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.

Codes as below:

```
1  # Review data
2  data(child)
3  head(child)
4  summary(child)
5
6  # Create a survival object
7  child_surv <- with(child, Surv(enter, exit, event))
8
9  # Fit the Cox Proportional Hazards model
10 cox_mod <- coxph(child_surv ~ m.age + sex, data = child)
11 summary(cox_mod)
```

Summary output:

<div align="center">

Table 1:

	Dependent variable:
	child_surv
m.age	0.008***
	(0.002)
sexfemale	−0.082***
	(0.027)
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

</div>

Interpretation:
1. m.age (mother's age): The coefficient is 0.008 (***). This indicates that for every one-year increase in mother's age, the risk of infant survival increases by a factor of 1.008 on average. The three asterisks denote that this result is highly statistically significant ($p < 0.01$).
2. sexfemale (infant's gender): The coefficient is -0.082. This suggests that compared to male infants, the survival risk of female infants decreases by a factor of 0.082 on average, or in other words. Similarly, the three asterisks denote that this result is highly statistically significant ($p < 0.01$).

In summary, increasing mother's age is associated with an increase in infant survival risk, while female infants have a higher survival risk compared to male infants.
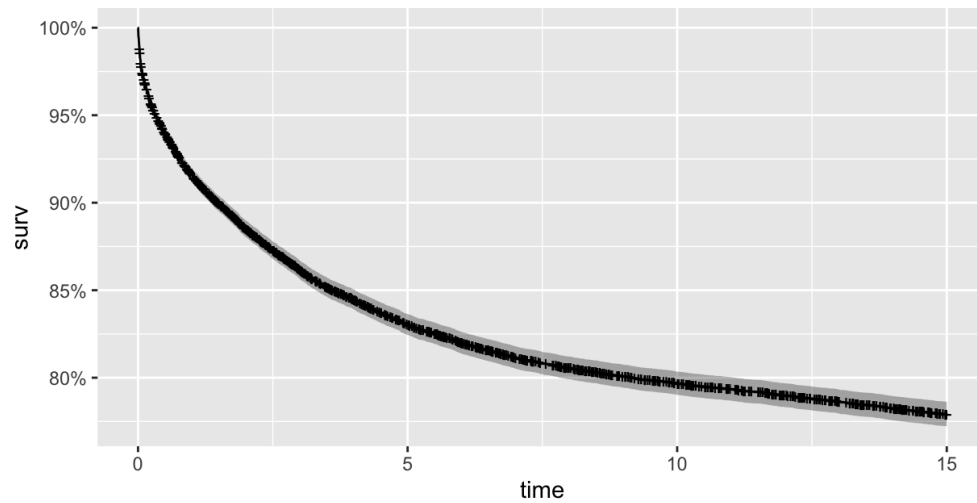
Model Evaluation and Visualization:

```
1  cox_drop <- drop1(cox_mod, test = "Chisq")
2  stargazer(cox_drop, type = 'latex')
3
4  cox_fit <- survfit(cox_mod)
5  autoplot(cox_fit)
```

Output:

Table 2:

Statistic	N	Mean	St. Dev.	Min	Max
Df	2	1.000	0.000	1	1
AIC	3	113,017.100	5.528	113,011.000	113,021.800
LRT	2	11.130	2.355	9.465	12.795
Pr(>Chi)	2	0.001	0.001	0.0003	0.002

Interpretation:

In the assessment results of the Cox Proportional Hazard model, the statistical significance p-values of the model are all less than 0.05, indicating that the variables in the model have a significant impact on survival time.

The survival curve plot displays estimated survival probabilities at specific time points. For example, at the 15-year timestamp, the survival probability drops to approximately 75