# Problem Set 4: Applied Stats II

2022-04-02

## 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 sing mother's age and infant's gender as covariates. Present and interpret the output.

```
# load the package
require(eha)
```

```
## Loading required package: eha
```

```
## Warning: package 'eha' was built under R version 4.1.3
```

```
require(survival)
```

```
## Loading required package: survival
```

```
require(survminer)
```

```
## Loading required package: survminer
```

```
## Warning: package 'survminer' was built under R version 4.1.3
```

```
## Loading required package: ggplot2
```

```
## Loading required package: ggpubr
```

```
## Warning: package 'ggpubr' was built under R version 4.1.1
```

```
require(dplyr)
```

```
## Loading required package: dplyr
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
data("infants")

#glimpse(infants)
```

```r
# create a time variable by subtracting exit and enter
infants <- infants %>% mutate(time = exit - enter)

# Fit survival data a Cox proportional hazards model
cox_model <- coxph(Surv(time, event) ~ age + sex, data = infants)

# summarize the model
summary(cox_model)
```

```
## Call:
## coxph(formula = Surv(time, event) ~ age + sex, data = infants)
##
##   n= 105, number of events= 21
##
##             coef exp(coef) se(coef)      z Pr(>|z|)
## age    -0.03800   0.96271  0.04868 -0.781    0.435
## sexboy -0.47582   0.62138  0.44186 -1.077    0.282
##
##        exp(coef) exp(-coef) lower .95 upper .95
## age       0.9627      1.039    0.8751     1.059
## sexboy    0.6214      1.609    0.2614     1.477
##
## Concordance= 0.591  (se = 0.055 )
## Likelihood ratio test= 1.85  on 2 df,   p=0.4
## Wald test            = 1.87  on 2 df,   p=0.4
## Score (logrank) test = 1.9  on 2 df,   p=0.4
```

The above summary is discussed below:

Statistical significance: From the results, the Wald statitic value in colum 'z' shows tha the variable age and sex are different from 0, therefore we can conclude that age and sex are not statistically significant coefficients.

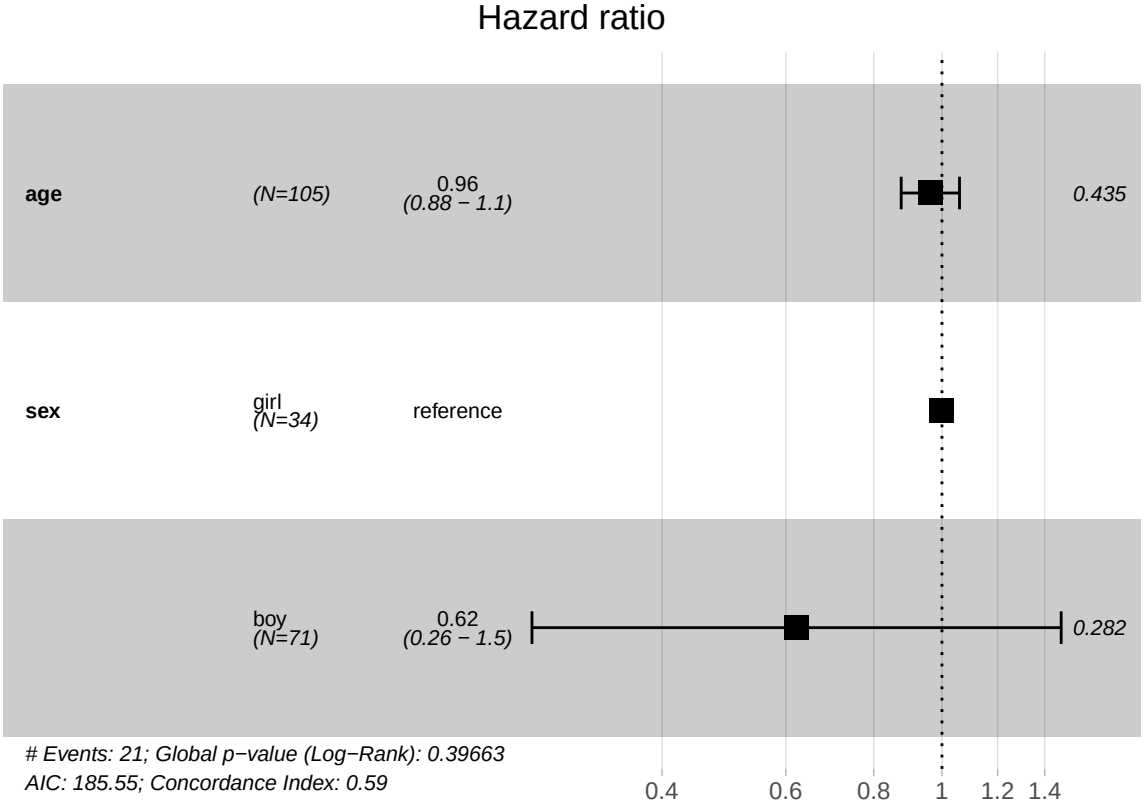
The regression coefficients: The beta coefficients for sexboy = -0.47582 indicate that infants boys have lower risk of death than girls. Also, beta coefficients for age = -0.03800 shows that no matter the age of mother there is lower risk of death of infants.

Hazard ratios: The exponentiated coefficient (exp(-0.47582) = 0.6214) indicates that being an infant boy reduces the hazard by a factor of 0.62. The exponentiated coefficient (exp(-0.03800) = 0.96271) indicates that age reduces the hazard by a factor of 0.96.

Confidence intervals: The upper and lower 95% confidence intervals for hazard ratios (exp(age) and exp(sexboy)), lower 95% bound = 0.8751, upper 95% bound = 1.059 and lower 95% bound = 0.2614, upper 95% bound = 1.477 respectively

Also the above explanation is as shown below by a diagram:

```r
# visualize the results
ggforest(cox_model, data = infants)
```


Hazard ratio
age
(N=105)
0.96
(0.88 − 1.1)
0.435
sex
girl
(N=34)
reference
boy
(N=71)
0.62
(0.26 − 1.5)
0.282
# Events: 21; Global p−value (Log−Rank): 0.39663
AIC: 185.55; Concordance Index: 0.59
0.4
0.6
0.8
1
1.2
1.4