

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 # Tolga Bag - 23371290
3 #####
4
5 #####
6 # load libraries
7 # set wd
8 # clear global .envir
9 #####
10
11 getwd()
12 setwd("C:/Users/tolga/OneDrive/Documents/GitHub/StatsII_Spring2024/problemSets/PS04")
13 #####
14 # Question 1: Model Fit
```

```

15 #####
16
17 # first I need to download the relevant packages
18 install.packages("eha") #I need this to access the child dataset.
19 install.packages("survival") #this will help me compute Cox model.
20 library(eha)
21 library(survival)
22 data(child) #I load the data set. I found information here: https://cran.r-
    project.org/web/packages/eha/eha.pdf
23 child_surv <- with(child, Surv(enter, exit, event)) #I create a survival
    object
24 cox <- coxph(child_surv ~ sex + m.age, data = child) #I fit the cox model as
    gender
25 #and mother's age as covariates.
26 summary(cox)
27 drop1(cox, test = "Chisq")
28 library(stargazer)
29 stargazer(cox, type = "text")
30
31 #####cox#####
32 # Question 1: Presentation and Interpretation
33 #####
34 # There is a 0.08 decrease in the expected log of the hazard for female babies
    compared to
35 # male, holding mother's age constant. There is a 0.008 increase in the
    expected log of the
36 #hazard for babies of mother's with older age compared to younger ones,
    holding gender constant.
37
38 #I exponentiate parameter estimates to obtain hazard ratios
39 exp(-0.007617)
40 # The hazard ratio of older mothers' babies is 0.007 of that of younger babies
    ,
41 cox_fit <- survfit(cox)
42 autoplot(cox_fit) #I plot the survival curve to visualize it.

```

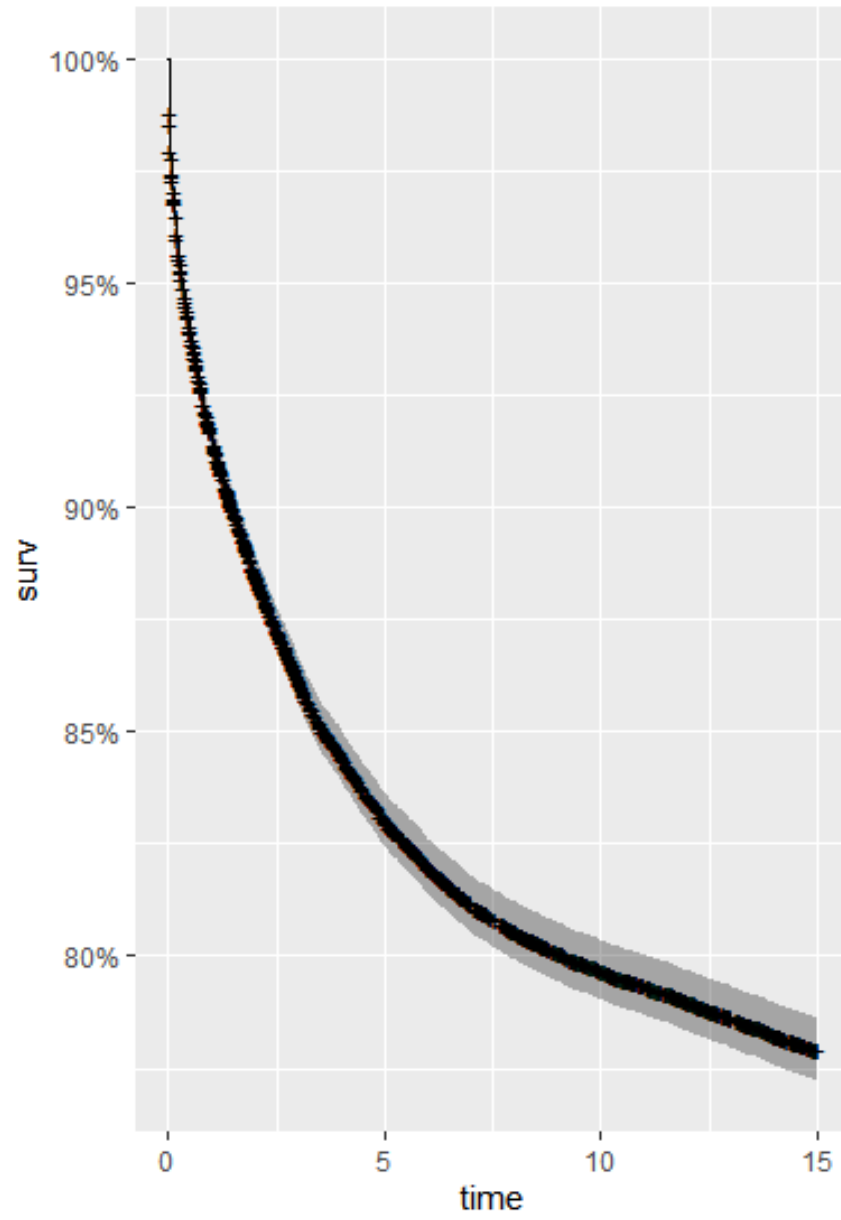


Figure 1: Survival Curve of Cox Model