Problem Set 4

Applied Stats II

Due: 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 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.

Importing data and packages

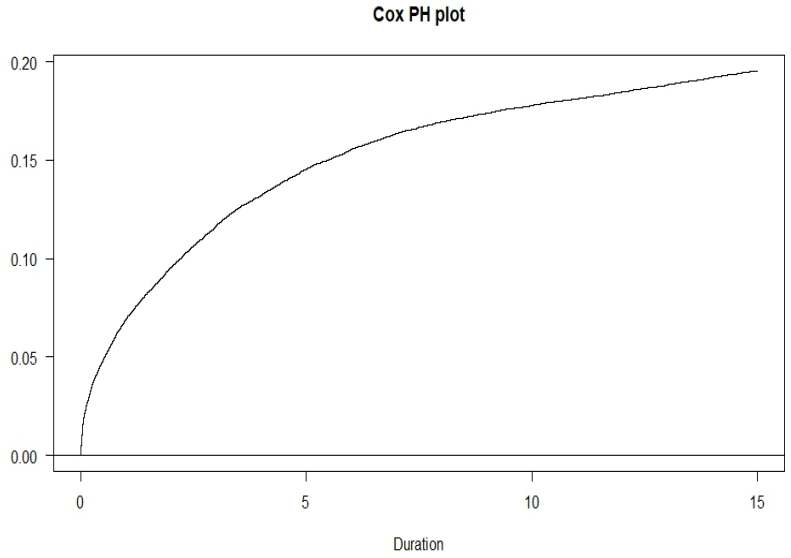
```
library(eha)
library(survival)
data <- child

```

Estimating a Cox PH model

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



> summary(add_surv)
Call:
coxph(formula = Surv(enter, exit, event) ~ m.age + sex, data = data)

n= 26574, number of events= 5616

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

exp(coef) exp(-coef) lower .95 upper .95
m.age         1.0076     0.9924     1.003     1.0119
sexfemale     0.9211     1.0857     0.874     0.9706

Concordance= 0.519  (se = 0.004 )
Likelihood ratio test= 22.52  on 2 df,   p=1e-05
Wald test            = 22.52  on 2 df,   p=1e-05
Score (logrank) test = 22.53  on 2 df,   p=1e-05
```

plotting Cox Pobability Harzard model

```
plot_coxph <- coxreg(Surv(enter, exit, event) ~ m.age + sex, data=data)
plot(plot_coxph)


```



Assessing model fit

```
drop1(add_surv, test = 'Chisq')

Model:
Surv(enter, exit, event) ~ m.age + sex
Df     AIC     LRT  Pr(>Chi)
<none>     113011
m.age    1 113022 12.7946 0.0003476 ***
sex      1 113018  9.4646 0.0020947 **
---

```

Low P-values suggest both explanatory variables are reliable predictors

Interpretation

With an increase in mother's age by 1 unit, there is increase of 1.0076 in the

log hazard rate at some fixed point in time t when all other covariates are held constant at observed means.

With a move up in one level in child's gender (to female) there is an increase of 0.9211 in the log hazard rate at some fixed point in time t when all other covariates are held constant at observed means