

# Lab\_Assignment\_1\_by\_Yiqiu\_Wang

Yiqiu Wang

2023-01-25

## A Table 1

Table 1 reports the descriptive statistics on the passengers on the Titanic.

There were a total of 891 passengers on Titanic of which 35.24%(314) were female and 64.76%(577) were male.

Due to missing data, the age of 714 passengers were accessible. The recorded ages ranged from 0.42 to 80. The average age of the 714 passengers was 29.70. The standard deviation is smaller than the mean which indicates that the ages of many passengers on Titanic were not spread too far from 30 (the ages are clustered near the mean).

38% of all passengers survived the ship crash.

24.24%(216) of all 891 passengers were of the first class, 20.65%(184) of all passengers were of the second class and 55.11%(491) of all passengers were of the third class.

The majority(72%) of all passengers embarked from Southampton.

## B Table 2

We ran two linear probability models to estimate the expected survival probability on Titanic.

The results of Model A shows that the expected likelihood of survival for a 0-year-old, male passenger was 23%. A passenger's age had no significant effect on his/her probability of survival. Being a female significantly increased a passenger's probability of survival by .55, which means that regardless of her age, a female passenger's expected likelihood to survive the ship crash was 78%. Women's expected probability of survival was about 3 times as high as that of men.

In conclusion, female passengers were indeed much more likely to survive than male passengers. However, children and adults of the same gender had the same probability of survival. So sadly, children were not benefiting from "women and children first".

As we add two more variables, passenger class and embarkation location, to the model, the results of Model B appears to be different from those of Model A.

The results of Model B shows that the expected probability of survival for a 0-year-old, male, first class passenger was 68%. A one-year increase in age lowered this probability by .01, holding all other variables constant, which means children were more advantaged than adults. Being a female increased this probability of survival by .48, holding all other variables constant, which means women were more advantaged than men. However, being of a second class lowered a passenger's chance to survive by .19 and being of a third class lowered the chance by .39, holding all other variables constant. A passenger's embarkation location had a negative but insignificant effect on the survival likelihood so a passenger would not benefit or suffer from whether he/she embarked from Southampton or not.

Table 1: Passengers on the Titanic: Descriptive Statistics

		Min	Max	Mean	SD	N	Percent
	Age	0.42	80.00	29.70	14.53	714	100.00
	Survived	0.00	1.00	0.38	0.49	891	100.00
	Southampton	0.00	1.00	0.72	0.45	891	100.00
Pclass	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

Source: Titanic R package.

Comments: Pclass refers to passenger class with 1 for the first class, 2 for the second class and 3 for the third class. Southampton indicates whether a passenger embarked from Southampton(=1 and 0 otehrwise).

Table 2: Survival from Titanic. Linear probability models

	Model A	Model B
(Intercept)	0.23*** (0.04)	0.68*** (0.06)
Age	0.00 (0.00)	-0.01*** (0.00)
female	0.55*** (0.03)	0.48*** (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
Log.Lik.	-382.608	-327.594
F	145.958	91.415
RMSE	0.41	0.38

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

Source: Titanic R package.

Comments: Age equals to 0 is the reference category. female refers to the gender of a passenger which equals to 1 for female and 0 for male, male is the reference category. Pclass refers to passenger class with 1 for the first class, 2 for the second class and 3 for the third class, first class is the reference category. Southampton indicates whether a passenger embarked from Southampton(=1 and 0 otherwise).

A woman of the first class was at least 19% more likely to survive than another woman of the second or third class who was as old as her. A 20-year-old, first class male and a 29-year-old, third class female had the same probability of survival. So just being one year younger than 20 could make this man more likely to survive than that woman.

In conclusion, only if the passengers were of the same class in Model B can we say that children were indeed more likely to survive than adults and women were more likely to survive than men so a 0-year-old girl (if there should be one) was expected to enjoy the highest probability of survival (though the probability is over 1). This did not happen in reality. Passenger class could possibly play a greater role than age and gender on the survival probability, especially when the advantage of being a female was reduced to some extent because of age. To our disappointment, passenger class served as an obstacle when people were trying to save as many women and children as possible after saying “women and children first”.

R-squared increases from 0.291 in Model A to 0.392 in Model B, but this does not necessarily indicate a better model fit as the model fit could look better only because we are including more variables.

The adjusted R-squared increases from 0.289 in Model A to 0.388 in Model B. This indicates a better model fit because the adjusted R-squared would not increase when the variables in the model does not provide the model fit by a sufficient amount. So Model A is biased because we neglected the variable Pclass which had significant effect on the probability of survival.