Problem Set 4

Applied Stats II
Dan Zhang 23335541

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.

Load data and check data structure.

```
data(child)
```

Table 1: Structure of Child Data

id	m.id	sex	socBranch	birthdate	enter	exit	event	illeg	m.age
3	246606	male	farming	1853-05-23	0	15.000	0	no	35.009
42	377744	male	farming	1853-07-19	0	15.000	0	no	30.609
47	118277	male	worker	1861-11-17	0	15.000	0	no	29.320
54	715337	male	farming	1872-11-16	0	15.000	0	no	41.183
78	978617	female	worker	1855-07-19	0	0.559	1	no	42.138
102	282943	male	farming	1855-09-29	0	0.315	1	no	32.931

Use Surv() function to create a survive object and run the Cox model.

```
# Create a survival object using the Surv() function
child_surv <- with(child, Surv(enter, exit, event))
# Fit the Cox model
cox <- coxph(child_surv ~ m.age + sex, data = child)
```

Table 2: The impact of maternal age and child's sex on child survival

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

Note: *p<0.1; **p<0.05; ***p<0.01

Before interpreting the results, let's run a drop1() function to check covariables' contribution to the model.

```
drop1(cox, test = "Chisq")
```

Table 3: Single Term Deletions from the Model Predicting Child Survival

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

Note: Significance codes: *** p<0.001, ** p<0.01, * p<0.05, · p<0.1

The model fits well. Have a look at the survival curve.

```
# Fit the survival curve
cox_fit <- survfit(cox)
# Plot the survival curve
pdf("Survial curve.pdf")
autoplot(cox_fit)
dev.off()
```

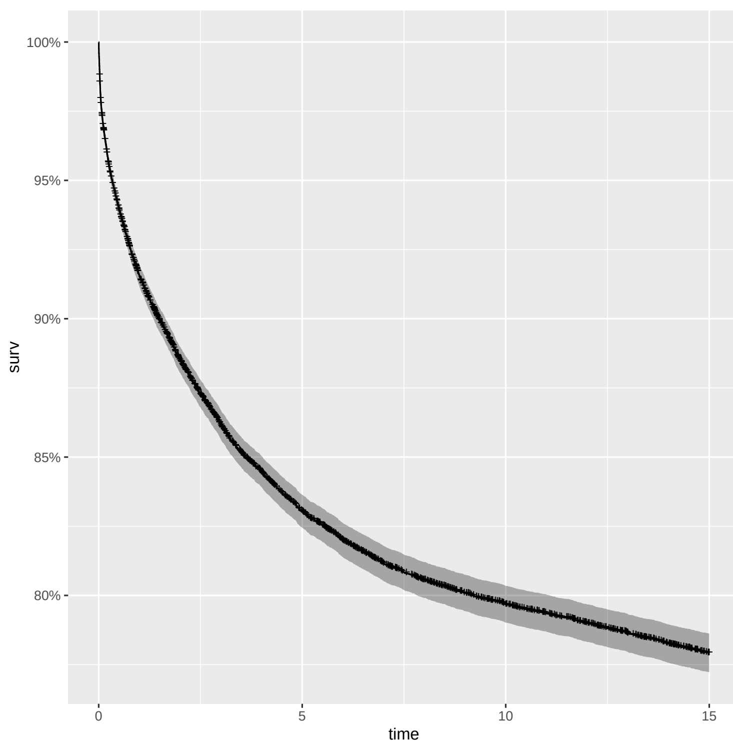

Figure 1: Survial curve

Calculate the exponential results of covariables to get the hazard ratio.

```
hrsex <- exp(-0.082215)
```

Interpretation:

- Coefficient of mother's age: The mother's age has a statistical significant positive impact on the infant's survival rate, with a coefficient of 0.008. For every one-year increase in maternal age, the log-hazard ratio of the infant's survival probability increases by an average of 0.008, holding other variables constant. The hazard ratio of female infant is 0.92 that of male infant, which means female infant are less likely to die, female deaths are 8% lower.

- Coefficient of infant's gender: Compared with the reference group (male infant), the logarithm ratio of the survival rate of female children is reduced by 0.082 on average, holding other variables constant.

In conclusion,
Run a interaction term with the covairables and check the model fit.

```
# Fit the model
cox.int <- coxph(child_surv ~ m.age*sex, data = child)
summary(cox.int)

# Test the model
drop1(cox.int, test = "Chisq")
```

Table 4: Single Term Deletions for interaction term model

Term	Df	AIC	LRT	Pr(>Chi)
None	-	113013	-	-
m.age:sex	1	113011	0.10623	0.7445

As $p > 0.05$. This means that adding the m.age:sex interaction term did not bring statistically significant model improvements.