Problem Set 4

Applied Stats II

Due: April 16, 2023

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 Sunday April 16, 2023. 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, the data was imported, after which a Cox Proportional Hazard Regression model was fit to the data:

```
#load in data
data(child)

#carrying out coxph regression
coxmod <- coxph(Surv(enter, exit, event) ~ sex + m.age, data = child)
```

This gave the following output:

Table 1: Cox Proportional Hazard Regression

	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

As can be seen, both of the covariates had significant effects on the log of the hazard, being the death of the child. In the case of the sex covariate, holding the age of the mother constant, there is a .08 decrease in the expected log of the hazard when the baby is female. This means that the hazard ratio for female babies is 0.92 to that of the male babies, meaning female deaths are lower by 8% compared to that of male babies

Holding sex constant, for every year increase in the mother's age, there is a .008 increase in the log of the hazard being the death of the child. This means that for every year increase in the mother's age, there is a 0.7% increase in the expected hazard, being the death of their child.

A likelihood ratio test was also carried out to assess whether the model is more useful than a null model. As shown in Table 2, the results indicate that using both covariates are more useful than a null model, with the p values being less than 0.05 for both.

Using this model, we can plot the Cumulative Hazard Function, indicating the probability of the hazard occurring as the time passed, as shown in Figure 1

Table 2: Likelihood Ratio Test

	Df	AIC	LRT	Pr(>Chi)
<none>		113,011.000		
sex	1	113,018.400	9.465	0.002
m.age	1	113,021.800	12.795	0.0003

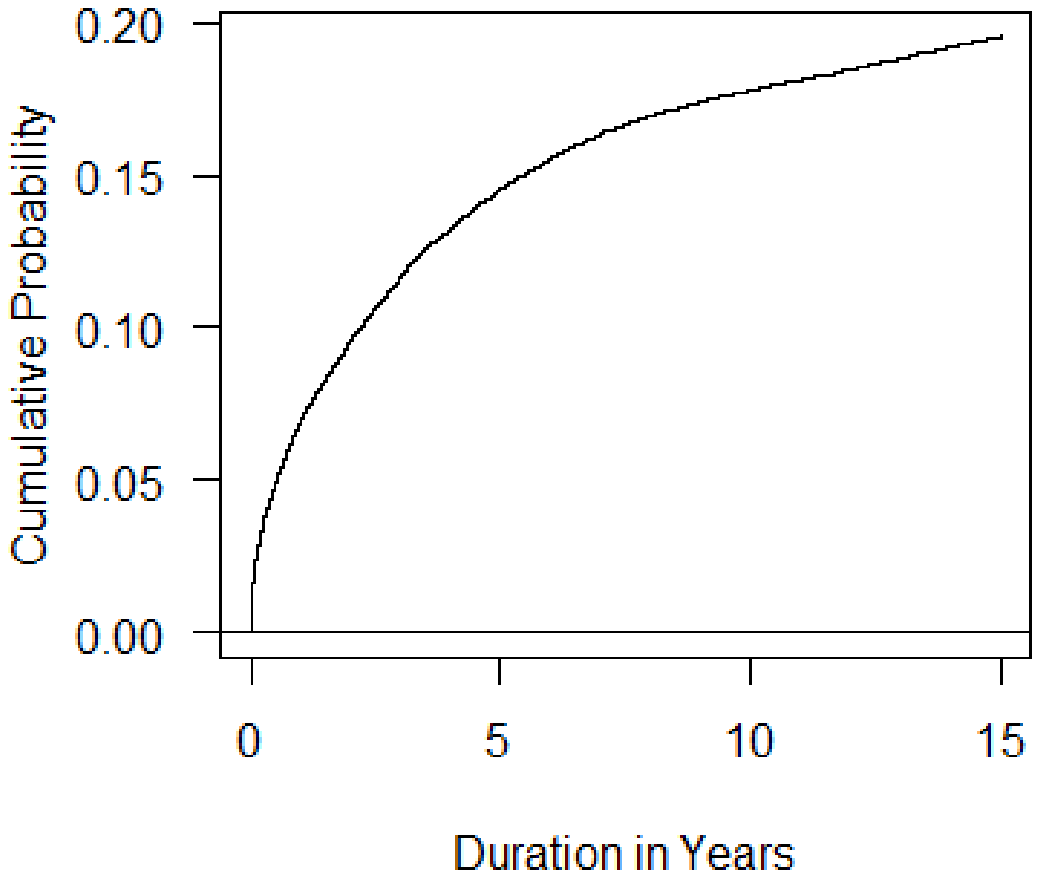

Figure 1: Cumulative Hazard Function - Probability of Death Occurring as Child Ages