EXECUTIVE

The RMS Titanic was a British passenger liner that sank in the North Atlantic Ocean in the early morning of April 15, 1912 after colliding with an iceberg during her maiden voyage from Southampton to New York City. The ship contained 2,224 passengers and crew, out of which 1,500 died in the unfortunate incident.

In this study, we intend to perform a statistical analysis of the fatalities on the ship using the Titanic dataset . The main question that we are addressing here is whether there is a statistically significance relation between the death of the person and their passenger class, age, sex and/or port where they embarked their journey.

INTRODUCTION

Titanic dataset consists of some variables and they are Passengerid, Survived, Pclass, Sex, Fare, Cabin, Embarked

We are going to analyse the dataset onto 4 variables and they are Pclass, Sex, Age, Embarked and the Response variable for these 4 variables is Survived

This report try to analyse the titanic dataset onto 4 variables and their effect on people survived or not

We are taking help of some visualizations to clear the analysis

PURPOSE

Titanic dataset analysis is done to find out the number of people survived and not survived .we are finding out that Four variables are actually affecting the survival of passengers or not. Age, sex, pclass are really affected the survival or not is going to calculated in this analysis using the visualization and then we are going to check that visualization results are actually true or not by using some functions of R.

METHODS

We have imported the Titanic dataset first

We have found the number of passengers according to group class, sex, age and embarked

Number of passengers survived or not survived

We have cleaned the data using na.omit() function

Four input variables and one response variable is taken into consideration

Computed main effects for all four factors validated effects are true or not by using the subset and nrow()

We have drawn the barplots and scatterplots

Scatterplots consist of Age vs Survived graph, Sex vs survived, Pclass vs survived and embarked vs Survived

RESULTS

Summary results

Class 1 consist of 216 people class 2 consist of 184 and class 3 consist of 491 people. Female passengers on titanic were less and they were 314 male passengers were more and they were 577

Age Summary consist of

Min :0.42

Median :28.00

Mean :29.70

Max:80

Embarked :

Cherbourg:168

Queenstown:77

Southampton:644

Main effect results:

1.

mean of class 1 people survived is 1.65,mean of class 2 people survived is 1.47 and mean of class 3 people survived is 1.23

From subset and nrow() functions

Class1 survived:122

Class2 survived: 83

Class3 survived:85

2.

Female survived :197

male survived:93

3.

Children survived: more (215)

Adult survived: medium(70)

senior survived: minimum(5)

4.

Cherbourg survived: more

Queenstown: medium

Southampton: minimum