

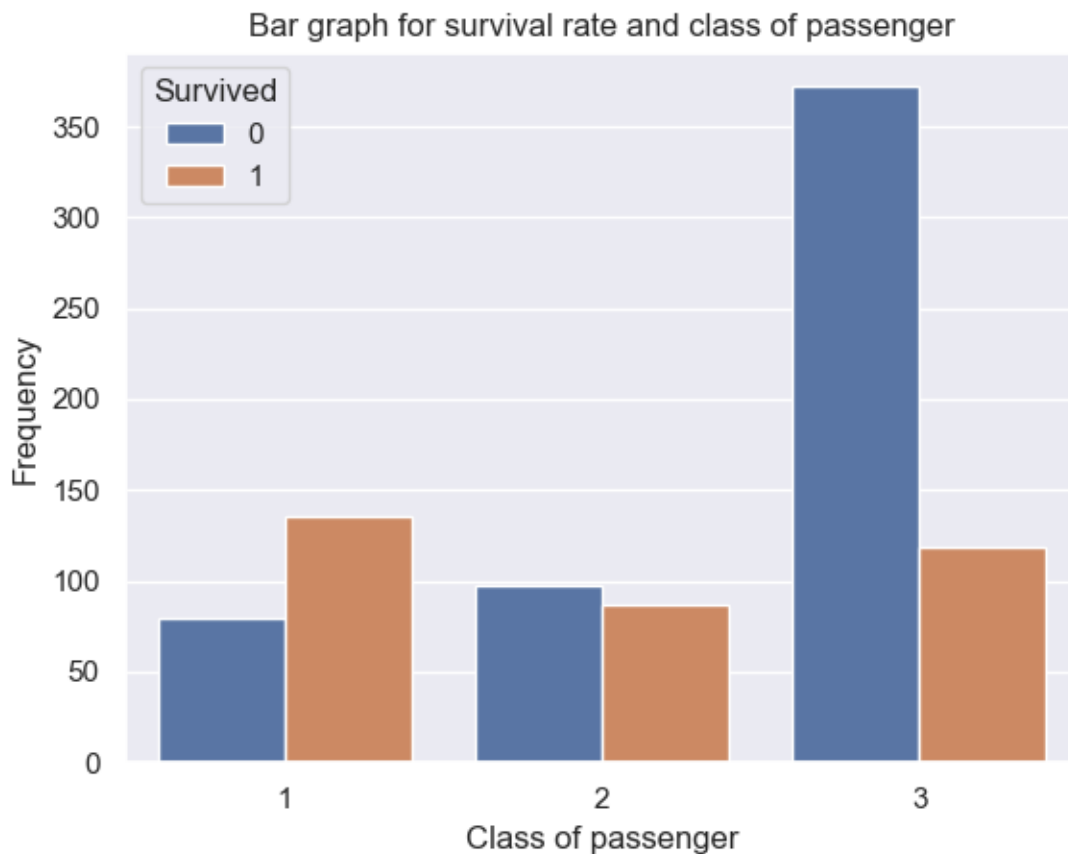
## Analytics for Titanic Dataset

Three selected hypotheses:

- Determine if the survival rate is associated to the class of passenger
- Determine if the survival rate is associated to the gender
- Determine the survival rate is associated to the age

### 1. Determine if the survival rate is associated to the class of passenger

#### 1.1. Bar plot:



Observations: the survival rate is different for each ticket class

- The number of deaths has been highest for passengers in class 3, and lowest for passengers in class 1
- The number of survivals is far outweighed by the number of deaths in class 3. It looks balance in class 2. In contrast, passengers in class 1 have a higher proportion in number of survivals.

### 1.2. Chi-square:

H0: Survival rate and ticket class are independent.

Ha: Survival rate and ticket class are not independent

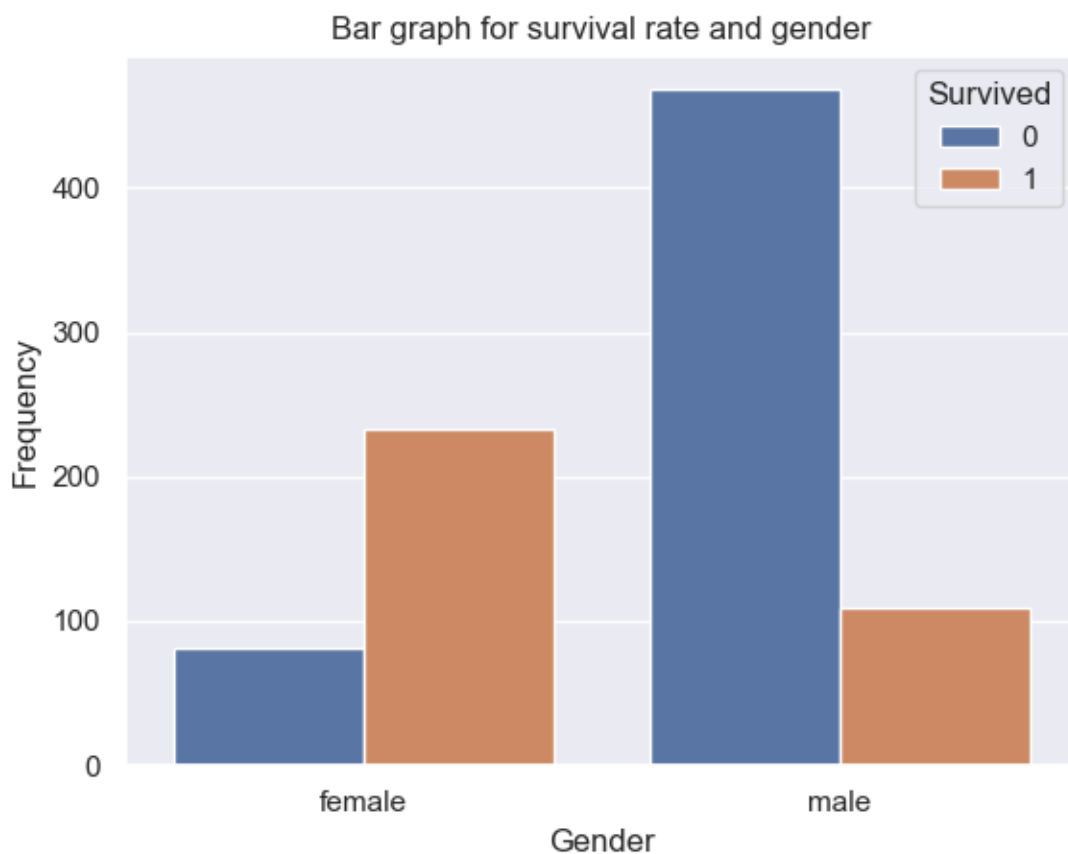
chi2 statistic = 102.89

df = 2  $\rightarrow$  dp = 5.99 12

Since the  $\chi^2 > dp$ , we reject the null hypothesis and conclude that; survival rate and ticket class are not significantly independent of each other among all subjects in the population.

## 2. Determine if the survival rate is associated to the gender

### 2.1. Bar plot:



Observations: the survival rate is different for each gender of passengers

- Most of the deaths are male.
- For male passengers, the number of deaths is much higher than that of survivals. In contrast, the survival rate is greater than death rate among female passengers.

## 2.2. Chi square

H0: Survival rate and ticket class are independent.

Ha: Survival rate and ticket class are not independent

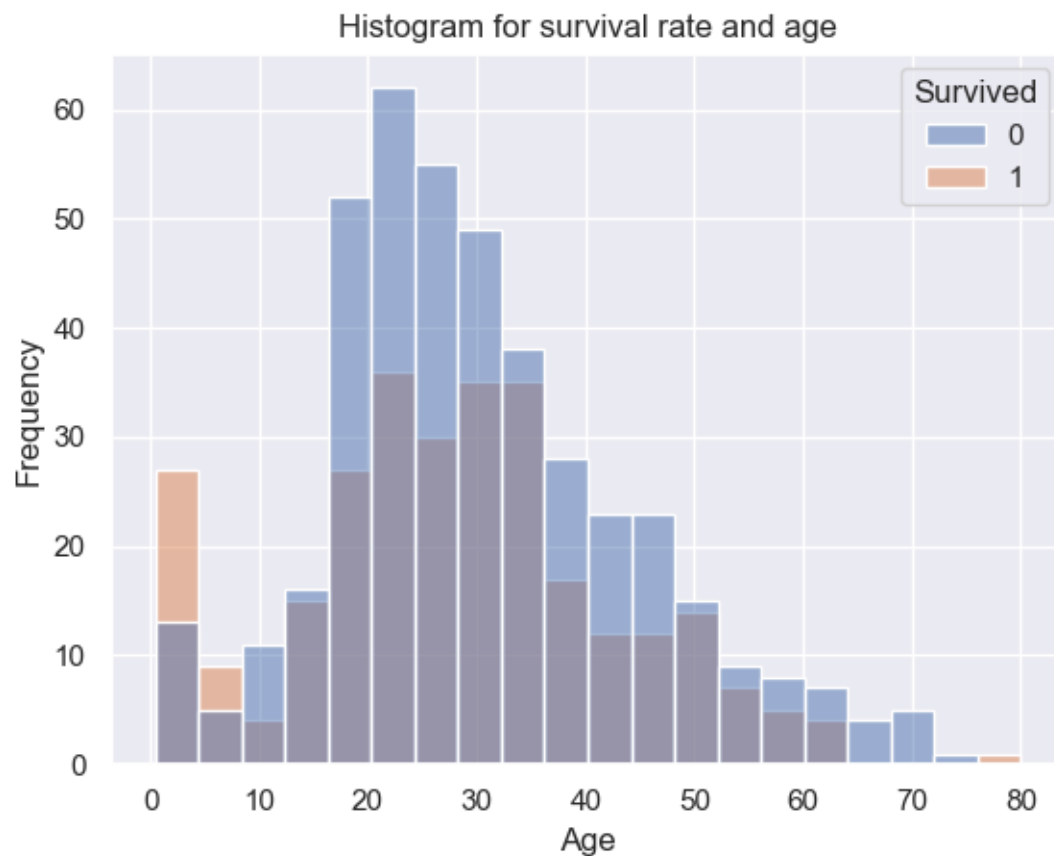
chi2 statistic = 260.72

df = 1  $\rightarrow$  dp = 3.84

Since the  $\chi^2 > dp$ , we reject the null hypothesis and conclude that; survival rate and gender are not significantly independent of each other among all subjects in the population.

## 3. Determine the survival rate is associated to the age

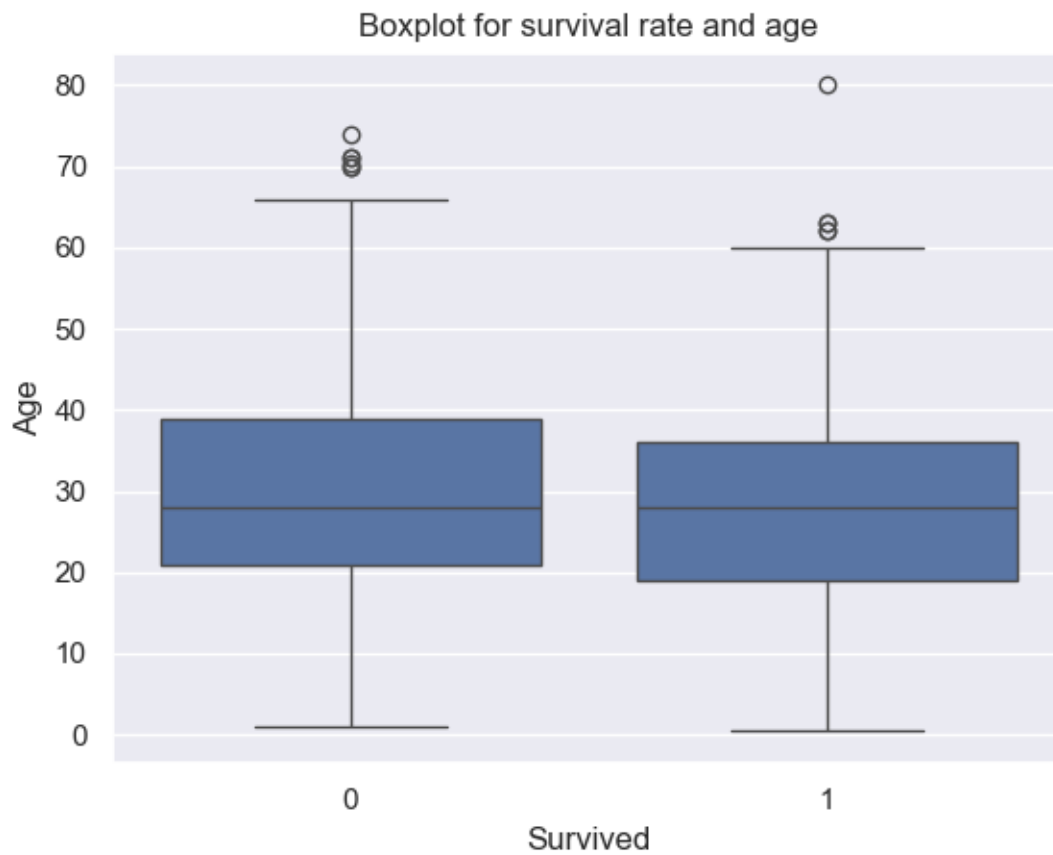
### 3.1. Histogram



Observations:

- The distribution of both survival and death rate are right-skewed.
- There death rate is greater than survival rate in all ages, except for the children less than 10 years old.

## 3.2. Box plot



Observations:

- The death rate and survival rate are quite similar distribution.