

# Analysis of Titanic dataset part 2

AUTHOR  
Michael Dang

PUBLISHED  
December 4, 2024

## Load the tidyverse library

```
library(epitools)
library(tidyverse)
```

## Comments on the code

For most of your programs, you should load the [tidyverse library](#). The messages and warnings are suppressed.

In previous programs, I put a label for each chunk inside the curly braces ({}). It is recommended instead to put the label on a separate line inside the program chunk. It is a bit more work to provide a unique label for each chunk, but it helps quite a bit to isolate where to look when your code produces an error.

## Read the data and view a brief summary

```
ti <- read_tsv(
  file="../data/titanic.txt",
  col_names=TRUE,
  col_types="ccncn",
  na="NA")
names(ti) <- tolower(names(ti))
glimpse(ti)
```

Rows: 1,313

Columns: 5

```
$ name      <chr> "Allen, Miss Elisabeth Walton", "Allison, Miss Helen Loraine"...
$ pclass    <chr> "1st", "1st", "1st", "1st", "1st", "1st", "1st", "1st", "1st"...
$ age       <dbl> 29.00, 2.00, 30.00, 25.00, 0.92, 47.00, 63.00, 39.00, 58.00, ...
$ sex       <chr> "female", "female", "male", "female", "male", "male", "female..."
$ survived  <dbl> 1, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1, 1, 1...
```

## Replace numeric codes for survived

```
ti$survived <-
  factor(
    ti$survived,
    level=1:0,
    labels=c("yes", "no"))
```

## Question 1

Create a new variable, `third_class` that indicates whether a passenger is in third class or not. What is the odds ratio comparing survival between third class passengers and first/second class passengers? Interpret this odds ratio and the associated confidence interval.

```
ti$third_class <-  
  case_when(  
    ti$pclass == "1st" ~ "no",  
    ti$pclass == "2nd" ~ "no",  
    ti$pclass == "3rd" ~ "yes")  
  
ti$third_class <- factor(ti$third_class, levels = c("no", "yes"))
```

Odd ratio comparing survival between third class passengers and first/second class passengers

```
table1 <- xtabs(~third_class+survived, data=ti)  
table1
```

```
      survived  
third_class yes  no  
      no   312 290  
      yes  138 573
```

```
oddsratio(table1)
```

\$data

```
      survived  
third_class yes  no Total  
      no   312 290   602  
      yes  138 573   711  
Total  450 863  1313
```

\$measure

```
      odds ratio with 95% C.I.  
third_class estimate    lower    upper  
      no  1.000000      NA      NA  
      yes  4.459216  3.496075  5.711579
```

\$p.value

```
      two-sided  
third_class midp.exact fisher.exact  chi.square  
      no      NA      NA      NA  
      yes      0 3.498409e-35  6.078882e-35
```

\$correction

```
[1] FALSE
```

```
attr(,"method")
```

```
[1] "median-unbiased estimate & mid-p exact CI"
```

- Odd ratio for third-class passengers (yes): 4.46
- Confidence Interval (95%): (3.5, 5.7)

## Confidence Interval Interpretation:

- The 95% confidence interval ranges from 3.50 to 5.71, indicating that the odds of not surviving for third-class passengers are at least 3.5 times and at most 5.71 times higher than for first/second-class passengers.
- Since the confidence interval does not include 1, the result is statistically significant, meaning that there is strong evidence that survival odds differ between third-class passengers and first/second-class passengers.

## P-value:

- The p-value is extremely small (close to 0), further confirming that the difference in survival odds between third-class and first/second-class passengers is statistically significant.

## Question 2

---

Calculate a chi-squared test of independence that examines the association between passenger class (third versus first/second) and mortality. Interpret the test result.

```
m1 <- chisq.test(table1, correct=FALSE)
m1
```

Pearson's Chi-squared test

data: table1

X-squared = 152.08, df = 1, p-value < 2.2e-16

- Chi-squared Statistics:  $X^2 = 315.24$ , indicates a strong deviation from the expected values under the null hypothesis.
- Since the p-value is very small (almost 0), we reject the null hypothesis. There is evidence for the association between passenger class (third vs. first/second) and mortality on the Titanic.