Assignment 1 - The Titanic - Discrete Choice Modelling

Marc Sparhuber

Table of contents

Task A

The passengers on Titanic embarked from different ports. Create a new variable called Family, that should be a dummy variable (i.e. that takes the value 1 or 0) that refers to if the passenger had any spouse or sibling on board the ship. Make a table with descriptive statistics for the variables: Sex, Age, Survived, Pclass, and your new variable Family. The continuous variables and the dummy variables (with 0 or 1) should have information on min, max, mean, SD, and number of observations (one could question the value of SD of a dummy variable, but include it anyway). The categorical variables should have information on number of observations and the distribution over the categories in percent. Describe the descriptives statistics in the table as you would do in a paper with full paragraphs. Use clear and concise English.

Table 1: Passengers on the Titanic: Descriptive statistics

		Mean	SD	N	Max	Min	Percent
	Age	29.70	14.53	714	80.00	0.42	100.00
	Survived	0.38	0.49	891	1.00	0.00	100.00
	Family	0.32	0.47	891	1.00	0.00	100.00
Pclass	1			216			24.24
	2			184			20.65
	3			491			55.11
Sex	male			577			64.76
	female			314			35.24

Comments: Data from the Titanic R package.

Task B

Let us say that we want to study if the old phrase "women and children first" accurately describes the event on Titanic, or rather if women and younger persons were more likely to survive. Create a new dummy variable, female, that indicates if a passenger is female (=1) or not (=0). Estimate a linear probability model with survival as the dependent variable and age and your new variable female as independent variables. Estimate a second model where you also include passenger class and your new variable Family as independent variables. Produce one table that includes both of the two estimated models. Describe the estimated models and your findings, including interpretation of the results (including at least one measurement of the model fit), in a way that is understandable for someone who is not familiar with the data, i.e. as you would do in an article with full paragraphs. Use clear and concise English.

Table 2: Survival from Titanic. Linear probability models

	Model 1	Model 2
(Intercept)	0.234***	0.671***
	(0.038)	(0.057)
female	0.547***	0.485***
	(0.032)	(0.031)
Age	−0.001	−0.006***
	(0.001)	(0.001)
Pclass2		−0.211***
		(0.042)
Pclass3		−0.414***
		(0.039)
Family		−0.044
		(0.031)
Num.Obs.	714	714
R2	0.291	0.392
R2 Adj.	0.289	0.388

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Titanic R package.

Comments: Pclass refers to passenger class, first class is the reference category.