

# SurveyHypothesisTestExample

Julia Child

March 10, 2021

## Hypothesis: Does the choice of computer vary with political affiliation?

I want to test the hypothesis that whether students use Apple or PC's varies with political affiliation. This is a comparison of nominal to nominal scale data, so it'll require making a bar plot of frequencies and running a chi-squared test for independence. I will choose an alpha value of .05.

## Analysis

The following R code loads in the survey data and creates our 2x2 table of frequencies

```
# First we'll clear the workspace and load in the survey data:
rm(list = ls())
survey <- read.csv("http://www.courses.washington.edu/psy315/datasets/Psych315W21survey.csv")

# Then create the table
fo <- table(survey$vote, survey$computer)

# The result is a table with both rows and columns, with labels:
fo

##
##           Apple Other PC
## Democrat      59    9 33
## I never (or can't) vote 31    1 7
## Independent     3    0 4
## Other           1    0 2
## Republican     1    1 0

# The labels can be pulled out using 'row.names' and 'colnames' (note
# the inconsistency using '.' in the function names)
row.names(fo)

## [1] "Democrat"           "I never (or can't) vote"
## [3] "Independent"         "Other"
## [5] "Republican"

colnames(fo)

## [1] "Apple" "Other" "PC"

# The first and fourth rows correspond to Apple and PC's, and the 1st and 5th columns
# correspond to Democrat and Republican. This pulls out the relevant subset of rows and
# columns:

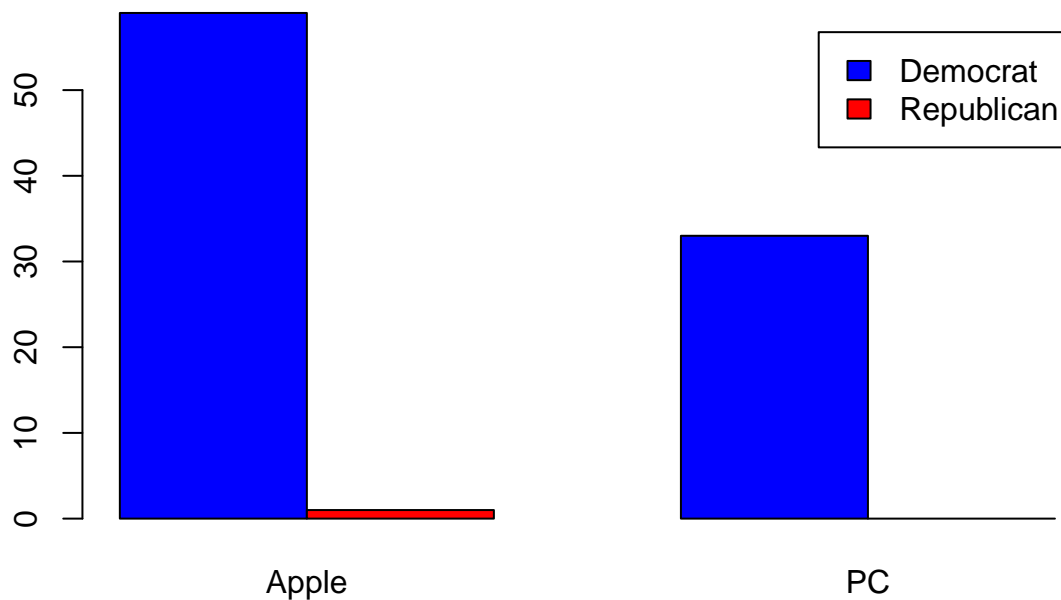
fo <- fo[c(1,5),c(1,3)]
```

## Results

```
# Here's the table of the results:  
fo
```

```
##  
##           Apple PC  
## Democrat    59 33  
## Republican   1  0
```

```
# And the bar graph (optional):  
barplot(fo,  
        beside=TRUE,  
        legend = row.names(fo),  
        col = c("Blue", "Red"))
```



```
# Here is the chi-squared test on the data  
out <- chisq.test(fo, correct = FALSE)
```

```
## Warning in chisq.test(fo, correct = FALSE): Chi-squared approximation may be  
## incorrect
```

```
# The chi-squared statistic is:  
out$statistic
```

```
## X-squared  
## 0.5559783
```

```
# The degrees of freedom is:
out$parameter

## df
## 1

# And the p-value is:
out$p.value

## [1] 0.4558852

# Writing in APA format can be done like this:
sprintf('Chi-Squared(%d,N=%d) = %5.2f, p = %5.4f',out$parameter,sum(fo),out$statistic,out$p.value)

## [1] "Chi-Squared(1,N=93) = 0.56, p = 0.4559"
```

## Summary

Our p-value of 0.5127 is much larger than .05 so our results are not statically significant. We therefore cannot conclude that the choice of computers varies with political affiliation.