# **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. **Bar plot:**

A graph of a passenger

Description automatically generated

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. **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 🡪 dp = 5.99 12

Since the chi2 > 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.

1. **Determine if the survival rate is associated to the gender**
   1. **Bar plot:**

A graph of a bar graph

Description automatically generated

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.
  1. **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 🡪 dp = 3.84

Since the chi2 > 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.

1. **Determine the survival rate is associated to the age**
   1. **Histogram**

A graph of a graph of survival

Description automatically generated with medium confidence

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.
  1. Box plot

A graph of a box plot

Description automatically generated with medium confidence

Observations:

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