

R Assignment - 2

Jasmeet Singh Saini - 0758054

2023-03-31

Background

Data Description

The data used in this analysis were obtained from the **Kaggle Titanic dataset** and it has been made available by Kaggle as part of a data science competition. The dataset consists of passenger information for **891** of the 2224 passengers and crew members who were aboard the Titanic during its maiden voyage in 1912, including their personal information and whether or not they survived the sinking of the ship. The dataset includes **training set** (also known as the “ground truth”) as *train.csv* with total size 61.19 KB.

Training Set

The *train.csv* file contains data on 891 passengers (891 rows), which includes 12 columns. The dataset provide the information of survival status of passengers, age of passengers (in years), gender of passengers (65% of male, 35% of female), passenger fare, socio-economic status and some other demographic information.

Below is a table summarizing the data:

Variable	Datatype	Key	Description
passengerId	Numeric	-	A unique identifier for each passenger
survived	Categorical	0=No, 1=Yes	A binary variable indicating whether the passenger survived or not
pclass	Categorical	1=1st, 2=2nd, 3=3rd	The ticket class of the passenger
name	Text	-	The name of the passenger
sex	Categorical	male, female	The gender of the passenger
Age	Numeric	-	The age of the passenger
sibsp	Numeric	-	The number of siblings and spouses the passenger had on board
parch	Numeric	-	The number of parents and children the passenger had on board
ticket	Text	-	The ticket number of the passenger
fare	Numeric	-	The fare paid by the passenger
cabin	Text	-	The cabin number of the passenger
embarked	Categorical	C=Cherbourg, Q=Queen -stown, S=Southampton	The port of embarkation of the passenger

In parch variable, some children travelled only with a nanny, therefore parch = 0 in the dataset. Moreover, sibsp variable the spouse mistresses and fiances were ignored but only brother, sister, stepbrother, stepsister and husband, wife were taken consider.

In the next sections, we will analyze the data using inferential statistics and visualize the analyses to draw insights and conclusions. For that, we will be testing whether the gender of the passenger is associated with their survival on the Titanic ship or not

Method

Results

Conclusion