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.

Table 1: Child Mortality

	Dependent variable:
	enter
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

When controlling for gender, each unit increase in the child's age increases the log of the risk of infant mortality occurring by an average of 0.007. The exponentiated coefficient suggests the children are at 0.76% risk per mortality per year of age. When controlling for age, being female, as compared to male, reduces the log of the risk of infant mortality occurring by an average of -0.082215. The exponentiated coefficient for gender suggests female children have 0.397 times the risk of mortality as male children. Each of the coefficients are significant (p <.01), meaning we can reject the null hypotheses that neither gender or age are associated with child mortality. The likelihood ratio test also suggests the model explains a non-random level of variance in child mortality (p <.01). However R2 of .001 suggests in practice the model explains less than 1% of variation in child mortality.