

16.6: Practice Chi-Square Test of Independence- College Sports

We will be using the frequencies of watching college sports and how much college sports teams affect the choice of college attendance again, so here is the contingency table of the Observed frequencies and the table with the Expected frequencies that have been presented before. They are being presented again so that you don't have to go back and forth to find all of the data.

Here's the contingency table with the Observed frequencies:

Table 16.6.1: Contingency Table of College Sports and College Decision

College Sports	Primary Affect	Somewhat Affected	Did Not Affect Decision	Total
Watched	47	26	14	87
Did Not Watch	21	23	37	81
Total	68	49	51	168

and the Expected frequencies:

Table 16.6.2: Table of EXPECTED Frequencies of College Sports and College Decision

College Sports	Primary Affect	Somewhat Affected	Did Not Affect Decision	Total
Watched	35.21	25.38	26.41	87
Did Not Watch	32.79	23.63	24.59	81
Total	68	49	51	168

Now we're ready to follow the same 4-step procedure that you've come to know.

Step 1: State the Hypotheses

Chi-Square tests on patterns of relationship, and this doesn't change when we have more than one variable. The research hypothesis does not need to specify how each cell relates to each other like in a factorial ANOVA because we generally don't do pairwise comparisons for Chi-Square analyses. Instead, you can describe a general pattern of relationship between the two variables.

✓ Example 16.6.1

What is the research hypothesis in words for this scenario? Make sure to describe a general pattern.

Solution

- Research hypothesis in words: There will be a pattern of difference such that there will be more people whose college decision was affected by college sports AND who watched college sports.

Remember, the hypotheses in symbols we can only say that the probabilities will not be equal. To determine that, let's figure out what the probability would be if all of the cells were equal. To find that out, we would divide a probability of 100% by the number of cells. There are six cells (Affected Decision=3; Watched=2; $2 \times 3 = 6$).

$$\frac{100}{6} = 16.67$$

So the probability that any random participant will fall into a specific cell is 0.167 for each cell.

✓ Example 16.6.2

What is the research hypothesis in symbols for this scenario?

Solution

- Research hypothesis in symbols: $P_{EachCell} \neq 0.167$.

If this research hypothesis in symbols doesn't make sense, it might be easier to start with a null hypothesis in words and symbols, then figure out how that works out for the research hypothesis. But honestly, it's not a huge deal if you don't get the probability part of the hypotheses. The important point is that you are testing the null hypothesis that all frequencies will be similar (no pattern of relationship), but that you actually expect a particular pattern (research hypothesis).

✓ Example 16.6.3

What is the null hypothesis in words and symbols for this scenario?

Solution

- Null hypothesis in words: There is no pattern of difference in watching college sports affect college decisions.
- Null hypothesis in symbols: $P_{EachCell} = 0.167$.

Before we move on to an easier step, let's stop and remind ourselves that, just like correlations, Chi-Square tests cannot show that watching sports *caused* people to choose their college differently. This scenario is set up to make you think that the IV is watching college sports and the DV is the person's choice of college, but a Chi-Square can't test whether one thing causes another; these statistical analyses can only show if there's a *pattern of relationship or not*. The design of the experiment (how we collect the data to rule out alternative causes) is how we can show that one variable causes changes in another. This scenario is basically asking the participants *if they think* that college sports affected their choice of college. It's a small distinction, but an important one.

Step 2: Find the Critical Value

Okay, moving on from the "correlation doesn't equal causation" rant applied to Chi-Square!

Our critical value will come from the same table that we used for the Goodness of Fit Chi-Square test, but our degrees of freedom will change. Because we now have rows and columns (instead of just columns) our new Degrees of Freedom use information on both. This is described at the bottom of the [Critical Value of Chi-Square Table page](#), and looks like this:

- χ^2_{Tot} Test of Independence: $(R - 1) \times (C - 1)$
 - R is the number of rows
 - C is the number of columns

What this means is that the number of rows minus one is multiplied by the number of columns minus one. In our example:

$$df = (2 - 1)(3 - 1) = 1 \times 2 = 2$$

? Exercise 16.6.1

What is the critical value for this scenario from the $p = 0.05$ column?

Answer

With our $df = 2$ ($df = (2 - 1) * (3 - 1) = 1 \times 2 = 2$), the critical value is 5.991.

Step 3: Calculate the Test Statistic

You probably won't believe it, but you finally have caught a break in learning formulas. The formula for Chi-Square's Goodness of Fit test is the same formula for the Chi-Square Test of Independence!

$$\chi^2 = \sum_{Each} \left(\frac{(E - O)^2}{E} \right)$$

If you find a way to combine Table 16.6.1 and Table 16.6.2 with the Differences, Differences Squared and divided by the Expected frequencies into one table, then you are a data visualization wizard! For now, we'll create a new table for each step of the formula.

✓ Example 16.6.4

Use the previous two tables (Table 16.6.1 and Table 16.6.2) to create a table of differences by subtracting the Observed frequencies from the Expected frequencies for each cell.

Solution

Here is a table of differences:

Table 16.6.3: Table of Difference of Observed & Expected Frequencies of College Sports and College Decision

College Sports	Primary Affect	Somewhat Affected	Did Not Affect Decision	Total
Watched	35.21-47=-11.79	25.38-26=-0.62 (or -0.63)	26.41-14=12.47	0
Did Not Watch	32.79-21=11.79	23.63-23=0.63	24.59-37=-12.41	0
Total	0	0 (or 0.01)	0	0 (or 0.01)

Notice that the row and column Totals are zero (or nearly so, depending on rounding). This is another good calculation check!

So far, we've accomplished the part in the parentheses of the formula:

$$\chi^2 = \sum_{Each} \left(\frac{(E - O)^2}{E} \right)$$

What is the next step in this formula?

✓ Example 16.6.4

Square the difference scores in each cell of Table 16.6.3

Solution

Table 16.6.4: Table of Squared Differences of Observed & Expected Frequencies of College Sports and College Decision

College Sports	Primary Affect	Somewhat Affected	Did Not Affect Decision
Watched	139.00	0.38	154.01
Did Not Watch	139.00	0.40	154.01

The row and column for the Total was removed because those sums aren't used for anything. You can calculate them for completeness, but it won't help you finish the formula. Unless it helps you not get lost!

Now we've finished the numerator of the formula:

$$\chi^2 = \sum_{Each} \left(\frac{(E - O)^2}{E} \right)$$

What is the next step?

✓ Example 16.6.5

Using the squared differences in Table 16.6.4, complete the formula by dividing each cell by its own Expected frequency (found in Table 16.6.2). Then, add up all of the Total rows and columns to get the calculated χ^2 .

Solution

Table 16.6.5: Table of Squared Difference of Observed & Expected Frequencies of College Sports and College Decision Divided by Expected Frequencies

--	--	--	--

College Sports	Primary Affect	Somewhat Affected	Did Not Affect Decision	Total
Watched	$\frac{139}{35.21} = 3.95$	$\frac{0.38}{25.38} = 0.01$	$\frac{154.01}{26.41} = 5.83$	$\sum_{Row} = 9.79$
Did Not Watch	$\frac{139}{32.79} = 4.24$	$\frac{0.40}{23.63} = 0.02$	$\frac{154.01}{24.59} = 6.26$	$\sum_{Row} = 10.52$
Total	$\sum_{Column} = 8.19$	$\sum_{Column} = 0.03$	$\sum_{Column} = 12.09$	$\sum_{Column} = 20.31$ $\sum_{Row} = 20.31$

The Total for summing the Totals for the columns is the same as the sum of the Totals for the rows, so we did it correctly!

Okay, let's see if, for the Test of Independence, doing your calculations in the formula might be easier.

✓ Example \(\backslash\PageIndex{6}\)

Use the Chi-Square formula to calculate the χ^2 statistic

Solution

Using the information in Table 16.6.1 and Table 16.6.2 we find:

$$\begin{aligned}\chi^2 &= \frac{(35.21 - 47)^2}{35.21} + \frac{(25.38 - 26)^2}{25.38} + \frac{(26.41 - 14)^2}{26.41} + \frac{(32.79 - 21)^2}{32.79} + \frac{(23.62 - 23)^2}{23.62} + \frac{(24.59 - 37)^2}{24.59} \\ \chi^2_{Diff} &= \frac{(-11.79)^2}{35.21} + \frac{(-0.62)^2}{25.38} + \frac{(12.41)^2}{26.41} + \frac{(11.79)^2}{32.79} + \frac{(0.63)^2}{23.62} + \frac{(-12.41)^2}{24.59} \\ \chi^2_{DiffSquared} &= \frac{139}{35.21} + \frac{0.38}{25.38} + \frac{154.01}{26.41} + \frac{139.00}{32.79} + \frac{0.40}{23.62} + \frac{154.01}{24.59} \\ \chi^2_{Division} &= 3.95 + 0.01 + 5.83 + 4.24 + 0.02 + 6.26 = 20.31 \\ \chi^2 &= 20.31\end{aligned}$$

What do you think? Was it easier to do the calculations in five different tables, or do it all in one formula? There's no right answer for this, it really is what's easier for you.

But now, we're ready to make a decision!

Step 4: Make the Decision

What is the final decision?

? Exercise 16.6.2

Should the null hypothesis be retained or rejected?

Answer

Our calculated $\chi^2=20.31$, and the critical χ^2 was 5.991, so we would reject the null hypothesis. Our calculated value is so extreme that we would expect it less than 5% of the time if there really was no pattern of relationship between the two qualitative variables.

So what would the statistical sentence look like?

? Exercise 16.6.3

What would our results look like in the statistical sentence?

Answer

$$\chi^2(2)=20.31, p<.05$$

Let's use all that we've done to let people know what we found in...

The Write-Up

Can you write this up with the [four requirements for reporting results](#) but without descriptive statistics? You can include all of the Observed frequencies, but that gets clunky. A good way around that is to refer to the original table of Observed frequencies.

✓ Example 16.6.7

Report the results in a concluding paragraph that includes the four requirements.

Solution

The research hypothesis was that there will be a pattern of difference such that there will be more people whose college decision was affected by college sports AND who watched college sports A pattern of difference was found ($\chi^2(2)=20.31, p<.05$). As can be seen in Table 16.6.1, this research hypothesis was not supported. People who watched college sports seem to believe that they used that to choose their college, and people who didn't watch college sports seem to believe that they did not use college sports to make their decision about which college to choose.

Did you notice all of the "seems like" and "they believe" in that concluding paragraph? Yeah, that's how scientists write. Because science is cumulative, each of one us adds one piece of evidence to a pile that supports one idea. In this case, the idea was that people thought that their choice of college was affected by whether they watched college sports or not. In the Goodness of Fit example, the idea that was supported was that there are about the same amount of people who like and dislike pineapples on pizza. But one study is never conclusive. Instead, many, many scientists conduct many, many studies. Some of them show reality, but some of them ($p<.05$) find results from their sample that do not match the reality in the population. It can be hard for non-scientists because they might just see us being wishy-washy about our results when we are really following the guidelines of the null hypothesis significance testing procedure.

Let's try one more example so that we've got this Chi-Square thing down.

This page titled [16.6: Practice Chi-Square Test of Independence- College Sports](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Michelle Oja](#).

- [14.7: College Sports](#) by Foster et al. is licensed [CC BY-NC-SA 4.0](#). Original source: <https://irl.umsl.edu/oer/4>.
- [14.5: Contingency Tables for Two Variables](#) by Foster et al. is licensed [CC BY-NC-SA 4.0](#). Original source: <https://irl.umsl.edu/oer/4>.
- [14.4: Pineapple on Pizza](#) by Foster et al. is licensed [CC BY-NC-SA 4.0](#). Original source: <https://irl.umsl.edu/oer/4>.
- [11.1: Chi-Square Tests for Independence](#) by Anonymous is licensed [CC BY-NC-SA 3.0](#). Original source: <https://2012books.lardbucket.org/books/beginning-statistics>.