born in seven Swedish parishes 1820-1895. Using the "infants" dataset in the `eha` library, fit a Cox Proportional Hazard model using mother's age and infant's gender as covariates. Present and interpret the output.

We will use R's plotting functions to produce a Kaplan-Meier plot of the data. Firstly, to plot the overall data, and secondly comparing the mortality of the female and male babies.



We use the product-limit survival plot to visualize and estimate the survival function from 5641 Swedish first-born lifetime data.

Then, we are comparing the female and male babies.



Cox Proportional Hazard model using mother's age and infant's gender (covariates).

```
cox <- coxph(dataset ~ sex + age, data = infants)
```

Table 1:

	Dependent variable:
	dataset
sexboy	−0.485
	(0.442)
age	−0.040
	(0.045)
Observations	105
R^2	0.019
Max. Possible R^2	0.800
Log Likelihood	−83.626
Wald Test	2.000 (df = 2)
LR Test	1.992 (df = 2)
Score (Logrank) Test	2.034 (df = 2)
Note:	*p<0.1; **p<0.05; ***p<0.01

The output:
There is a 0.485 decrease in the expected log of the hazard for male first-borns compared to female, holding their mother's age constant.
For every one year increase in the mother's age there is a 0.04 decrease in the expected log of the hazard for first born, holding sex constant.

Assessing model's quality:

```
drop1(cox, test = "Chisq")
```

Exponentiated parameter estimates to obtain hazard ratios:

```
    exp(-0.083546)
```

The hazard ratio of male babies is 0.62 that of female babies, i.e. female babies are less likely to do (62 male babies die for every 100 female babies; male deaths are 38% lower, etc.)