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
2 # Using the child dataset
3
4 data(child)
5
6 child_surv <- with(child, Surv(enter, exit, event))
7
8 # fit a Cox Proportional Hazard model
9 cox <- coxph(child_surv ~ sex + m.age, data = child)
10 summary(cox)
11
12 # check fit
13 drop1(cox, test = "Chisq")
14 stargazer(cox, type = "text")
```

Results are as follows:

```
> stargazer(cox, type = "text")

================================================
Dependent variable:
---------------------------
child_surv
------------------------------------------------
sexfemale                      -0.082***
(0.027)

m.age                          0.008***
(0.002)

------------------------------------------------
Observations                     26,574
R2                                0.001
Max. Possible R2                  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

```

```

  > exp(-0.082)
  [1] 0.921272
  > exp(0.008)
  [1] 1.008032

```

Interpretation: There is a 0.082 decrease in the expected log of the hazard for female babies compared to male, holding mother's age constant. There is a 0.008 increase in the expected log of the hazard for babies of mothers as they age by a year, holding sex constant.

By exponentiating the parameter estimates to obtain hazard ratios, we find that the hazard ratio of female babies is 0.92 that of male babies, i.e. female babies are less likely to die (92 female babies die for every 100 male babies; female deaths are 8 per cent lower, etc.)

Presenting the results:

```
cox_fit <- survfit(cox)
autoplot(cox_fit)
```

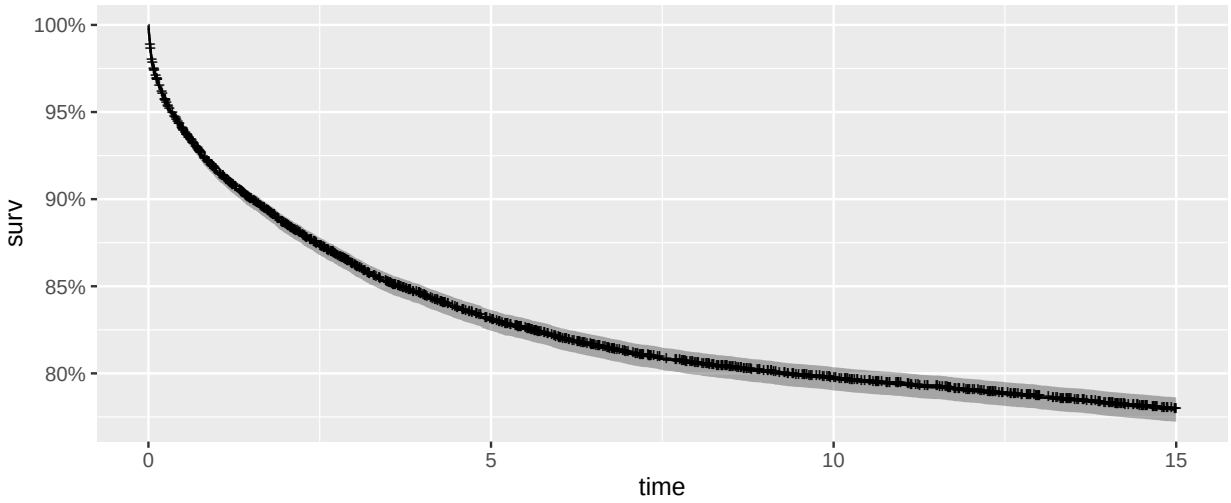

Figure 1: Hazard Function Results

We can also visualise the cumulative hazard function.

```
plot_COXPH <- coxreg(child_surv ~ sex + m.age, data = child)
plot(plot_COXPH)
```

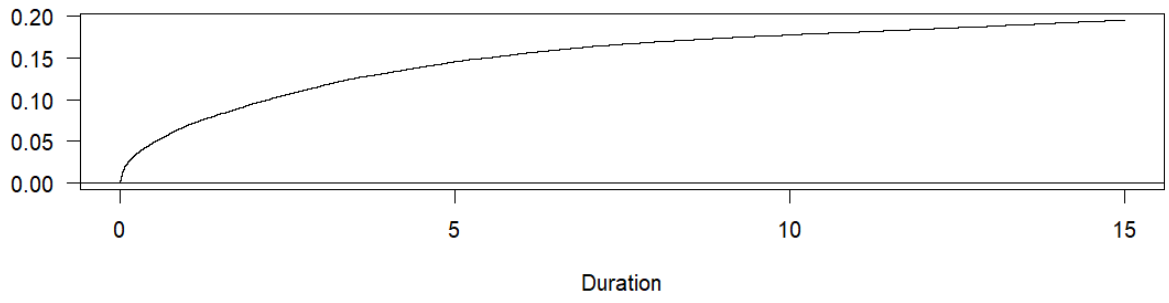

Figure 2: Cumulative Hazard Function

The following plot provides a visualisation of the difference across genders, using an average age for the mother.

```
1
2 newdat <- with(child,
3 data.frame(
4 sex = c("female", "male"), m.age = rep(mean(m.age,na.rm = TRUE, 2))
5 )
6 )
7
8 # using an average age
9 fit <- survfit(cox, newdata = newdat)
10
11 plot(survfit(cox, newdata = newdat), xscale = 1,
12 conf.int = T,
13 ylim = c(0.75, 1),
14 col = c("red", "blue"),
15 xlab = "Time",
16 ylab = "Survival proportion",
17 main = "")
18 legend("bottomleft",
19 legend=c("Male", "Female"),
20 lty = 1,
21 col = c("red", "blue"),
22 text.col = c("red", "blue"))
```

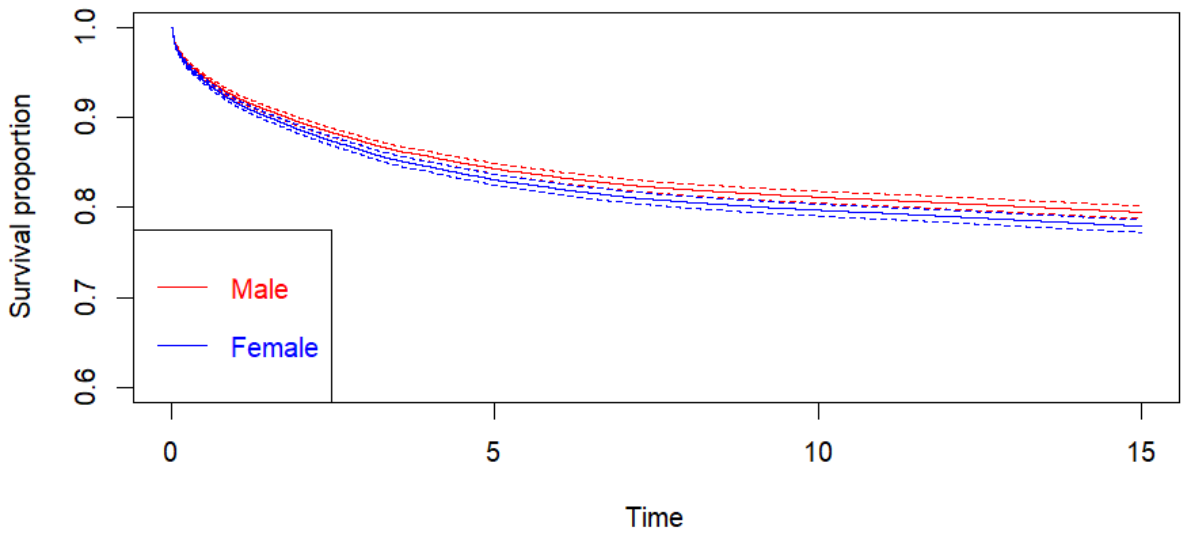

Figure 3: Comparison of male and females using an average age for mothers

The following graph provides a visualisation of the difference in life expectancy for males and female at 20 and 40.

```
newdat20 <- with(child, expand.grid(m.age = c(20, 40), sex = c("male", "female")))

plot(survfit(cox, newdata = newdat),
xscale = 0.30,
conf.int = TRUE,
ylim = c(0.75, 1),
col = c("red", "blue", "green", "black"), # Specify colors for each age group and sex
xlab = "Time",
ylab = "Survival proportion",
main = "")
legend("topright",
legend=c("Male (20)", "Male (40)", "Female (20)", "Female (40)"),
lty = 1,
col = c("red", "blue", "green", "black"),
text.col = c("red", "blue", "green", "black"),
cex = 0.6)
```

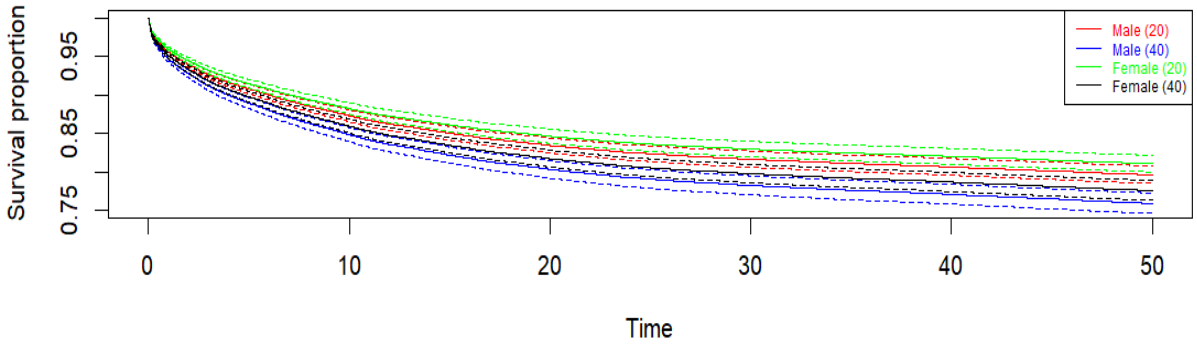

Figure 4: Difference in life expectancy for males and female at 20 and 40