Problem Set 4

Applied Stats II

Due: April 4, 2022

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 class on Monday April 4, 2022. No late assignments will be accepted.
- Total available points for this homework is 80.

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.

Code:

Here is the R code I used to answer this question:

```
# remove objects
rm(list=ls())
# detach all libraries
detachAllPackages <- function() {
  basic.packages <- c("package:stats", "package:graphics", "package:grDevices"
    , "package:utils", "package:datasets", "package:methods", "package:base")
  package.list <- search()[ifelse(unlist(gregexpr("package:", search()))==1,
    TRUE, FALSE)]
```

```
  package.list <- setdiff(package.list, basic.packages)
  if (length(package.list)>0)  for (package in package.list) detach(package,
    character.only=TRUE)
}
detachAllPackages()

# load libraries
pkgTest <- function(pkg){
  new.pkg <- pkg[!(pkg %in% installed.packages()[,  "Package"])]
  if (length(new.pkg))
    install.packages(new.pkg,  dependencies = TRUE)
  sapply(pkg,  require,  character.only = TRUE)
}

#install.packages("Rtools")

install.packages("ggfortify", dependencies = TRUE, INSTALL_opts = '--no-lock')

#Loading necessary packages

lapply(c("survival", "eha", "tidyverse", "ggfortify", "stargazer"),  pkgTest)

data(infants) #calling data w/ eha package


infant_surv <- with(infants, Surv(enter, exit, event)) #creating survival
    object

#creating additive cox ph model with mother and
#sex as covariates
infants_cox <- coxph(infant_surv ~ age + sex, data = infants)

summary(infants_cox) #summary output of model

drop1(infants_cox, test = "Chisq") #assessing model quality
#by using this line to LRT both explanatory variables -
#see Table 2 for output

#Chisq = expected versus observed

#plotting
cox_fit <- survfit(infants_cox)

autoplot(cox_fit, main = "Survival Rate Over Time")
```

Outputs:

The tables and plot I got as outputs from the above code were as follows:

Table 1:

	Dependent variable:
	infant_surv
age	−0.040
	(0.045)
sexboy	−0.485
	(0.442)
Observations	105
R^2	0.019
Max. Possible R^2	0.800
Log Likelihood	−83.626
Wald Test	2.000 (df = 2)
LR Test	1.992 (df = 2)
Score (Logrank) Test	2.034 (df = 2)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2:

Statistic	N	Mean	St. Dev.	Min	Max
Df	2	1.000	0.000	1	1
AIC	3	170.598	0.586	170.124	171.253
LRT	2	1.018	0.208	0.871	1.166
Pr(>Chi)	2	0.315	0.050	0.280	0.351

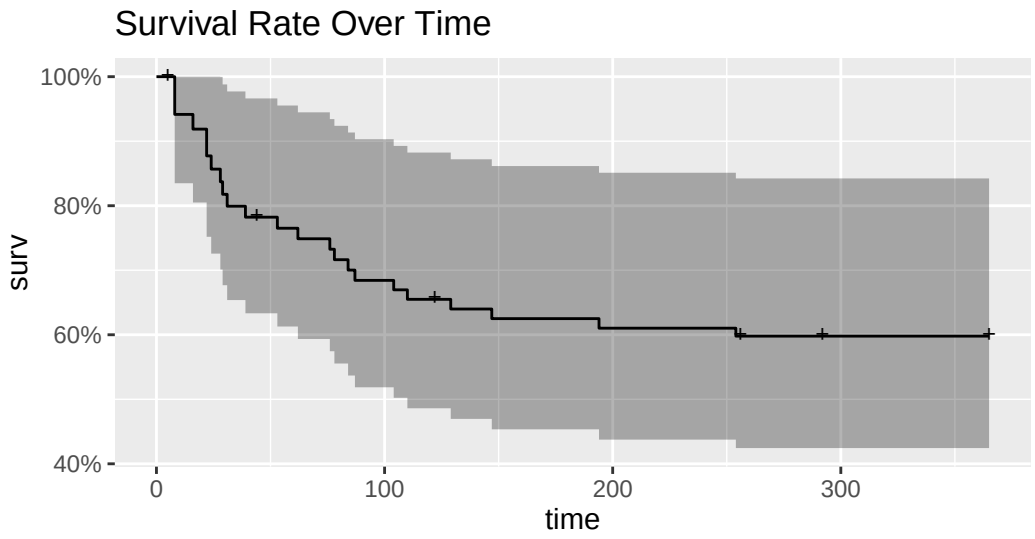

Figure 1:

Interpretation

- There is a 0.485 decrease in the expected log of the hazard for male babies compared to female babies, holding the age of the mother constant.
- There is a 0.040 decrease in the expected log of the hazard for each time the mother's age increases by 1 year/unit, holding the sex of the infants constant.
- None of the coefficients are statistically significant in their p values according to the output table.

Looking at Hazard Ratios:

```
# exponentiate parameter estimates to obtain hazard ratios
exp(coef(infants_cox))

```

Table 3:

age	sexboy
0.960	0.616

- The hazard ratio of male babies is 0.6156972 that of female babies, i.e. 62 male babies die for every 100 female babies; male deaths are 38 per cent lower.

- The hazard ratio for the age of the mother is 0.9603673.
- This means that there is a 4 per cent decrease in the expected hazard for infants relative to a one year increase in mother's age (or the expected hazard is 0.9603673 times higher for a mother who is one year older than another), holding sex constant.