

# Problem Set 4

Kaley Burg

April 11, 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.

- First I imported the data

```
1 data(child)
2
3 child_surv <- with(child, Surv(enter, exit, event))
4 #creating km object to use for plots
5 km <- survfit(child_surv ~ 1, data = child)
```

- Next I fit a Cox Proportional Hazard model using mother's age and infant's gender as covariates

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

- We can see the results of the overall Kaplan Meier plot in Figure 1 and the results of the Kaplan Meier plot by sex in Figure 2

- The code for these plots is as follows:

```
1 plot_km <- autoplot(km)
```

```
1 sex_km_plot <- autoplot(km_sex)
```

- The results for the Cox Proportional Hazard model are shown in Table 1, with the hazard ratios shown in Table 2.
- I also exponentiated the results for Table 2

```
1 exp(cox$coefficients)
```

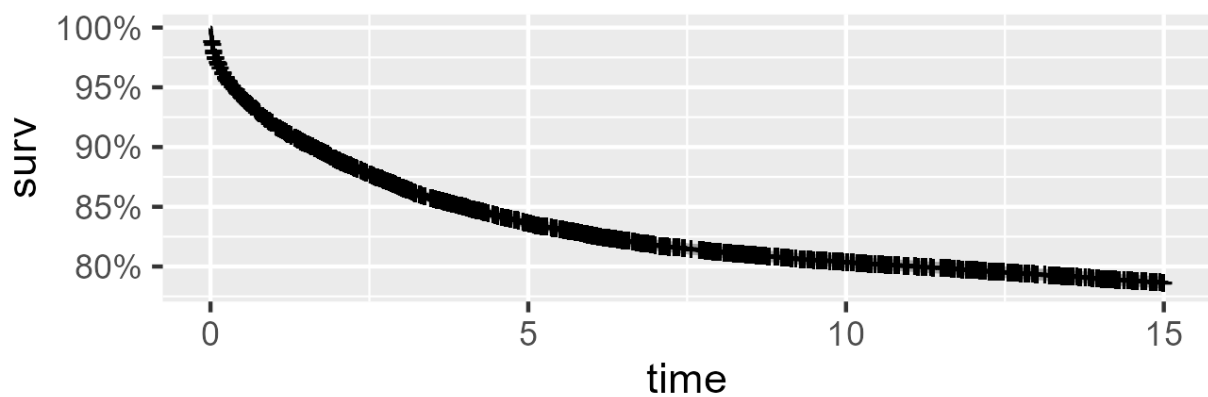


Figure 1: Kaplan-Meier Survival Plot

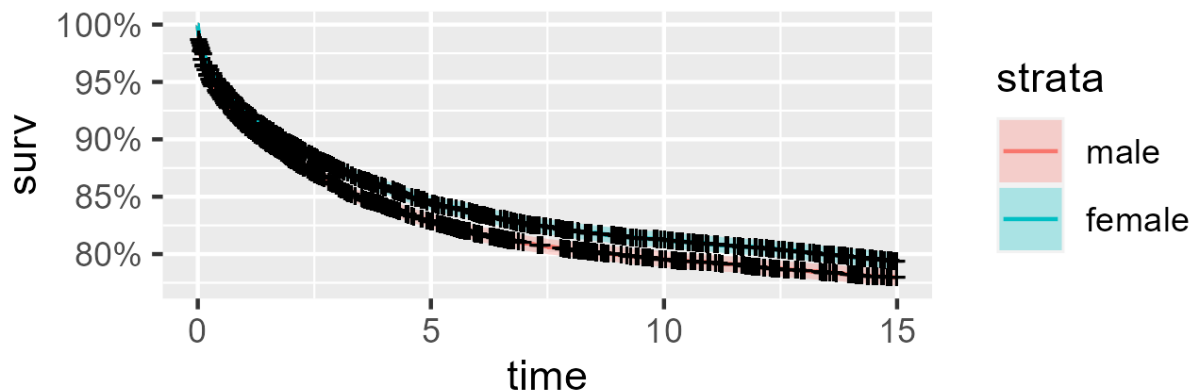


Figure 2: Kaplan-Meier Survival Plot by Sex

Table 1:

<i>Dependent variable:</i>	
	child_surv
sexfemale	-0.082*** (0.027)
m.age	0.008*** (0.002)
Observations	26,574
R <sup>2</sup>	0.001
Max. Possible R <sup>2</sup>	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

- From the table above, we can see that there is a 0.082 decrease in the expected log of the hazard for female babies compared to male, holding mother's age constant. There is a 0.008 increase in the expected log of the hazard for babies for every one unit increase in mother's age, holding sex constant.
- We can also exponentiate these results to obtain hazard ratios

Table 2:

sexfemale	m.age
0.921	1.008

- From the hazard ratio table (Table 2) we can see that:
- The hazard ratio of female babies is 0.92 that of male babies, meaning that female babies are less likely to die (92 female babies die for every 100 male babies; female deaths are 8% lower)
- For every one unit increase in mother's age, the hazard ratio increases by 1.01, i.e. babies with older mothers are more likely to die (101 babies with older mothers for every 100 babies with younger mothers, older mother deaths are 0.01% higher)