Problem Set 4 - Laura McPhillips

Applied Stats II

Due: April 4, 2022

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:

```
install.packages("eha")
library(eha)
data(infants)
```

Cox P-H model:

```
cox <- coxph(infant_surv ~ sex + age, data = infants)
summary(cox)
cox_chisq <- drop1(cox, test = "Chisq")
cox_fit <- survfit(cox)
autoplot(cox_fit)
```

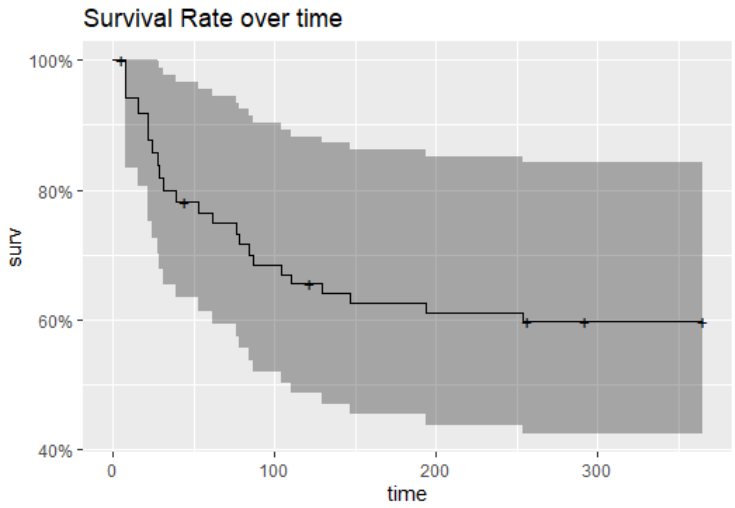

Table 1:

	Dependent variable:
	infant_surv
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

Table 2:

Statistic	N	Mean	St. Dev.	Min	Max
Df	2	1.000	0.000	1	1
AIC	3	170.598	0.586	170.124	171.253
LRT	2	1.018	0.208	0.871	1.166
Pr(>Chi)	2	0.315	0.050	0.280	0.351

Interpretation:
There is a 0.49 decrease in the expected log of the hazard for male babies compared to female, holding age constant.
For each unit increase in age, there is a 0.4 decrease in the expected log of the hazard for babies, holding sex constant. (neither of these are statistically significant)

Hazard Ratios (exponential of coefficients):
The hazard ratio for male babies is 0.62 that of female babies, i.e. 62 male babies die for every 100 females.

The hazard ratio for age is 0.96, i.e. the expected hazard decreases by 0.96 relative to one day increase in age.