Lab Assignment 1

Yan Gao

2023-01-26

Task A

#1. The passengers on Titanic embarked from different ports. Create a new variable, Southampton, that refers to if they embarked from Southampton (=1) or not (=0).

#2. Make a table with descriptive statistics for the variables: Sex, Age, Survived, Pclass, and your new variable Southampton. 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 table as you would do in a paper. Use clear and concise English.

#In this table, we can see there are 891 people embarked from Southampton, ranges from infant to elderly.The oldest passenger was 80 years old and the average of survived rate is 38%. In addition,the mean age of passengers embarked from Southampton is 29.7 with 14.53 standard deviation.The female passengers accounted around for 35.24% and male passengers accounted for 64.76% respectively. The largest number of Passengers were from the 3rd class with 491 people accounted for 55% of total number of passengers, and the second large embarked people were from first class with 216 people accounted for 24.24% in total passengers.   Most of the passengers (72%) boarded from Southampton, there are also a large number of missing values for age.

#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.

#1.Create a new dummy variable, female, that indicates if a passenger is female (=1) or not (=0).

#2. Estimate a linear probability model with survival as dependent and age and your new variable female as independent variables.

Table 1: Passengers embarked from Southampton: Descriptive statistics

		Mean	SD	Min	Max	N	Percent
	Age	29.70	14.53	0.42	80.00	714	100.00
	Survived	0.38	0.49	0.00	1.00	891	100.00
	Southampton	0.72	0.45	0.00	1.00	891	100.00
Passenger class	1					216	24.24
	2					184	20.65
	3					491	55.11
Sex	female					314	35.24
	male					577	64.76
	All					891	100.00

Comments: Data from the Titanic R package.

	Model A	Model B
(Intercept)	0.234	0.678
	(0.038)	(0.058)
Age	−0.001	−0.005
	(0.001)	(0.001)
female	0.547	0.475
	(0.032)	(0.031)
Pclass2	(omitted first class)	−0.189
		(0.043)
Pclass3		−0.393
		(0.039)
Southampton		−0.057
		(0.036)
Num.Obs.	714	714
R2	0.291	0.392
R2 Adj.	0.289	0.388
AIC	773.2	669.2
BIC	791.5	701.2
Log.Lik.	−382.608	−327.594
F	145.958	91.415
RMSE	0.41	0.38

year of age associated with a 0.5% lower probability to survival.

a 10 years old children

R2 adjusted from 0.289 to 0.388, which increased.

#3.Estimate a second model where you also include passenger class and the variable Southampton as an independent variables.

#4.Produce one table that includes both of the two estimated models.

#5. 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. Use clear and concise English.

	Model A	Model B
(Intercept)	0.23***	0.68***
	(0.04)	(0.06)
Age	0.00	−0.01***
	(0.00)	(0.00)
female	0.55***	0.48***
	(0.03)	(0.03)
Pclass2		−0.19***
		(0.04)
Pclass3		−0.39***
		(0.04)
Southampton		−0.06
		(0.04)
Num.Obs.	714	714
R2	0.291	0.392
R2 Adj.	0.289	0.388
AIC	773.2	669.2
BIC	791.5	701.2

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

Source: Titanic R package.

We use survival as dependent variable, age and female as independent variables to test whether women and child were more likely to be rescued in the Titanic. Regression analyses reported in model A confirm the old phrase "women and children first"; e.g., a 1 SD(0.03) increase in survived people is associated with a 55 per cent increase survival probability of female. The age as the independent variable we control at model A shows insignificant related. When controlling for prestige in Model B, the passenger class and Southampton as independent variables are contoled in this model, the age is significant and negative for survived people, which could means that age is significant consideration but compare with other variable such as gender, it may not account for much. The gender variable still significant when controlling for pclass and Southampton but decreased to 48 per cent.The passenger class variables indicated that the probability of women and children survived from second class decreased but still got rescued about 19 per cent and 39 per cent from the third class,e.g.,1 SD(0.04) increase in survived people is associated with a 19 per cent decrease survival probability of female passengers from sencond class and 39 per cent decreased from third class.

In addition, Southampton shows negative and insignificant for survived people. After adjust R2, the figure turn small, so there may have other confounders affect the probability of women and child survived so there

have room for us to create the model c to observe the result.

Both model A and B indicated that female and younger person were more likely to survied, especially the women, had more chance to survived. The age variable at model A shows insignificant and shows decreased trend at model B probably because at that sitution, rescue plan considered women first, no matter what age they are.