

# 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.

```
1 #Load library and dataset
2
3 library(aha)
4 library(survival)
5 data(child)
6
7 #Explore dataset
8 str(child)
9 View(child)
10
11 #Fit the Cox Proportional Hazard model:
12 cox_model <- coxph(Surv(enter, exit, event) ~ m.age + sex, data = child)
13 print(cox_model)
```

```
> str(child)
'data.frame': 26574 obs. of 10 variables:
 $ id      : int  9 150 158 178 263 342 363 393 408 486 ...
 $ m.id    : int  246606 377744 118277 715337 978617 282943 341341 840879 586140 564736
 $ sex     : Factor w/ 2 levels "male","female": 1 1 1 1 2 1 1 1 2 2 ...
 $ socBranch: Factor w/ 4 levels "official","farming",...: 2 2 4 2 4 2 2 2 2 2 ...
 $ birthdate: Date, format: "1853-05-23" "1853-07-19" ...
 $ enter    : num  0 0 0 0 0 0 0 0 0 0 ...
 $ exit     : num  15 15 15 15 0.559 0.315 15 15 15 15 ...
 $ event    : num  0 0 0 0 1 1 0 0 0 0 ...
 $ illeg    : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 2 ...
 $ m.age    : num  35 30.6 29.3 41.2 42.1 ...
```

Call:

```
coxph(formula = Surv(enter, exit, event) ~ m.age + sex, data = child)
```

|           | coef      | exp(coef) | se(coef) | z      | p        |
|-----------|-----------|-----------|----------|--------|----------|
| m.age     | 0.007617  | 1.007646  | 0.002128 | 3.580  | 0.000344 |
| sexfemale | -0.082215 | 0.921074  | 0.026743 | -3.074 | 0.002110 |

```
Likelihood ratio test=22.52 on 2 df, p=1.289e-05
n= 26574, number of events= 5616
```

We use the "event" variable as indicator in the survival analysis (indicates the occurrence of an event, in this case death).

The Cox Proportional Hazard model results indicate that both mother's age (m.age) and infant's gender (sex) are significantly associated with child mortality in Skellefteå, Sweden from 1850 to 1884.

Mother's age has a hazard ratio of  $\exp(0.007617) = 1.0076$ , indicating that for each one-unit increase in mother's age, the hazard of child mortality increases by approximately 0.76 percentage. This association is statistically significant ( $p < 0.001$ ).

Infant gender also shows a significant association with child mortality. Female infants have a hazard ratio of  $\exp(-0.082215) = 0.9211$  compared to male infants, indicating that female infants have a lower hazard of mortality compared to male infants. This association is also statistically significant ( $p = 0.0021$ ).

The results indicate that both mother's age and infant's gender are significantly associated with child mortality. Older mothers and male infants have higher hazards of mortality compared to younger mothers and female infants, respectively.