assignment2\_word

Sorawan Tiratrakoonwichaya

2024-09-10

The Titanic dataset is the information about 891 passengers that consists of passenger ID, name, survival status (0 = dead, 1 = survive), ticket class (Pclass = 1st, 2nd, 3rd), sex, age, number of siblings/spouses aboard (sibsp), number of parents/children aboard (parch), ticket number, passenger fare, cabin number, and port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton).

Summarizing available data on the Titanic passengers (Table 1) is Titanic passenger information, including the median and interquartile range (IQR) for quantitative variables, and the number of people (n) and percentage for qualitative variables. Moreover, there are boxplots with outliers for quantitative variables. The quantitative variables consist of age and fare, while the qualitative variables consist of ticket class, sex, number of siblings/spouses aboard, and number of parents/children aboard. There are 342 survivors out of 891 passengers, representing approximately 38% of the total. The median age of the survivors is 28 years, with an IQR of 17 years. There are 52 individuals with unknown ages. The median fare is 26 pounds, with an IQR of 45 pounds. Survivors who bought Class 1 tickets are about 40%, and the Class 2 and 3 tickets are 25% and 35%, respectively. 68% of survivors are female, and the others (32%) are male. There are 61% of survivors who traveled alone (without a sibling or spouse aboard), 33% for who had a sibling or spouse aboard, 3.8% for 2 siblings or spouses, 1.2% for 3, 0.9% for 4, and 0% for 5 and 8 siblings or spouses. From the information on survivors by the number of parents or children aboard, there are 68% of survivors who traveled alone (without a parent and child aboard), 19% for who had a parent or child aboard, 12% for 2 parents or children, 0.9% for 3, 0% for 4, 0.3% for 5, and 0% for 6. Looking at the data of the non-survivors (dead), there are 549 non-survivors of 891 passengers, which is about 62%. The median age of the non-survivors is 28 years, with an IQR of 18 years. There are 125 individuals with unknown ages. The median fare is 11 pounds, with an IQR of 18 pounds. Approximately 15%, 18%, and 68% of non-survivors purchased Class 1, 2, and 3 tickets, respectively. 15% of non-survivors are female, while 85% are male. There are 72% of non-survivors who traveled alone (without a sibling or spouse aboard). There are roughly 18%, 2.7%, 2.2%, 2.7%, 0.9%, and 1.3% of non-survivors who have one, two, three, four, five, and eight siblings or spouses aboard. The information regarding non-survivors based on the number of parents or children traveling indicates that 81% of survivors traveled alone (without a parent or child), 9.7% had a parent or child aboard, 7.3% had two, 0.4% had three, 0.7% had four and five, and 0.2% had six.

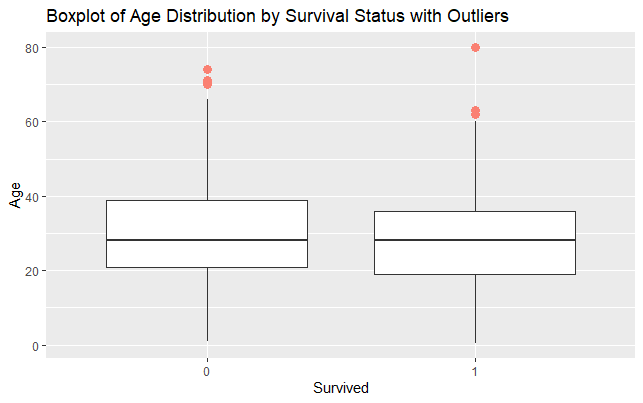
Table 1 shows that the median and IQR ages of survivors and non-survivors are similar, indicating that age does not affect survival (ignoring unknown age). Meanwhile, passengers who paid higher fares are more likely to survive than passengers who paid lower fares, which corresponds to the class of tickets purchased by passengers. Passengers who purchased first-class tickets have the highest survival rate, followed by passengers who purchased second-class and third-class tickets. A comparison of survival rates by sex shows that women have a higher chance of survival than men, which may indicate the attitudes and morals of people at the time. When comparing the number of siblings and parents together, it is found that although the percentage of survivors traveling alone is the highest, the percentage of non-survivors traveling alone is also the highest. When comparing the number of survivors (n), it is found that the number of survivors traveling alone is lower than the number of non-survivors. Meanwhile, those traveling with one or two other family members have similar numbers of survivors to the number of non-survivors. Therefore, it can be concluded that traveling with one or two other family members would have a higher chance of survival. However, if information on ticket class was used in conjunction with the number of family members traveling, a more reliable conclusion could be drawn. All the above conclusions show that the variables that affect survival are fare, ticket class, sex, and the number of people that traveling with.

“If you are a woman, buying a first-class ticket and traveling with 1 or 2 family members, you have the best chance of surviving the Titanic disaster.”

# Table 1: Summarizing available data on the Titanic passengers

|  | Survival Status | |
| --- | --- | --- |
| Variable | **0**  N = 5491 | **1**  N = 3421 |
| Age | 28 (18) | 28 (17) |
| Unknown | 125 | 52 |
| Fare | 11 (18) | 26 (45) |
| Ticket class |  |  |
| 1 | 80 (15%) | 136 (40%) |
| 2 | 97 (18%) | 87 (25%) |
| 3 | 372 (68%) | 119 (35%) |
| Sex |  |  |
| female | 81 (15%) | 233 (68%) |
| male | 468 (85%) | 109 (32%) |
| Number of siblings/spouses aboard |  |  |
| 0 | 398 (72%) | 210 (61%) |
| 1 | 97 (18%) | 112 (33%) |
| 2 | 15 (2.7%) | 13 (3.8%) |
| 3 | 12 (2.2%) | 4 (1.2%) |
| 4 | 15 (2.7%) | 3 (0.9%) |
| 5 | 5 (0.9%) | 0 (0%) |
| 8 | 7 (1.3%) | 0 (0%) |
| Number of parents/children aboard |  |  |
| 0 | 445 (81%) | 233 (68%) |
| 1 | 53 (9.7%) | 65 (19%) |
| 2 | 40 (7.3%) | 40 (12%) |
| 3 | 2 (0.4%) | 3 (0.9%) |
| 4 | 4 (0.7%) | 0 (0%) |
| 5 | 4 (0.7%) | 1 (0.3%) |
| 6 | 1 (0.2%) | 0 (0%) |
| 1Median (IQR); n (%) | | |

# Boxplot 1: Age Distribution by Survival Status with Outliers



# Boxplot 2: Fare Distribution by Survival Status with Outliers

