

# The Titanic - A disaster reflecting social inequality

Damien Dupré

## Introduction

This document is an example of what can be done using Rmarkdown file to produce an academic/professional report. It focuses on the titanic dataset which is one of the most classic dataset to practice data analytic and data science. Indeed the catastrophe of the titanic is not only interesting from a data point of view but also from an human point of view.

While the James Cameron movie did not focus too much on the influence of classes on the final outcome, it appears that passengers were not equals in age and gender. For these reasons, this report will investigate what are the influence of age and gender on passengers' ticket fare.

## Literature Review

According to Saidatunnisa, Sili and Nasrullah (2019), the inequality issues towards women existed in one of the most outstanding movies of all times, Titanic. Titanic movie began with the 84 years old woman named Rose DeWitt Bukater who told about Titanic's tragic story on its first and last voyage April 1912. Rose DeWitt Bukater was regardless of her status as the upper class, also a woman who experienced the inequality issues. Rose as a woman had to be obedient towards the male supremacy society she lived in.

In addition, Researchers interested in inequalities in healthcare wont to quote data from the Titanic disaster" (Marmot 1999, p. 16). This ritual stems from Antonovsky (1967), who opened his classic paper with an illustration based on data from the Titanic, highlighting the higher survival rate of the first- and second-class passengers, versus the second- and third-class passengers, to demonstrate that social class influences mortality (Pearson, 2008).

- **Hypothesis 1:** The average tickets fare of female passengers is higher than the average ticket fare of male passengers.
- **Hypothesis 2:** Passengers' tickets fare increase when their age increase.

## Method

The method section usually describes the participants where the data was collected, the material and the procedure used to collect the data, and the analysis performed to test the hypotheses.

### Participants

The data were collected from 1309 passengers known for having embarked on the Titanic. The average age of passengers was 29.9 years old ( $SD = 14.4$ ) but the age from 263 passengers is unknown.

n	missing	m_age	sd_age
1309	263	29.88114	14.41349

From them, 466 where female passengers and 843 were male passengers.

Sex	n
female	466
male	843

## Material

Little is know over how these data were collected but it lists the passengers name, age, gender, ticket class, number of siblings / spouses aboard the Titanic, number of parents / children aboard the Titanic, ticket number, passenger fare, cabin number, port of Embarkation (C = Cherbourg, Q = Queenstown, S = Southampton) and ultimately their survival (0 = No, 1 = Yes).

## Procedure

This section usually describes how the data were collected but here they were found online on the Kaggle website <https://www.kaggle.com/c/titanic/data>.

## Data Analyses

To test these hypotheses, the following linear regression model will be used:

$$Fare = \beta_0 + \beta_1 Gender + \beta_2 Age + \epsilon$$

This model uses Fare as a continuous outcome variable. Age is a continuous predictor and Gender a categorical predictor with two categories (i.e., female and male). Age and Gender are tested in the model as a main effect as indicated in Figure 1.

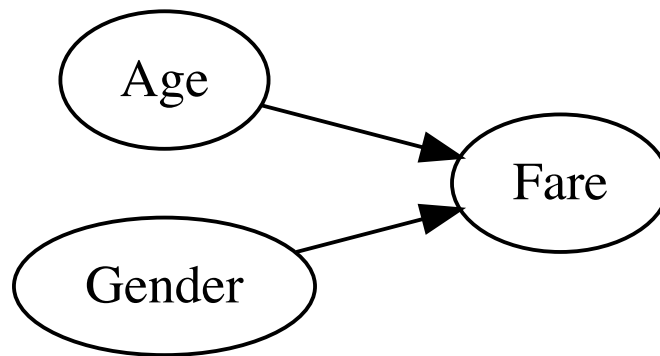


Figure 1: Representation of the model tested.

## Results

The average fare for female passengers is slightly higher than the average fare for male passengers (£46.2 for female passengers vs. £26.2 for male passengers) as show by the Figure 2.

It also appears that the ticket fare increases when age increases as well as shown in Figure 3.

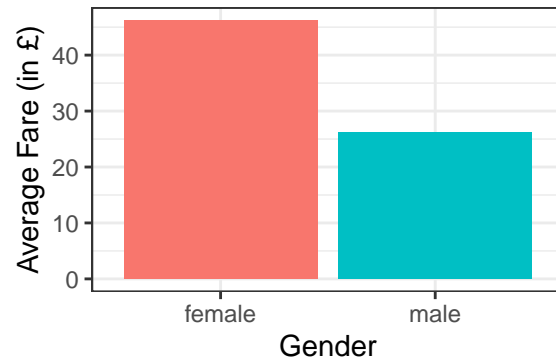


Figure 2: Average ticket fare for female and for male passengers.

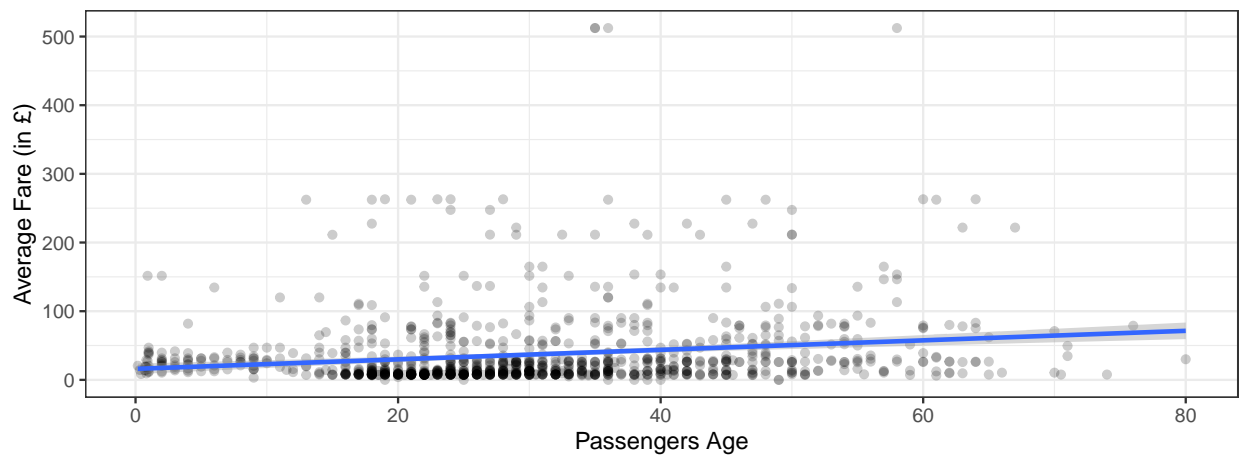


Figure 3: Relationship between passenger's age and ticket fare.

We fitted a linear model (estimated using OLS) to predict Fare with Age and Sex (formula:  $\text{Fare} \sim \text{Age} + \text{Sex}$ ). The model explains a statistically significant and weak proportion of variance ( $R^2 = 0.07$ ,  $F(2, 1042) = 40.26$ ,  $p < .001$ , adj.  $R^2 = 0.07$ ). The model's intercept, corresponding to Age = 0 and Sex = , is at 29.07 (95% CI [20.63, 37.50],  $t(1042) = 6.76$ ,  $p < .001$ ). Within this model:

- The effect of Age is statistically significant and positive ( $\beta = 0.74$ , 95% CI [0.51, 0.97],  $t(1042) = 6.39$ ,  $p < .001$ ; Std.  $\beta = 0.19$ , 95% CI [0.13, 0.25])
- The effect of Sex [male] is statistically significant and negative ( $\beta = -23.04$ , 95% CI [-29.81, -16.27],  $t(1042) = -6.68$ ,  $p < .001$ ; Std.  $\beta = -0.41$ , 95% CI [-0.53, -0.29])

Standardized parameters were obtained by fitting the model on a standardized version of the dataset.

## Discussion and Conclusion

An obvious reason why age of passengers influenced their ticket fare is due to negotiated rates for children being cheaper than adults. There is also the fact that older passengers prefer more comfortable cabins and therefore will pay a higher price. However, the reason why female passengers paid a higher price on average is still a mystery. Probably, male passengers were satisfied with more simple cabins. However, female passengers were usually travelling with their children and therefore required a larger cabin. This explanation provide supports for a latent gender inequality in western society. Female have much more side expenses than males, this phenomenon is known as pink tax (Lafferty, 2019) or tampon tax (Bennett, 2017).

## References

- Antonovsky, A. (1967). Social class, life expectancy, and overall mortality. *Milbank Quarterly*, 45(2), 31–73.
- Bennett, J. (2017). The tampon tax: Sales tax, menstrual hygiene products, and necessity exemptions. *Business Entrepreneurship & Tax Law Review*, 1, 183.
- Lafferty, M. (2019). The pink tax: The persistence of gender price disparity. *Midwest Journal of Undergrad Research*, 11, 56–72.
- Marmot, M. G. (1999). Epidemiology of socioeconomic status and health: Are determinants within countries the same as between countries? *Annals of the New York Academy of Sciences*, 896, 16–29.
- Pearson, J. A. (2008). Can't buy me whiteness: New lessons from the Titanic on race, ethnicity, and health. *Du Bois Review: Social Science Research on Race*, 5(1), 27–47.
- Saidatunnisa, N. E., Sili, S., & Nasrullah, N. (2019). The inequality issues of male supremacy towards rose character in titanic movie. *Ilmu Budaya: Jurnal Bahasa, Sastra, Seni dan Budaya*, 3(4), 393–403.