R Homework 12

Fisher Ankney

November 27th, 2018

Statistics 5193

Question 1

Gender vs. Pinterest for th Class Data:

Create a 2 x 2 table of gender and pinterest with gender on the rows and Pinterest on the Columns. Print it to the console.

```
library(readxl)
StudentData <- read_excel("/Users/fisher/Documents/data_science/r_stat_5193/data/StudentData.xlsx")
```

```
ex_table <- table(StudentData$Gender, StudentData$Pinterest)
ex_table
```

```
##
##      N  Y
##   F 10 13
##   M 10  2
```

Question 1b

Construct a table that displays to proportion of Pinterest users among females and the proportion of Pinterest users among males. Print it to the console.

```
ex_prop_table <- prop.table(margin.table(ex_table, 1:2), 1)
ex_prop_table
```

```
##
##             N         Y
##   F 0.4347826 0.5652174
##   M 0.8333333 0.1666667
```

Question 1c

Get the Pearson residual for female Pinterest users and for male Pinterest users and interpret each.

```
resid(chisq.test(ex_table[,2]))
```

```
##         F         M
##  2.008316 -2.008316
```

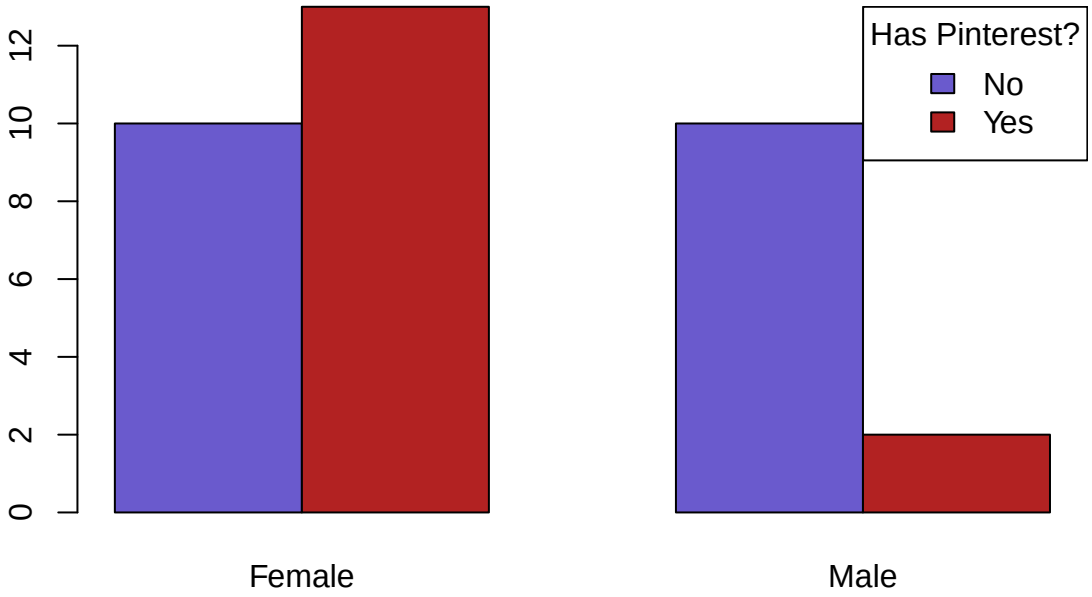
More women have pinterest than is expected, and less men have pinterest than is expected, if the expectation is that gender has no effect on pinterest use.

Question 1d

Construct a side by side bar plot. It should have female bars then males bars, and whether or not a person has a Pinterest account should be displayed with different colors and labeled with a titled legend.

```
barplot(t(ex_table),
        beside = T,
        names = c('Female', 'Male'),
        col = c('slateblue', 'firebrick', 'slateblue', 'firebrick'))

legend('topright',
       title = 'Has Pinterest?',
       legend = c('No', 'Yes'),
       fill = c('slateblue', 'firebrick'))
```

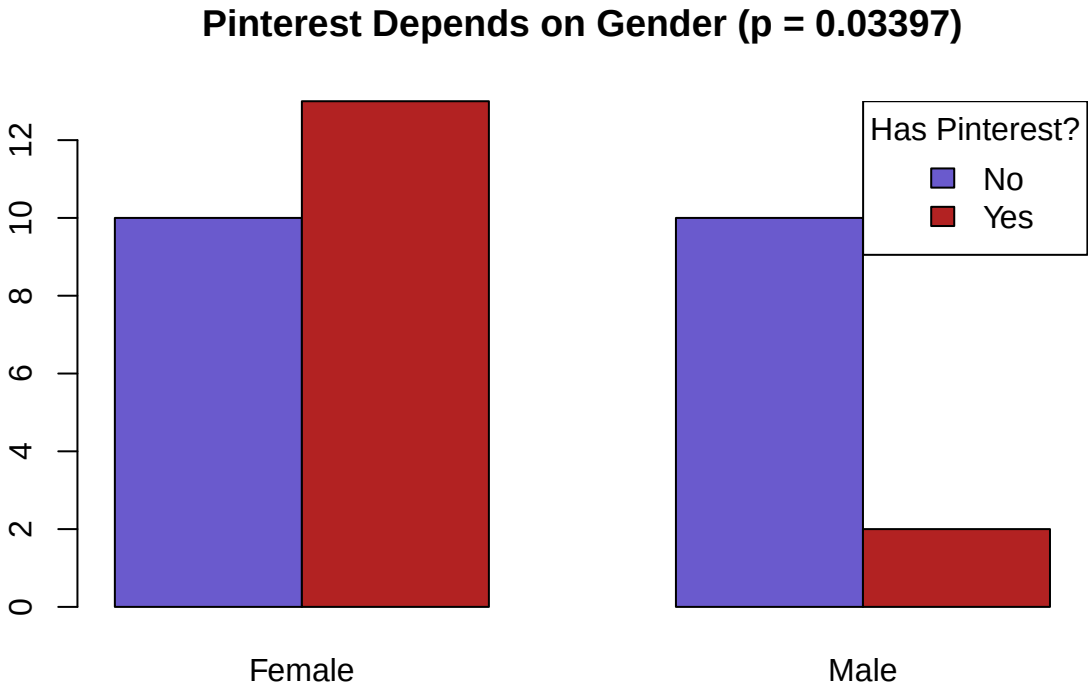


Question 1e

. Add a title to the plot that states "Pinterest Depends on Gender (p = _____)" with Fisher's exact test p-value inserted in the blank.

```
barplot(t(ex_table),
        main = 'Pinterest Depends on Gender (p = 0.03397)',
        beside = T,
        names = c('Female', 'Male'),
        col = c('slateblue', 'firebrick', 'slateblue', 'firebrick'))

legend('topright',
       title = 'Has Pinterest?',
       legend = c('No', 'Yes'),
       fill = c('slateblue', 'firebrick'))
```



Question 2a

The Titanic data set in R contains counts the number of people who survived and didn't survive the Titanic sinking across several factors including class (1st class, 2nd class, 3rd class, crew).

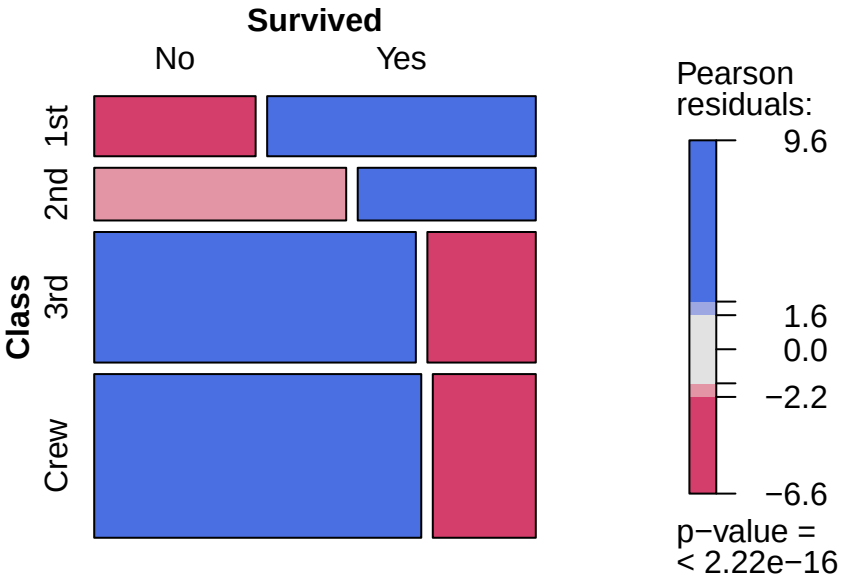
- Get a 4 x 2 table that displays the proportion of survivors and deaths among 1st class passengers, proportion of 2nd class passenger survivors and deaths and so on. Print it to the consol.
- Generate a colorful plot that depicts the Pearson standardized residuals and p-values for testing the null hypothesis that survival rate is independent of class.
- Did survival rate depend on class? Justify your answer with a p-value.

```
class_surv <- apply(Titanic, c(1,4), sum)

library(vcd)
```

```
## Loading required package: grid
```

```
mosaic(class_surv, legend = T, gp = shading_max)
```

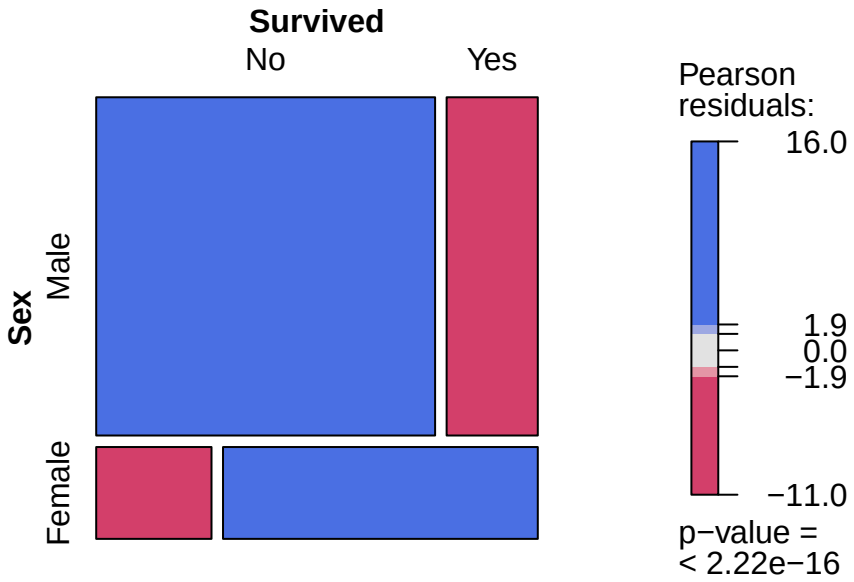


Yes, survival rate depends on class, with a p-value of $< 2.22 * 10^{-16}$

Question 3

Did survival rate depend on gender? Justify your answer with a p-value.

```
sex_surv <- apply(Titanic, c(2,4), sum)
mosaic(sex_surv, legend = T, gp = shading_max)
```

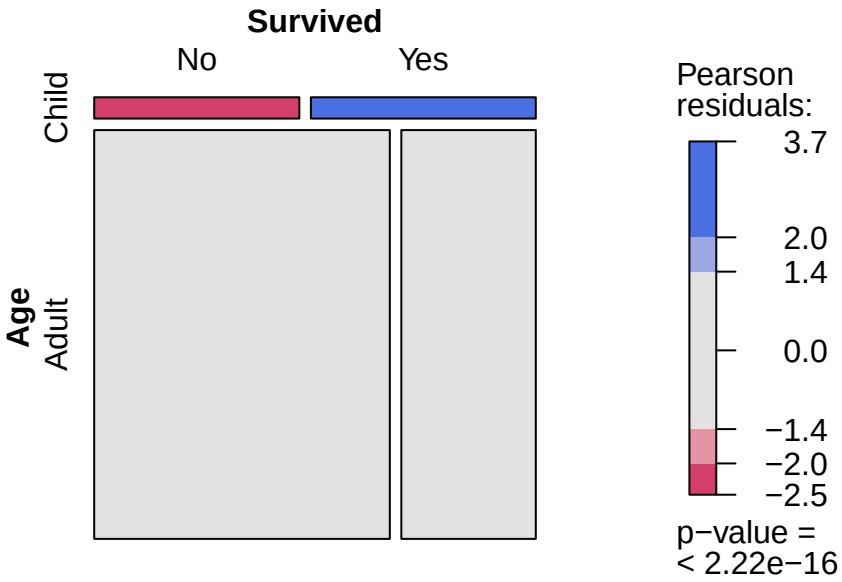


Yes, survival rate does depend on gender, with a p-value $< 2.22 * 10^{-16}$.

Question 4

Did survival rate depend on age? Justify your answer with a p-value.

```
age_surv <- apply(Titanic, c(3,4), sum)
mosaic(age_surv, legend = T, gp = shading_max)
```



Yes, survival rate does depend on age, with a p-value $< 2.22 * 10^{-16}$.