Problem Set 4

Applied Stats II

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.

```
# Load data
data(infants)

# Creating a Surv object
infant_surv <- Surv(enter, exit, event)
# infant_surv <- with(infants, Surv(enter, exit, event)) # Alternative option

# Plot Kaplan-Meier curves
# Overall survival
km <- survfit(infant_surv ~ 1, data = infants)
autoplot(km,
         main = "Overall survival Rate of Infants",
         xlab = "Time",
         ylab= "Survival Rate")
```

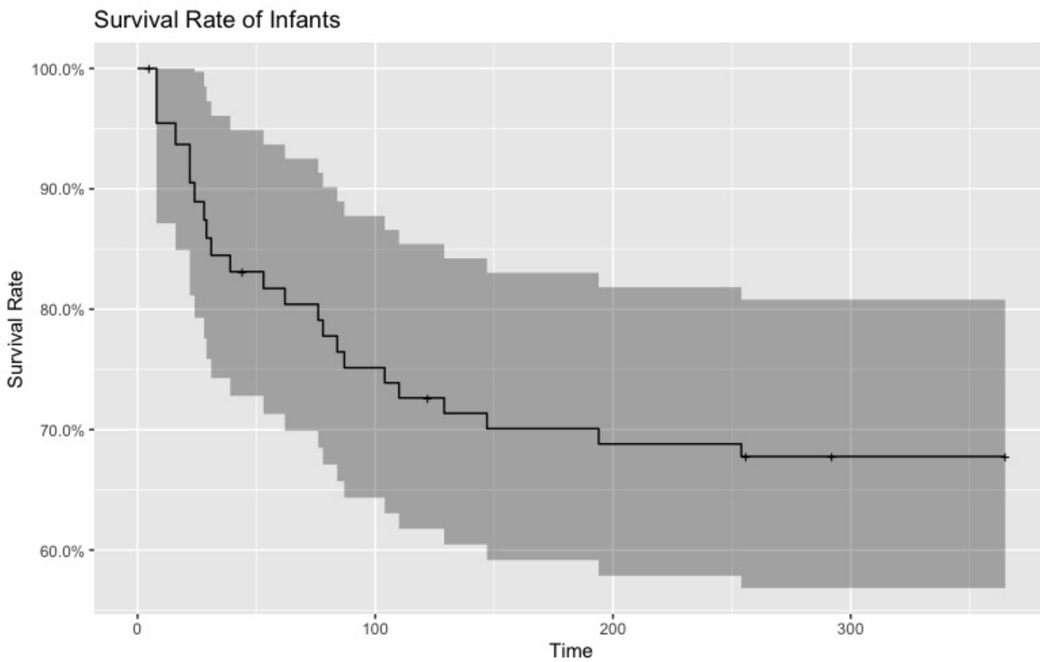

Figure 1: Survival Rate of Infants

Survival divided between control group and study group

```
km_groups <- survfit(Surv(infants$enter, infants$exit, infants$event) ~ mother
    , data = infants)
autoplot(km_groups,
            main = "Survival Rate of Infants Whose Mothers are Alive or Dead",
            xlab = "Time",
            ylab= "Survival Rate")
```

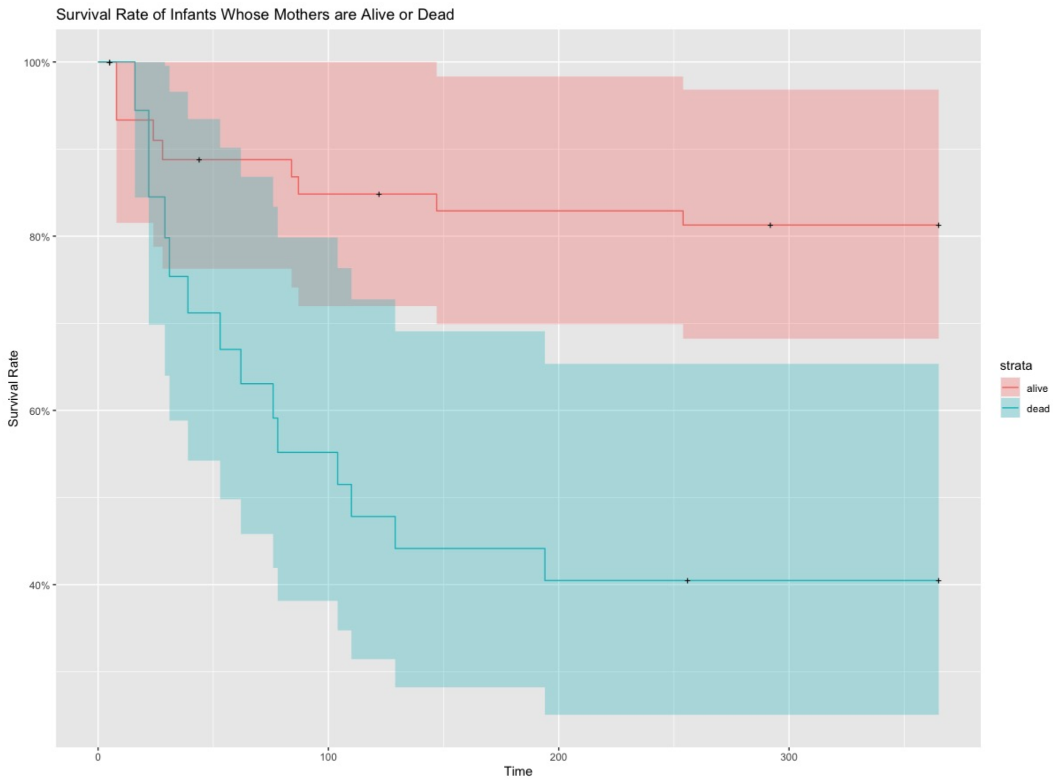

Figure 2: Survival Rate of Infants based on mother's status

As we can see, the survival rate for infants whose mothers had died is much lower than infants whole mothers were alive.

Cox Proportional Hazard Regression

```
cox <- coxph(Surv(enter, exit, event) ~ age + sex, data = infants)
survfit(cox)
summary(cox)
```

Table 1: Cox PH Regression Results

	exp(coef)	exp(-coef)	lower .95	upper .95
age	0.9604	1.041	0.8792	1.049
sexboy	0.6156	1.624	0.2587	1.465

Interpretation of the Cox PH regression results
A 1 unit increase in age results in a decrease in hazard rate by a factor of 0.96, holding everything else at its mean. A 1 unit increase in sex (from female to male) results in a decrease in hazard rate by a factor of 0.62, holding everything else at its mean.

Assessing the model with a' Chisq' test

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

Table 2: Assessing the Chisq test results

	Df	*AIC*	*LRT*	Pr(>Chi)
(none)		171.25		
age	1	170.12	0/87104	0.3507
sexboy	1	170.42	1.16548	0.2803

Results
Running a Chisq test on the model our p-scores in this test are 0.2803 and 0.3507 for sex and age respectively, suggesting that neither is significant using this model.

Plot Cumulative Hazard Function

```
plot_cox <- coxreg(Surv(enter, exit, event) ~ age + sex, data = infants)
plot(plot_cox,
     main = "Cumulative Hazard Function",
     xlab = "Duration",
     ylab = "Hazard Rate",
     xlim = c(0, 365))
```

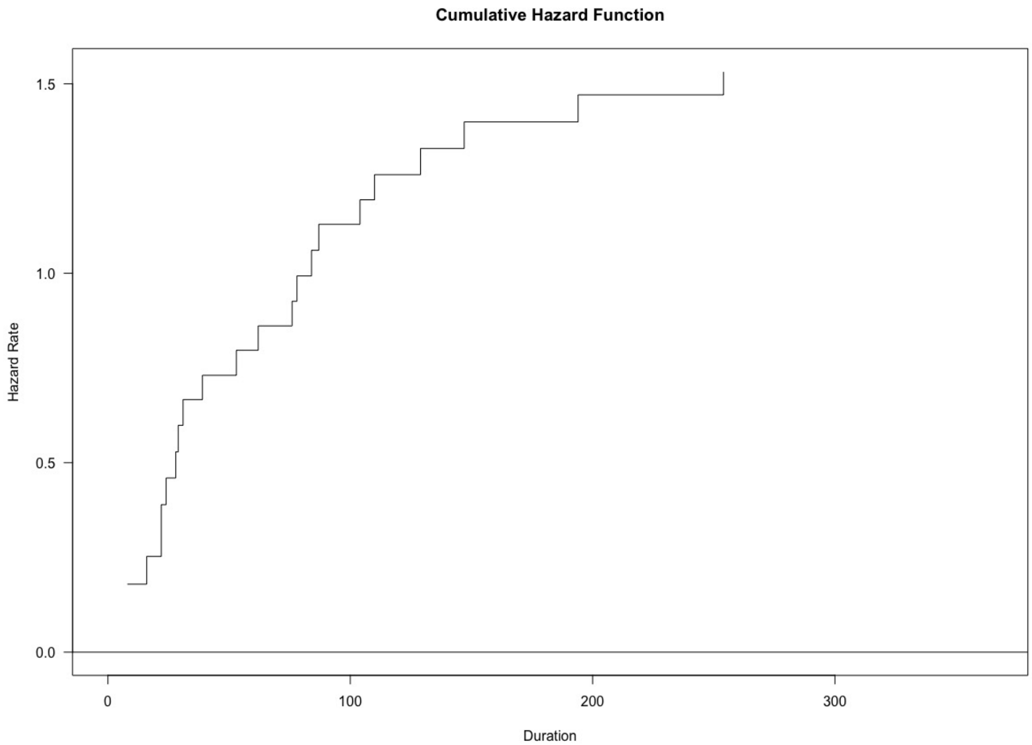

Figure 3: Cox PH Function Plot

Interpretation of the Cumulative Hazard Function
As can be seen from the plot, the hazard rate increases sharply over the first hundred days then becomes more gradual. Overall the rate is quite low, topping out at around 1.5%.