# Problem Set 4

Applied Stats II

**Maiia Skrypnyk 23371609**

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.

**Loading the data:**

```
child_data <- eha::child
```

**Fitting a Cox Proportional Hazard Model:**

```
model <- coxph(Surv(enter, exit, event) ~ m.age + sex, data = child_data)
summary(model)
```

Table 1: Model 1 (log odds)

| | *Dependent variable:* |
|---|---|
| | enter |
| m.age | $0.008^{***}$ |
| | (0.002) |
| sexfemale | $-0.082^{***}$ |
| | (0.027) |
| 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 |

**Exponentiating the coefficients:**

```
exp(model$coefficients)
```

*Please see the Model summary on the next page*

Table 2: Model 1 (odds ratio)

| | *Dependent variable:* |
|---|---|
| | enter |
| m.age | 1.008*** |
| | (0.002) |
| sexfemale | 0.921*** |
| | (0.027) |
| 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 |

**Interpretation:**

- `m.age`, **1.008**: On average and keeping gender predictor constant, a one year increase in mother's age is associated with an increase in the hazard of child mortality by a multiplicative factor of 1.008, or by approximately 0.8%. This coefficient is statistically differentiable from zero at the $\alpha$-level 0.01.

- `sexfemale`, **0.921**: On average and keeping mother's age predictor constant, a shift in child's gender from male to female is associated with an increase in hazard of child mortality by a multiplicative factor of 0.921, or a decrease by approximately 8%. This coefficient is statistically differentiable from zero at the $\alpha$-level 0.01.

  **In other words,** the older the mother, the higher the hazard of child mortality (by 0.8%). Moreover, female babies are 8% less likely to die than male babies (lower hazard of child mortality). (On average, keeping other predictors constant).