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 Chi Square in Excel.xlsx file contains the data we will use to conduct the chi square test of independence. The table at the top shows the observed counts from the EvalFest data set. Next, we need to compute the expected values.
2. Creating the table of expected values
   a. In order to compute the expected values, we need to multiply the column total by the row total and divide that product by the grand total. We will do that for each crosstab value in the Expected table.
   b. Once we've computed the expected values, we can run the chi square test that compares the observed counts and expected values.
3. Conducting the chi square test
   a. The chitest function in Microsoft Excel conducts a chi square test.
   b. Type =chitest(, then highlight the observed counts, type a comma, then highlight the expected values, and then type a close parentheses and hit Enter.
   c. This computes the p-value associated with the chi square test. You may want to label this p-value.
   d. 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.

e.   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).

f.   However, that isn't the case in this example. If you did compute a p-value less than .05, you could look back at the data 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.