

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.

Answer:

The following code was used to load the data and produce the model:

```
1 # Load data.
2 data(package = "eha")
3 data("child")
4
5 # Fit Cox Proportional Hazard model
6 add_surv <- coxph(Surv(enter, exit, event) ~ m.age + sex,
7 data = child)
```

Table 1: Table of Additive Cox Proportional Hazard Model with Hazard Ratios:

	<i>Dependent variable:</i>
	enter
m.age	0.007617*** (0.002128)
exp(m.age)	1.007646***
sexfemale	−0.082215*** (0.026743)
exp(sexfemale)	0.921074***
Observations	26,574
R ²	0.001
Max. Possible R ²	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)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Interpretation of Coefficients:

m.age: The estimated coefficient for mother's age is 0.007617, with an exponentiated coefficient (hazard ratio) of 1.007646. This means that with each additional year of age, the hazard (risk of the event occurring) increases by approximately 0.76%. The p-value (0.000344) suggests that this effect is statistically significant at $\alpha = 0.05$. The older the mother the greater the probability of the child dying.

sex: The estimated coefficient for sexfemale is -0.082215, with a hazard ratio of 0.921074. This indicates that the hazard for females is approximately 7.9% lower compared to males (because reference group male is coded as 0). So a one unit increase in sex is equal to changing sex from male to female. The p-value (0.0021) shows that this is statistically significant at $\alpha = 0.05$. Male children have a higher probability of dying than female children.

Assessing Model Quality:

The following code was used to test the quality of the model fit.

```
1 assessment <- drop1(add_surv, test = "Chisq")
```

Below are the results of the drop1 assessment.

Table 2: Model Assessment for Cox Proportional Hazards Model

Term	Df	AIC	LRT	Pr(>Chi)
<none>		113011		
m.age	1	113022	12.7946	0.0003476 ***
sex	1	113018	9.4646	0.0020947 **

The low p-values ($< \alpha = 0.05$) suggest that dropping either of the predictors would result in a worse model fit. The likelihood ratio test results also suggest that both the predictors are significant and result in a better model fit than a model without them.

Also, fitting a model with an interaction effect in R resulted in a worse model with lower p-values, the interaction effect was shown to be statistically insignificant, indicating that the additive model is better.

Bibliography:

1. Dr. Jeffrey Ziegler's Lecture material.