



Chi Square Test of Independence

1. Setting up the data
 - 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. The chi square test then compares these observed counts with expected values for each category.
 - b. In the conceptual video, we provide the example of tweens and teens and whether they would come back to a similar event in the future. The contingency table, or table of observed counts, for that example showed the number of tweens who said they would and would not come back and the number of teens who said they would and would not come back.
 - c. In this example, we are going to compare Awareness of STEM due to Festival Attendance by age groups – those under 35 years of age, and those older than 35. 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, Age group 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.
 - d. The EvalFest Database for SPSS.sav file contains the data we will use to conduct the chi square test. We will be working with the Under35 variable (that is our predictor) and the Aware variable (that is our response).
 - e. You want to ensure that you only use the cases for which you have data for both variables. You can select data in SPSS by going to Data, Select Cases. You'll then click If..., and indicate here which cases you want to include. Here, you should indicate that you want to include cases where Under35 is either 0 (not under 35) or 1 (under 35) and where the answer to the awareness item is either Yes or No. That is, both criteria have to be satisfied in order for the case to be used in my analysis.
2. Conducting the Chi Square Test of Independence
 - a. In order to conduct a chi square test of independence, we go to Analyze, Descriptive Statistics, and choose Crosstabs...
 - b. Now, we need to indicate which variable is the predictor variable and which is the outcome variable. We will click on Under35 on the left and click the arrow next to the Rows box to indicate that this is the predictor variable. Similarly, we will click on Aware on the left and click on the arrow next to the Columns

- box to indicate that this is the outcome variable.
- c. Next, click Statistics... on the right. This allows us to indicate which tests we want to conduct. Click the box next to Chi-square to run a chi square test of independence. Then click Continue.
 - d. Next, click Cells...on the right. Make sure the Observed box is checked to indicate that the data we are using are observed counts. In the Percentages area, click Row, Column, and Total. Then click Continue.
 - e. Click OK at the bottom right.
3. Interpretation of the Chi Square Test of Independence Output
- a. SPSS will populate the Output window with the results of the chi square test of independence. This includes a contingency, or two-way table, as well as the chi square value, the degrees of freedom, and the p-value for various tests. You're most interested in the Asymptotic Significance (2-sided) p-value for the Pearson Chi-Square test to determine there are significantly different patterns in Awareness of STEM by Age group.
 - b. If you've watched the conceptual video, you may recall that the p-value is a measure of the risk you take in making a decision based on your data. If the p-value is less than .05, that means there is less than 5% chance that the results are based on something other than the data. Thinking about this another way, there is a 1 minus the p-value chance that there is a significant pattern in the way the data cluster.
 - c. In more statistical terms, a p-value less than .05 means that you have distinct patterns in your outcome variable (STEM awareness) based on your predictor variable (age group).
 - d. However, that isn't the case in this example - our p-value is 0.860, indicating that there is no significantly different pattern in Awareness of STEM based on Age group.
 - e. If we did compute a p-value less than .05, we could look back at the contingency table to make more specific statements about how the variables relate. Interpreting the results of a 2x2 relationship is pretty easy because as one thing goes up the other goes down. Things get a bit trickier when you have more than two groups. Please see the Bonferroni in SPSS video on the EvalFest website for more information about this additional analysis.



UNC
MOREHEAD PLANETARIUM
AND SCIENCE CENTER



karen peterman
CONSULTING

