SurveyHypothesisTestExample

Anthony S. Fauci M.D.

March 10, 2021

## Hypothesis: Does the choice of computer vary with choice of superpower?

I want to test the hypothesis that whether students use Apple or PC’s varies with their choice of superpower. 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$superpower,survey$computer)

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

##
## Apple Other PC
## 1 0 2
## Flight 47 7 17
## Invisibility 47 4 27

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

## [1] "" "Flight" "Invisibility"

colnames(fo)

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

# The second and third correspond to 'Flight and Invisibility', and the 1st
# and 3rd columns correspond to Apple and PC. This pulls out the relevant
# subset of rows and columns:

fo <- fo[c(2,3),c(1,3)]

## Results

# Here's the table of the results:
fo

##
## Apple PC
## Flight 47 17
## Invisibility 47 27

# 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)

# The chi-squared statistic is:
out$statistic

## X-squared
## 1.556262

# The degrees of freedom is:
out$parameter

## df
## 1

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

## [1] 0.2122134

## Summary

Our p-value of 0.2122 is much larger than .05 so our results are not statically significant. We therefore cannot conclude that the choice of computers varies with choice of superpower.

# 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=138) = 1.56, p = 0.2122"