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.

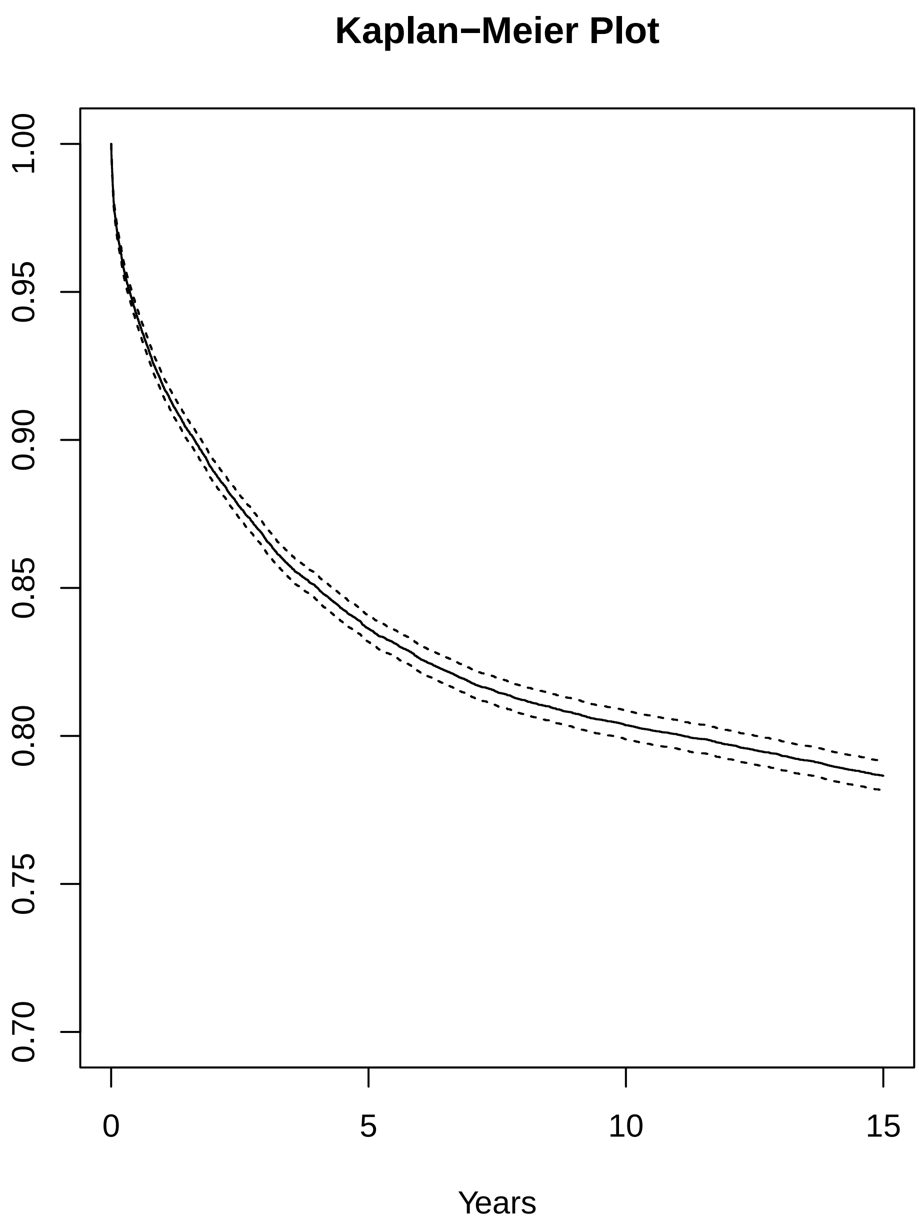
Question 1

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.

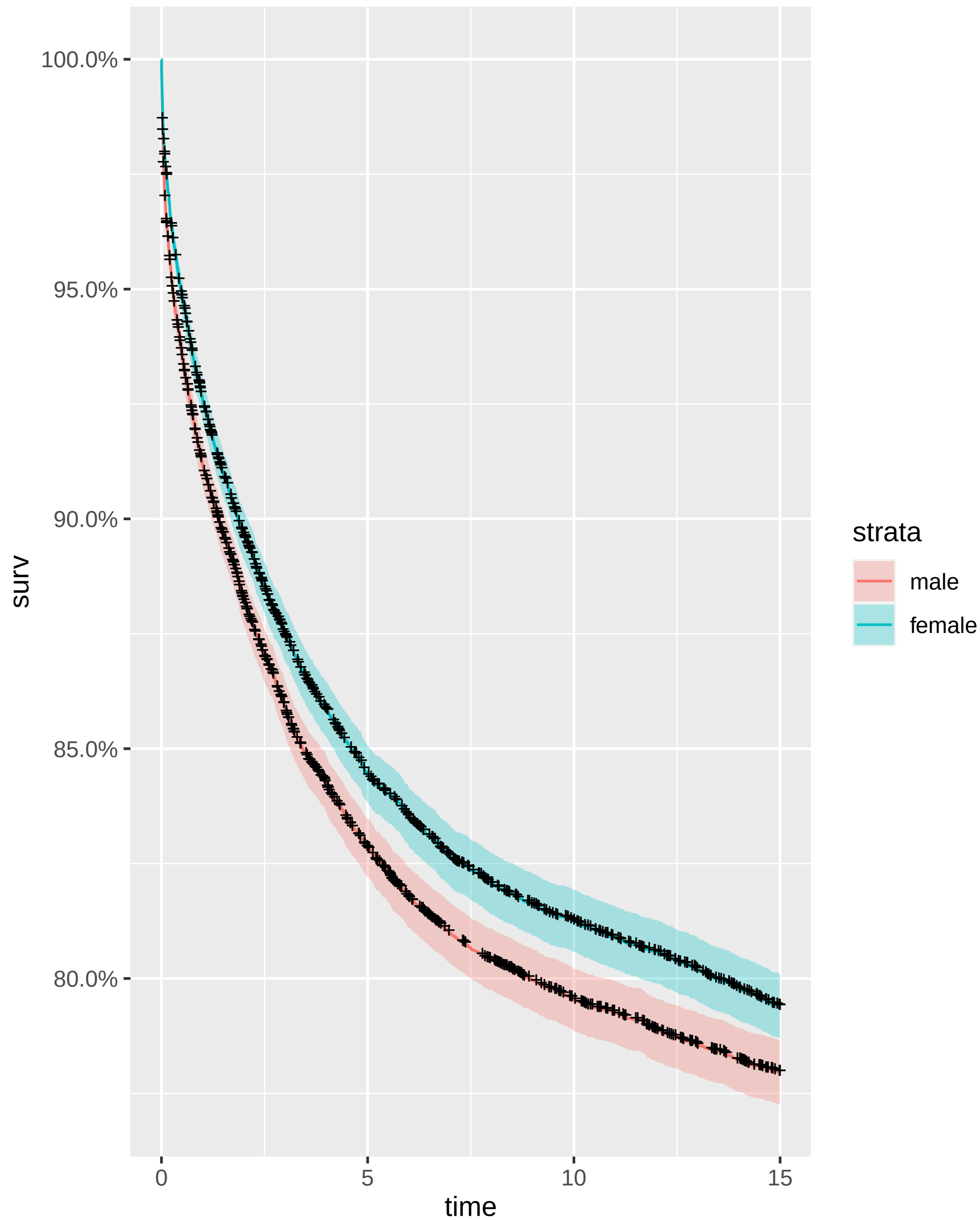
```
data(child) #our dataset

child_surv <- with(child, Surv(enter, exit, event)) #building a survival
    object for children
km <- survfit(child_surv ~ 1, data = child) #children end at the age of 15
summary(km, times = seq(0, 15, 1))
plot(km, main = "Kaplan-Meier Plot", xlab = "Years", ylim = c(0.7, 1))
autoplot(km) #visualization of the survival rate of children in our dataset
    over time period of 15 years.

km_m.age <- survfit(child_surv ~ m.age, data = child) #doesn't give a readable
     output due to formatting issues
autoplot(km_m.age)

```

```
km_sex <- survfit(child_surv ~ sex, data = child) #gives a visualization of
    the survival rates of male vs female children over first 15 years
autoplot(km_sex) #we can see that female children have a higher survival rate,
    when compared to their male counterparts
```

First, we assign a survival object for the children using entry, exit, and event data to the age of 15. We are then able to get a graph of the total population's survival rate. As seen below.



We are also able to graph the difference in survival rates between males and females.



The next stage is to fit a Cox Proportional Hazard Model, which we do with the following code.

```
cox <- coxph(Surv(enter, exit, event) ~ sex + m.age, data = child)
summary(cox)
drop1(cox, test = "Chisq") #9.4646
stargazer(cox, type = "latex")
```

There is a 0.0822 decrease in the expected log of the hazard for female babies compared to male babies, holding m.age constant. There is a 0.0076 increase in the expected log of the hazard for babies of mothers with one extra unit of age, holding sex constant. We also note that m.age has a p-value that is statistically significant, as is the p-vlaue for gender.

Table 1:

	Dependent variable:
	enter
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

```
exp(-0.082215) #exponentiate parameter estimates to obtain hazard ratios for
    gender
```

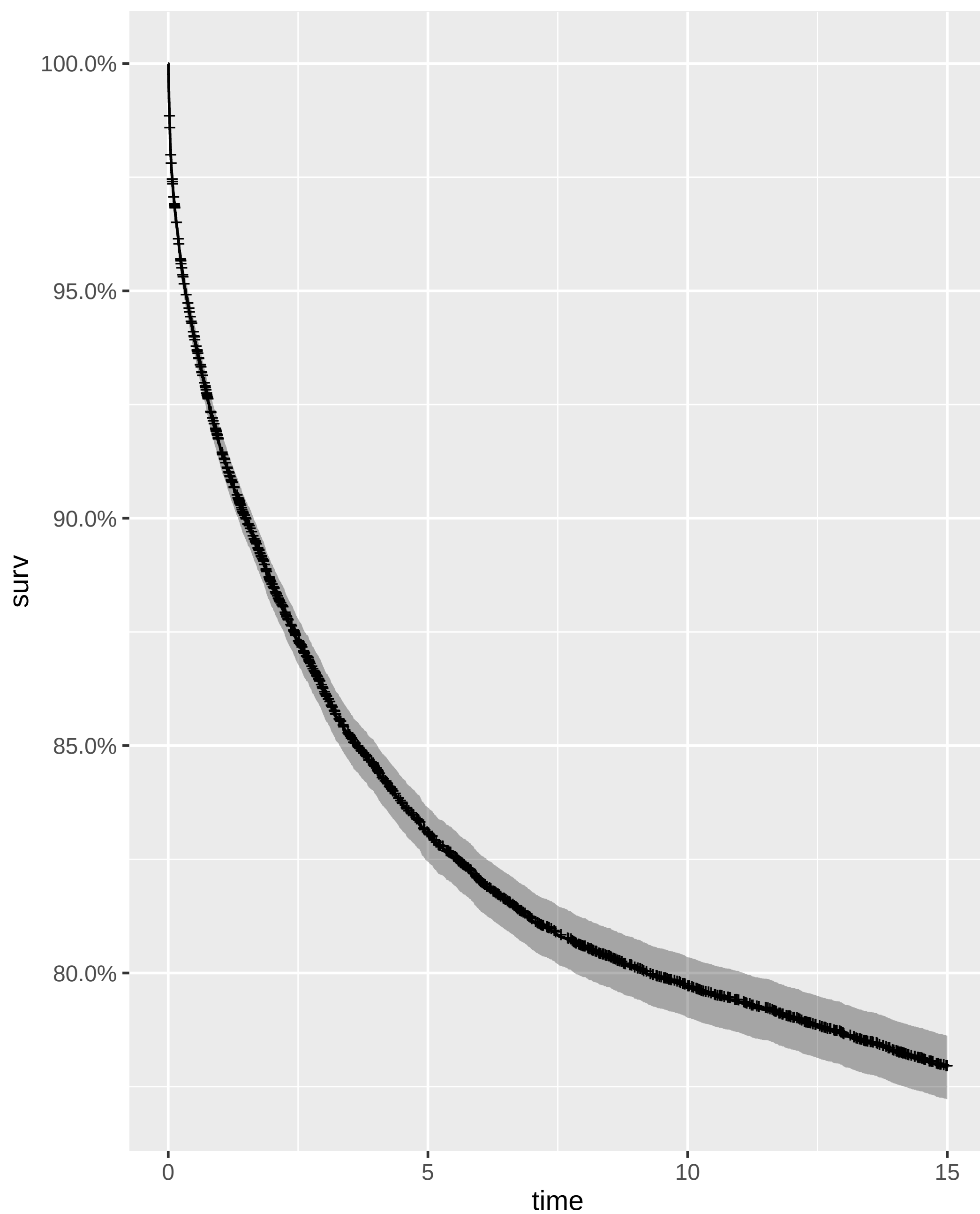
The hazard ratio of female babies is 0.9210739 that of male babies. This means that for every 100 male babies that die, we would expect that only 92 female babies would die. This could be interrupted in another sense as female babies have a 8 percent lower death rate than equivalent male babies.

```
exp(0.007617) #exponentiate parameter estimates to obtain hazard ratios for
    mother's age
```

The hazard ratio of a mother aged 0 is 1.0076466, this number is not quite accurate as no mother can have a child a age zero but it does allow us to analyse a one unit increase in age, and its affect on infant mortality. This could be interrupted in another sense as for each unit increase in a mothers age, the chance of that child dying is 0.7 percent higher than if the mother was one unit of age younger.

```
cox_fit <- survfit(cox)
autoplot(cox_fit)
```

We now graph the two covariets in our model. Which shows that as age increase the bounds of certinly also expand.



We now will run an interaction model to compare with additive model.

```
# Adding an interaction model
cox.int <- coxph(child_surv ~ sex * m.age, data = child)
summary(cox.int)
drop1(cox.int, test = "Chisq") #0.10623
stargazer(cox.int, type = "latex")
```

There is a 0.127105 decrease in the expected log of the hazard for female babies compared to male babies, holding m.age constant.

There is a 0.006963 increase in the expected log of the hazard for babies of mothers with one extra unit of age, holding sex constant. However the P-Values are at level that they are not statistically significant when compared to the additive model.

Table 2:

	Dependent variable:
	child_surv
sexfemale	−0.127
	(0.140)
m.age	0.007**
	(0.003)
sexfemale:m.age	0.001
	(0.004)
Observations	26,574
R^2	0.001
Max. Possible R^2	0.986
Log Likelihood	−56,503.430
Wald Test	22.530*** (df = 3)
LR Test	22.624*** (df = 3)
Score (Logrank) Test	22.562*** (df = 3)
Note:	*p<0.1; **p<0.05; ***p<0.01

```
exp(-0.127105) #exponentiate parameter estimates to obtain hazard ratios for
    mother's age
```

The hazard ratio of a mother aged 0 is 0.8806412 that of male babies. This means that for every 100 male babies that die, we would expect that only 88 female babies would die. This could be interrupted in another sense as female babies have a 8 percent lower death rate than equivalent male babies.

```
exp(0.006963) #exponentiate parameter estimates to obtain hazard ratios for
    mother's age
```

The hazard ratio of a mother aged 0 is 1.006987, this number is not quite accurate as no mother can have a child a age zero but it does allow us to analyse a one unit increase in age, and its affect on infant mortality. This could be interrupted in another sense as for each unit increase in a mothers age, the chance of that child dying is 0.6987 percent higher than if the mother was one unit of age younger.

While the interactive model had a Chisq value of 0.10623, and the additive has a value of 9.464. The P values in the additive model were statisically significant and better explain the covariets affect on child mortality in this case.