

# Problem Set 4

## Applied Stats II

Due: April 12, 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 and created the child survival object:

```
1 data(child)
2 child_surv <- with(child, Surv(enter, exit, event))
```

Then I fit the Cox PH model:

```
1 cox <- coxph(child_surv ~ sex + m.age, data = child)
2 summary(cox)
3 drop1(cox, test = "Chisq")
4 stargazer(cox, type = "latex")
```

This generated the following results:

Table 1: Cox PH Model Results

	<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)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Based on the above coefficients, going from male babies to female ones is associated with a decrease of 0.082 in the log hazard, on average, while controlling for the age of the mother. This means that male babies have higher mortality rates. As for the age of the mother, one additional year of age is associated with an increase of 0.008 in the log hazard for the babies, on average, while controlling for the gender of the babies. This means that as mothers become older, the babies face worse chances of survival.

Then I exponentiated the coefficients to also interpret in terms of the hazard ratio:

```

1 # Exponentiating Coefs for Interpretation
2 exp(cox$coefficients)
3 stargazer(exp(cox$coefficients))

```

This yielded the following:

Table 2: Exponentiated Coefficients

sexfemale	m.age
0.921	1.008

From the table, we can interpret that the hazard ratio for female babies is 0.921 times that of male babies (so for every 92.1 females who die, 100 male babies do). Again, female babies have slightly better chances of survival. With regard to age, we can interpret that the hazard ratio for children with mothers with one additional year of age is 100.8 times that of children whose mothers are one year younger. Again, babies with younger mothers face better chances of survival.