



EvalFest How-Tos for Conducting a Chi Square Analysis using RStudio

Chi Square Test of Independence

1. Creating the contingency (or two-way table)
 - a. If you've watched our video that explains the concept of chi square, you may recall the example of the four corners activity and that the chi square test counts the number of people in each category. The number of people in each group is an example of the observed count in the data. To get started with this analysis in RStudio, we have to first create a table of the observed counts to show the pattern of the data. This is called a contingency table or a two-way table. It shows how data from two different items relate.
 - b. In the conceptual video, we provide the example of boys versus girls and whether they would come back to a similar event in the future. The contingency table for that example showed the number of girls who said they would and would not come back and the number of boys who said they would and would not come back.
 - c. See the file containing the code you will use to tell RStudio how to create a contingency table. We have made all of this code available for you so that you can download the file and then change as needed to work with your data. If you've watched our video that shows you how to set up the data and variables for R, you will notice that lines 1-3, 6, and 7 are part of the initial setup.
 - d. In this example, we are going to compare Gender and Awareness of STEM due to Festival Attendance. The question asked festival attendees, "Did today's event make you more aware of how STEM is part of your daily life?" In this example, Gender is the predictor variable – we want to know if it predicts different patterns in Awareness of STEM. Awareness is our outcome variable – it is the variable that we want to learn more about.
 - e. In this file, Line 10 shows the code that creates this table, and Line 11 will show the values in the table.
2. Running the Chi Square Test of Independence
 - a. Once you have your contingency table, you can tell RStudio to run the chi square test of independence. That's what Line 14 does. In this case, it tells RStudio to run the chi square test for the contingency table that was created in Line 10.
3. Interpretation of the Chi Square Test of Independence Output
 - a. You will see output appear in the Console window – these are your results. RStudio provides the chi square value, the degrees of freedom, and the p-value. You're most interested in the p-value to determine if the outcome variable is significantly dependent on the predictor variable. If the p-value is

less than .05, then that means you have different patterns in your outcome variable based on your predictor variable.

- b. You can look back at the data to make more specific statements about how the variables relate. In our example, when you look at the percentage breakdown of Awareness by Gender, you can see that the percentage of Females who said that festival attendance made them more aware of how STEM is part of their daily lives is significantly more than the percentage of Males who indicated that festival attendance helped to increase their awareness. Interpreting the results of a 2x2 relationship are pretty easy because as one thing goes up the other goes down. Things get a bit more tricky when you have more than two groups.



UNC
MOREHEAD PLANETARIUM
AND SCIENCE CENTER



karen peterman
CONSULTING

