

# Problem Set 4

Brendan Walsh

Submitted: 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 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.

### Additive Survival Model R Code:

```
1 # Making dummy variables of the 'sex' column, male = 1, female = 0
2 child$sex <- ifelse(child$sex == 'male', 1, 0)
3
4
5 # Making the additive model:
6 Additive.hazard.model <- coxph(Surv(exit, event) ~ m.age + sex, data =
7   child)
8
9
10
11
12
```

```

13 summary(Additive.hazard.model)
14
15 # Additive model summary:
16 Call:
17 coxph(formula = Surv(exit, event) ~ m.age + sex, data = child)
18
19 n = 26574, number of events = 5616
20
21
22      coef exp(coef) se(coef)      z Pr(>|z|)
23 m.age    0.007617  1.007646  0.002128  3.580 0.000344 ***
24 sex      0.082215  1.085689  0.026743 -3.074 0.002110 **
25
26
27      exp(coef) exp(-coef) lower .95 upper .95
28 m.age      1.008      0.9924      1.003      1.012
29 sexfemale  1.086      0.9211      1.030      1.114
30
31 Concordance= 0.519 (se = 0.004 )
32 Likelihood ratio test= 22.52 on 2 df, p=0.00001
33 Wald test              = 22.52 on 2 df, p=0.00001
34 Score (logrank) test = 22.53 on 2 df, p=0.00001

```

### Testing the additive model's goodness of fit with a chi-squared test:

```

1 drop1(Additive.hazard.model, test = "Chisq")
2
3 # Chi-squared additive goodness of fit test results:
4 Single term deletions
5
6 Model:
7 Surv(exit, event) ~ m.age * sex
8      Df      AIC      LRT Pr(>Chi)
9 <none>      113013
10 m.age:sex  1 113011 0.10623  0.7445

```

### Interpreting the additive model:

From the additive model, gender (sex) and mother's age (m.age) are statistically significant at the 99% and 99.99% levels, with Wald statistics of 3.580 and 3.074 respectively. Judging by the confidence intervals, we can say that 9,999 times out of 10,000, the true coefficient for m.age falls between 1.003 and 1.012; and 99 times out of 100, the true coefficient for sex falls between 1.030 and 1.114.

The exponentiated coefficient (i.e., the hazard ratio) for gender (sex) - with male = 1 with sex as a dummy variable - of 1.086 indicates being male decreases survival rates by over 8.5% relative to being female, the reference category. In other words, being male in this model is a bad prognostic for surviving.

The exponentiated coefficient (i.e., the hazard ratio) for mother's age (m.age) of 1.0076 indicates for each additional year of age a mother is decreases survival rates of children 0.76%. In other words, the higher the mother's age in this model, the worst the prognostic for surviving.

The Chi-Squared Goodness of Fit test further confirms the significance of these two variables: m.age has a p-value of 0.0003476, and sex has a p-value of 0.0020947. These indicate we can accept the alternative hypothesis that m.age and sex are significant covariates in predicting survival.

I also ran my model using the Breslow and Extra methods (given that the default method for survival regression in R is the Efron method); these gave me very similar results, only differing in exponentiated coefficients and p-values for the model and the goodness of fit test after the fourth decimal points.

### Interactive Survival Model R Code:

```

1 Interactive.hazard.model <- coxph(Surv(exit, event) ~ age * sex, data =
  infants)
2
3
4 summary(Interactive.hazard.model)
5
6 # Interactive model summary:
7 Call:
8 coxph(formula = Surv(exit, event) ~ m.age * sex, data = child)
9
10      n= 26574, number of events= 5616
11
12      coef exp(coef) se(coef)      z Pr(>|z|)
13 m.age      0.008352  1.008387  0.003101  2.694  0.00706 **
14 sex        0.127105  1.135536  0.140304  0.906  0.36498
15 m.age:sex -0.001389  0.998612  0.004263 -0.326  0.74447
16
17      exp(coef) exp(-coef) lower .95 upper .95
18 m.age      1.0084      0.9917      1.0023      1.015
19 sex        1.1355      0.8806      0.8625      1.495
20 m.age:sex   0.9986      1.0014      0.9903      1.007
21
22 Concordance= 0.519 (se = 0.004 )
23 Likelihood ratio test= 22.62 on 3 df, p=0.00005
24 Wald test = 22.53 on 3 df, p=0.00005
25 Score (logrank) test = 22.56 on 3 df, p=0.00005

```

### Testing the interactive model's goodness of fit with a chi-squared test:

```

1 drop1(Interactive.hazard.model, test = "Chisq")
2
3 # Chi-squared additive goodness of fit test results:

```

```

4   Single term deletions
5
6   Model:
7   Surv(exit, event) ~ m.age * sex
8               Df      AIC      LRT Pr(>Chi)
9   <none>          113013
10  m.age:sex    1  113011  0.10623   0.7445

```

### Interpreting the interactive model:

The interactive model has no statistical significance, indicating mother's age (m.age) and the child's gender (sex) interact in any consequential manner. Thus, we can't reject the null hypothesis which states m.age and sex don't interact. This makes because there's insufficient medical evidence to suggest the age of a mother and her baby's sex are related, hence they wouldn't be related in a survival model such as this.

The Chi-Squared Goodness of Fit test further confirms the interaction of m.age and sex is useless for this survival analysis: the p-value on the goodness of fit test is 0.7445, indicating we can't accept the alternative hypothesis that m.age and sex interact in the model.

I also ran the Breslow and Exact methods for my interactive model. Like with the additive model above, the coefficients and p-values were basically similar, barely differing, but most importantly not being statistically significant.

### Evaluating the models and their results:

The additive model is more informative than the interactive model, both from summarising the models and performing Chi-Squared Goodness of Fit tests on both models. As I discussed above, mother's age (m.age) and the child's gender (sex) are both highly statistically significant for survival - mother's age increased the hazard (i.e., the child dying) by 0.0076%, and being male increased the hazard by over 8.5% relative to being female (the dummy variable's reference category).