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.

My Answer - Question 1

Looking at the variables in the imported data(infants)

- enter - Age (in days) of case when its mother died
- exit - Age (in days) at death or right censoring (at age 365 days)
- event - Follow up ends with death (1) or right censoring (0)
- mother - dead for cases, alive for controls
- age - Mother's age at infant's birth

- sex - The infant's sex
- parish - Birth parish, either Nedertornea or not Nedertornea (other)
- civst - Civil status of mother, married or unmarried
- ses - Socio economic status of mothe, either farmer or not farmer (other)
- year - Year of birth of the infant

In order to fit a Cox Proportional Hazard Model the coxph() function from the 'survival' package is required. This needs to be in a Surv() format which represents time to event data. The first argument is time variable.

Testing with Weibull proportional hazard model. Proportional hazards model with parametric baseline hazard(s). Allows for stratification with different scale and shape in each stratum, and left truncated and right censored data.

```
1
2
3 fit <- coxreg(Surv(enter, exit, event) ~ strata(stratum) + mother, data
4               = infants)
5 fit
6 fit.w <- phreg(Surv(enter, exit, event) ~ mother + parish + ses, data =
7                  infants)
8 summary(fit.w) ## Weibull proportional hazards model.
9 plot(fit.w)
10
11 coef(fit.w) # for extracting the coefficients from the model
12
13
14 infants_s <- with(infants, Surv(enter, exit, event))
```

The Kaplan-Meier method is a non parametric statistic that allows you to estimate the survival function.

```
1
2
3 # plotting Kaplan-Meier method
4 kaplan <- survfit(infants_s ~ 1, data = infants)
5 summary(kaplan, times = seq(0, 15, 1))
6 plot(kaplan, main = "Kaplan-Meier Method", xlab = "Days", ylim = c(0.5, 1))
7 autoplot(kaplan)
8
9 # plotting the sex covariates using the infants gender , girl and boy
10 kaplan_sex <- survfit(infants_s ~ sex, data = infants)
11 autoplot(kaplan_sex)
12
13 # running the Cox Proportional Hazard Model, using mothers age and infants
     gender as the covariates
14 cox <- coxph(Surv(enter, exit, event) ~ age + sex, data = infants)
15 summary(cox)
```

Interpretation

```
1
2
3  # there is a .485 decrease in the expected log of the hazard for male babies
       in comparison to female babies,
4  # while holding the age of the moether constant
5
6  # there is .04 decrease in the expected log of of the hazrd each time the
       mothers age increases by 1 year or 1 unit increase
7  # while holding the sex of infant constant
8
9  # coef    exp(coef) se(coef)                  z Pr(>|z|)
10 # age     -0.04044   0.96037  0.04507 -0.897    0.370
11 # sexboy  -0.48518   0.61559  0.44224 -1.097    0.273
12
13
14 ### hazard ratio for male babies is .61 compared to that of female babies or
       62 male babies die for every 100 female babies
15 exp(coef(cox))
16 # age       sexboy
17 # 0.9603673 0.6155879
18
19 ########################################## Dependent variable  is : infant_s
20 # testing the out of the coefficients
21 coef(cox_s)
22 # age       sexboy
23 # -0.04043946 -0.48517752
24
25 exp(coef(cox_s))
26 # age     sexboy
27 # 0.9603673 0.6155879
28
29 cox.w <- phreg(Surv(enter, exit, event) ~ age + sex, data = infants)
30 summary(cox.w)
31 plot(cox.w)
32
33 # Covariate             Mean       Coef     Rel.Risk   S.E.    LR p
34 # age                  27.127    -0.050     0.951     0.045   0.2376
35 # sex                                                        0.4029
36 # girl                0.317     0          1 (reference)
37 # boy                 0.683    -0.375     0.687     0.444
38
39 # Events                    21
40 # Total time at risk         21616
41 # Max. log. likelihood      -154.85
42 # L3 R test statistic         2.10
```

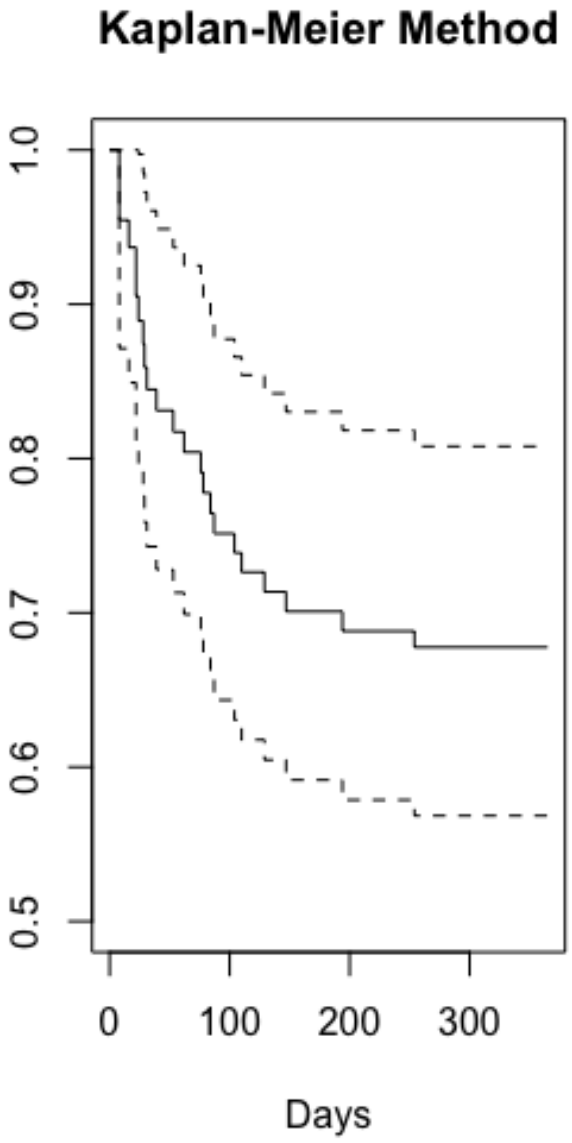

Figure 1: Kaplan-Meier Method

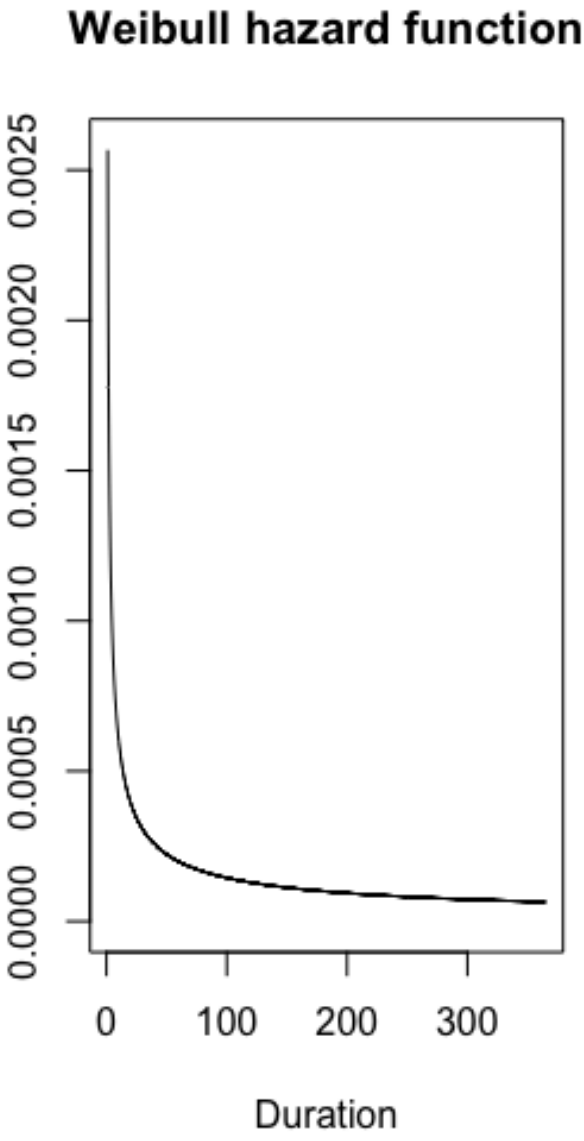

Figure 2: Weibull hazard function

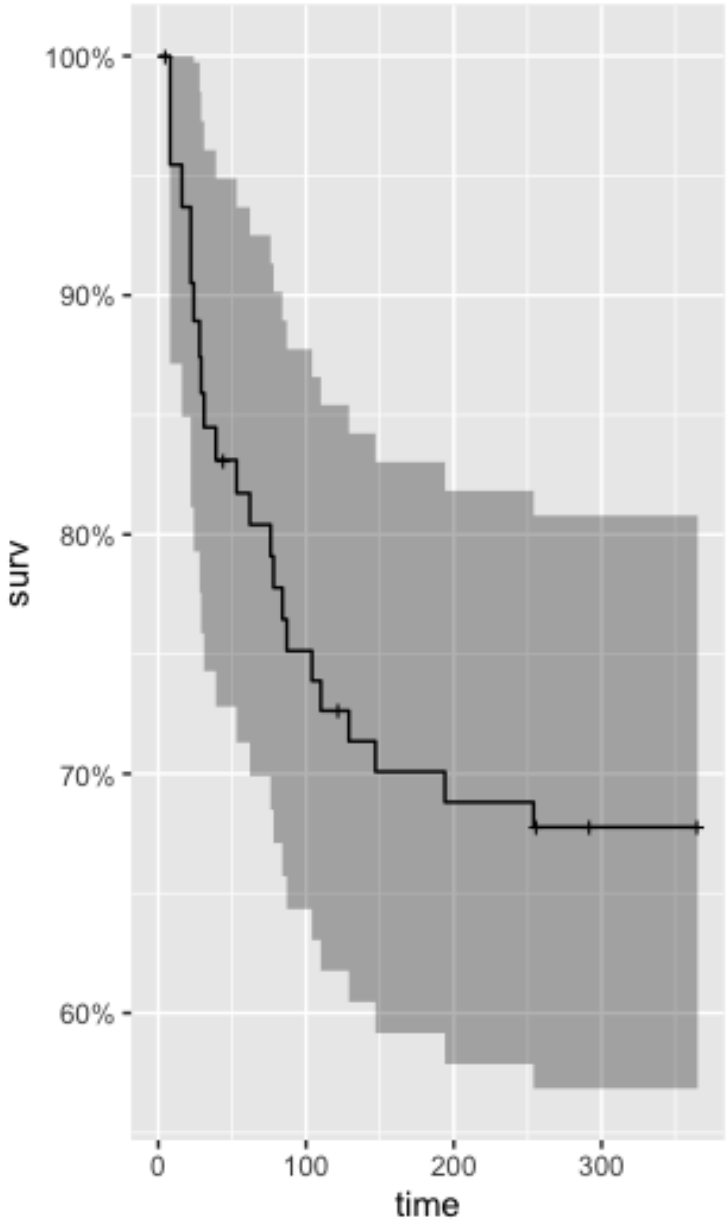

Figure 3: Kaplan-Meir Plot

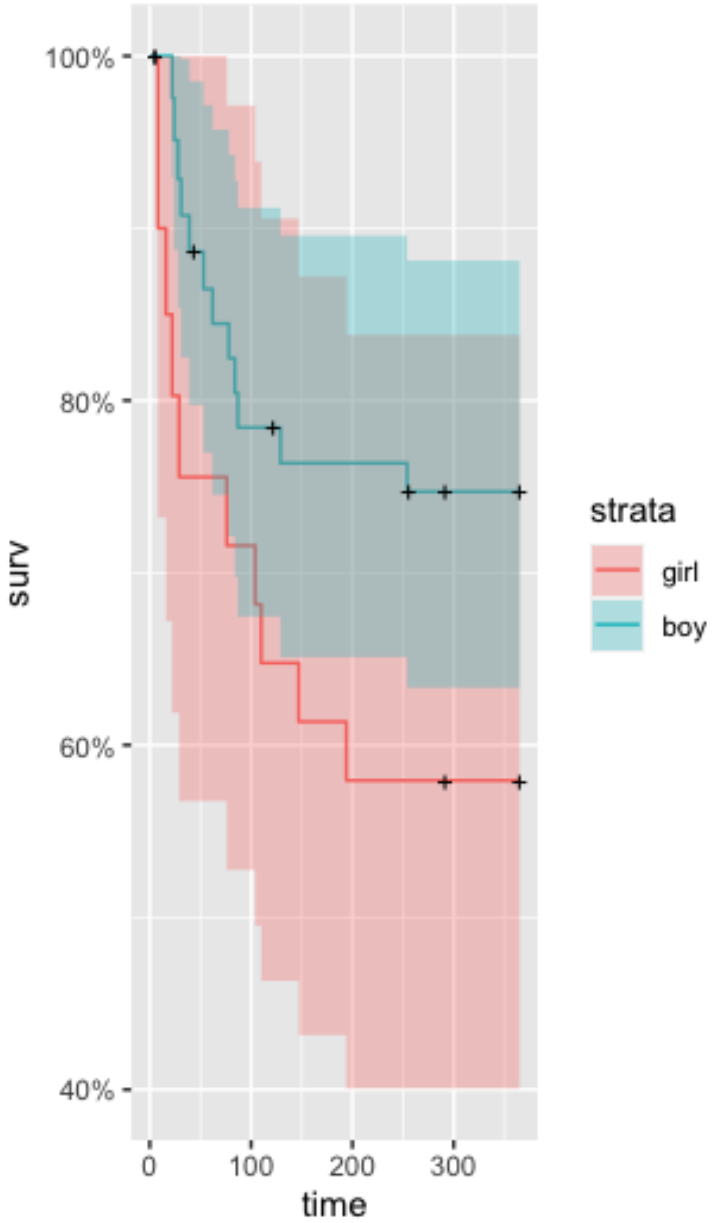

Figure 4: Kaplan-Meir Plot