Analysis of Titanic dataset Module 14

This program reads data on survival of passengers on the Titanic. Find more information in the [data dictionary](#).

This code was written by Steve Simon and Leroy Wheeler on 2024-11-19 and is placed in the public domain.

Load the tidyverse library

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

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")
```

Get counts of sex by passenger class

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

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

Odds ratio calculation

```
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"
```

Interpretation of the odd ratio output

We are 95% confident that the odds ratio of survival for 1st/2nd class passengers is at least 3.5 and possibly as large as 5.7, after accounting for sampling error. This interval excludes the value of 1, so we conclude that the odds of survival is 4.5 times higher for 1st and 2nd class passengers than for third class passengers.

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
```

Interpretation of the chi-square output

Because the chi-squared statistic is much larger than the degrees of freedom and the p-value is small, we will reject the null hypothesis and conclude that passenger class status and survival are related (not independent)