

Chi-Squared Test in R

Hannah Waddel

Chi-Squared Test

Recall from Bios 500 that a Chi-Squared test is used when we want to test the relationship between two categorical variables. In this case, we are going to be testing whether the proportion of patients who get the flu is different between those who got the vaccine and those who got a placebo.

Our null and alternative hypotheses are as follows:

$$H_0 : p_1 = p_2$$

$$H_A : p_1 \neq p_2$$

Where:

p_1 = proportion of vaccine recipients who get the flu

p_2 = proportion of placebo recipients who get the flu

Chi-Squared test with Contingency Table

There are two types of data that we can use for the Chi-Squared test in R. The first is contingency table data, where the data are in a 2x2 table (or any number of rows and columns), where each cell in the table contains a count of subjects.

This is our data for the example:

Outcome	Population	
	Placebo	Vaccine
No	49	75
Yes	31	25

To create contingency table data, we create a matrix with the counts. We give the matrix a list of the counts, and set it to have 2 rows and 2 columns. We also tell R to fill in the matrix along the rows.

```
fluvac.table <- matrix(c(49,75,31,25),nrow=2,ncol=2,byrow=T)
print(fluvac.table)
```

```
##      [,1] [,2]
## [1,]   49   75
## [2,]   31   25
```

Now, we conduct the chi-square test with the `chisq.test()` function.

```
chisq.test(fluvac.table)
```

```
##  
## Pearson's Chi-squared test with Yates' continuity correction  
##  
## data: fluvac.table  
## X-squared = 3.3053, df = 1, p-value = 0.06906
```

The output of `chisq.test()` is actually a data frame which contains more information about the test. Let's save the output as an object in order to see what the expected counts were, which will allow us to check the chi-squared assumptions.

```
fluchi.table <- chisq.test(fluvac.table)
```

We can use the `str()` function, short for “structure”, to see what information the chi-squared test object contains.

```
str(fluchi.table)
```

```
## List of 9  
## $ statistic: Named num 3.31  
## ..- attr(*, "names")= chr "X-squared"  
## $ parameter: Named int 1  
## ..- attr(*, "names")= chr "df"  
## $ p.value : num 0.0691  
## $ method : chr "Pearson's Chi-squared test with Yates' continuity correction"  
## $ data.name: chr "fluvac.table"  
## $ observed : num [1:2, 1:2] 49 31 75 25  
## $ expected : num [1:2, 1:2] 55.1 24.9 68.9 31.1  
## $ residuals: num [1:2, 1:2] -0.823 1.225 0.736 -1.096  
## $ stdres : num [1:2, 1:2] -1.98 1.98 1.98 -1.98  
## - attr(*, "class")= chr "htest"
```

The test has a lot of information, but we are interested in the “\$ expected” part. This tells us that there is a part of the test data frame which we can access with the `$` operator.

```
fluchi.table$expected
```

```
##           [,1]      [,2]  
## [1,] 55.11111 68.88889  
## [2,] 24.88889 31.11111
```

We can see that the expected values were greater than 5 in each of the cells, so it was appropriate to use the chi-squared test rather than Fisher's exact test.

We now make our conclusion as follows: “Fail to reject H_0 . With a p value greater than 0.05, we do not have evidence to conclude that the probability of getting the flu was different between vaccine and placebo recipients.”

Chi-Squared test with Patient Data

We can also conduct the chi-squared test when we have individual-level patient data. The null and alternative hypotheses are the same. The data “fluvaccine.csv” contains individual-level patient data corresponding to the contingency table above.

We first set our working directory, read in, and save the data as an R object, here called “fluvac.data”.

```
setwd("~/Work/BIOS 591P 2021/R Materials/1 Introduction/Chi-Square Test")  
  
fluvac.data <- read.csv("fluvaccine.csv")
```

We can view the data to verify that we read it in correctly.

```
View(fluvac.data)
```

Now, we can still use the `chisq.test()` function, but we give the function two arguments: the two categorical variables for which we are testing the association. We extract the data from the fluvac.data data frame with the `$` operator.

```
chisq.test(fluvac.data$Flu, fluvac.data$Vaccine)  
  
##  
## Pearson's Chi-squared test with Yates' continuity correction  
##  
## data: fluvac.data$Flu and fluvac.data$Vaccine  
## X-squared = 3.3053, df = 1, p-value = 0.06906
```

Observe that the test results in an identical test statistic and p value as before, and thus we come to the same conclusion.

We can save the chi-squared test again, in order to check that the expected values in each cell are greater than 5.

```
fluchi.data <- chisq.test(fluvac.data$Flu, fluvac.data$Vaccine)  
  
fluchi.data$expected
```

```
##                fluvac.data$Vaccine  
## fluvac.data$Flu Placebo Vaccine  
##                No  55.11111 68.88889  
##                Yes  24.88889 31.11111
```