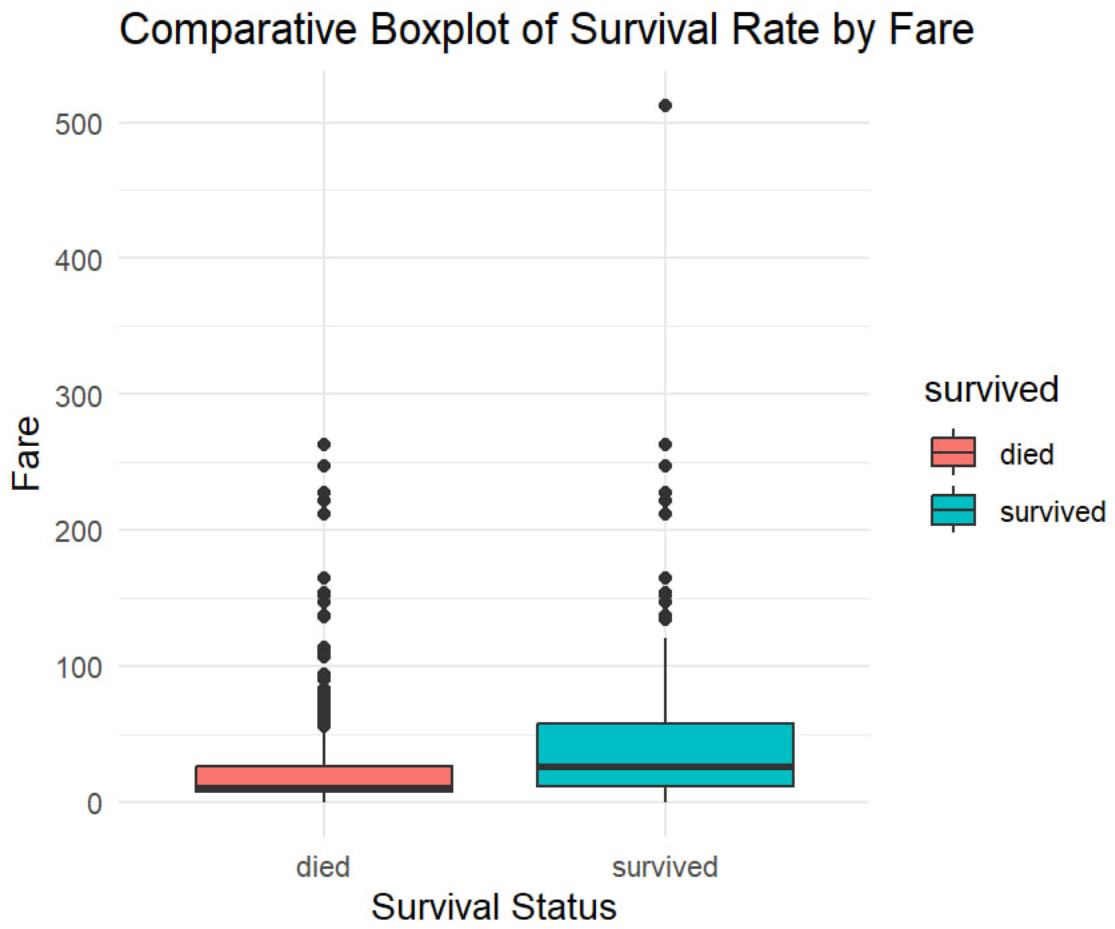
Chi Square for Categorical Variables with Titanic Dataset

1. Classifying Quantitative or Qualitative Variable

	Variable	Quantitative V.s. Qualitative
a)	pclass	Quantitative
b)	survival	Quantitative
c)	name	Quantitative
d)	Gender	Quantitative
e)	age	Qualitative
f)	sibsp	Qualitative
g)	parch	Qualitative
h)	ticket	Quantitative
i)	fare	Qualitative
j)	cabin	Quantitative
k)	embarked	Quantitative
l)	boat	Quantitative
m)	body	Quantitative
n)	home.dest	Quantitative
o)	Residence	Quantitative

3.




```r
# Load ggplot2 for visualization
library(ggplot2)

# Create a comparative boxplot of survival rate by fare
ggplot(TitanicData, aes(x = survived, y = fare, fill = survived)) +
  geom_boxplot() +
  labs(title = "Comparative Boxplot of Survival Rate by Fare",
       x = "Survival Status",
       y = "Fare") +
  theme_minimal()
```

According to the comparative boxplot, the median fare for passengers who survived is higher. It may indicate that wealthier passengers had a higher chance of survival.

4. Creating Two-way Tables

```
> # Two-way table for survival by gender
> table_bygender = table(TitanicData$survived, TitanicData$Gender)
> colnames(table_bygender)=c("Male", "Female")
> table_bygender

          Male Female
  died     682    127
  survived 161    339
>
> # Two-way table for survival by passenger class
> table_bypclass = table(TitanicData$survived, TitanicData$pclass)
> colnames(table_bypclass)=c("1st", "2nd", "3rd")
> table_bypclass

          1st 2nd 3rd
  died    123 158 528
  survived 200 119 181

>
> # Two-way table for survival by residence
> table_byres =table(TitanicData$survived, TitanicData$Residence)
> colnames(table_byres)=c("American", "British", "Other")
> table_byres

          American British Other
  died          113     206   490
  survived      145      96   259
>
> # Two-way table for gender by passenger class
> table_genbypclass = table(TitanicData$Gender, TitanicData$pclass)
> rownames(table_genbypclass)=c("Male", "Female")
> colnames(table_genbypclass)=c("1st", "2nd", "3rd")
> table_genbypclass

        1st 2nd 3rd
  Male   179 171 493
  Female 144 106 216
```

i. Survival by Gender

- Women had a significantly higher survival rate than men.

ii. Survival by Passenger Class

- First class passengers had a better chance of survival than those in second and third class.

iv. Gender by Passenger Class

- Wealthier women were more likely to afford first-class tickets.

5. Performing Chi Square Analysis

I. Hypothesize:

- Null Hypothesis (H0): There is no association between the two variables.
- Alternative Hypothesis (HA): There is an association between the two variables.

II. Prepare:

- Significance Level (Alpha): 5% or 0.05
- Will reject H0 if p-value is less than 0.05

III. Compute & Compare in R-studio:

```
> # Chi-square test for survival by gender
> chisq.test(table_bygender)

        Pearson's Chi-squared test with Yates' continuity correction

data:  table_bygender
X-squared = 363.62, df = 1, p-value < 2.2e-16

>
> # Chi-square test for survival by passenger class
> chisq.test(table_bypclass)

        Pearson's Chi-squared test

data:  table_bypclass
X-squared = 127.86, df = 2, p-value < 2.2e-16

>
> # Chi-square test for survival by residence
> chisq.test(table_byres)

        Pearson's Chi-squared test

data:  table_byres
X-squared = 44.835, df = 2, p-value = 1.838e-10

>
> # Chi-square test for gender by passenger class
> chisq.test(table_genbypclass)

        Pearson's Chi-squared test

data:  table_genbypclass
X-squared = 20.379, df = 2, p-value = 3.757e-05
```

IV. Interpret the Results:

i. Survival by Gender

- Since p-value < alpha, reject H0.
- There is a statistically significant association between gender and survival rate.

ii. Survival by Passenger Class

- Since p-value < alpha, reject H0.
- There is a statistically significant association between passenger class and survival rate.

iii. Survival by Residence

- Since p-value < alpha, reject H0.
- There is a statistically significant association between residence and survival rate.

iv. Gender by Passenger Class

- Since p-value < alpha, reject H0.
- There is a statistically significant association between passenger class and gender.

Conclusion:

The Chi-square tests show that there are significant associations between survival and gender, passenger class, and residence.