

Chi-Square Test of Independence Tutorial

FMU Biology Department

January 3, 2021

Chi-Square Test in R

This documents includes code for creating a table object in R, displaying summary statistics, creating an informative plot, and running Chi-Square test of independence.

Background

There has been an outbreak in the deadly Pokemon disease, Pokemonitis. You and Pokemon Professor have decided to investigate the issue. Based on your extensive work in Pokemon genomics, you suspect the presence of the Poke-X gene is linked to the development of Pokemonitis. To evaluate this relationship you take a random sample of 2000 Pokemon to test for the Poke-X gene and Pokemonitis.

Question: Is there an association between the Poke-X gene and Pokemonitis?

Hypothesis: There is an association between the Poke-X gene and Pokemonitis (i.e., Pokemon with the Poke-X gene tend to develop Pokemonitis)

Null Hypothesis: There is no association between the two variables (i.e., they are independent)

Data

The table created below contain 2 rows and 2 columns. Rows represent the frequency of Pokemon with the Poke-X gene and columns represent the frequency of Pokemon that develop Pokemonitis.

Important points to note. The function “as.table” creates a table object that is being assigned the name “PokeData”. Everything after “as.table” on the same line are the arguments being passed to the function. The “rbind” function is combining the data by rows. In this example the first row of data (793 and 325) represent the number of Pokemon with the Poke-X gene that developed Pokemonitis (N=793) or did not develop Pokemonitis (N=325). The second row of data (524 and 358) represent the number of Pokemon without the Poke-X gene that develop Pokemonitis (N=524) or did not develop Pokemonitis (N=358).

Coding Note: This section demonstrates using functions within functions. As you see, the “rbind” function is embedded in the “as.table” function.

Enter the following code in R:

```
#Create data table
PokeData <- as.table(rbind(c(793, 325), c(524, 358)))

#Assign names to rows and columns
dimnames(PokeData) <- list(PokeX = c("With", "Without"),
                           Pokemonitis = c("Yes", "No"))
```

The next line of code will print the data contained in the dat3 object

```
PokeData
```

```
##          Pokemonitis
## PokeX      Yes  No
##   With    793 325
##   Without 524 358
```

Run Chi-Square Test

This code chunk performs a Chi-Square Test of Independence.

```
#Perform a chi-square test and print x-squared test statistics, degrees of freedom, and p-value  
chisq.test(PokeData)
```

```
##  
## Pearson's Chi-squared test with Yates' continuity correction  
##  
## data:  PokeData  
## X-squared = 28.585, df = 1, p-value = 8.966e-08
```

Full code block

```
#Create data table  
PokeData <- as.table(rbind(c(793, 325), c(524, 358)))  
  
#Assign names to rows and columns  
dimnames(PokeData) <- list(PokeX = c("With", "Without"),  
                           Pokemonitis = c("Yes", "No"))  
  
#Print data frame  
PokeData  
  
#Perform a chi-square test and print x-squared test statistics, degrees of freedom, and p-value  
chisq.test(PokeData)
```