

Section 4F – Categorical Relationships: Categorical Association Test

In this chapter, we looked at the Goodness of Fit test. The Goodness of Fit Test determines if a single proportion is related to some other categorical variable. In this section, we will look at situations where more than one proportion is involved. When we have multiple different proportions in multiple groups, we call this the Categorical Association Test.

In the Categorical Association Test, we will be determining if categorical variables are related or not. Many students confuse the Goodness of Fit test and the Categorical Association Test because they are both categorical relationship tests. Look at your sample counts. If you have a single observed count for each group, you are doing a Goodness of Fit test since we are only looking at one proportion in the groups. If your observed counts are summarized in a contingency table, then more than one proportion is involved in your groups. That makes it a Categorical Association Test.

For example, suppose we wanted to see if the amount of education a person has is related to their health. Notice the amount of education has multiple options and the health status has multiple options. This cannot be a Goodness of Fit test. The observed counts are summarized in a contingency (two-way) table so we will be using the Categorical Association Test.

	Excellent Health	Good Health	Fair Health	Poor Health
Less than High School	72	202	199	62
High School Diploma	465	877	358	108
Some College/Associates Degree	80	138	49	11
Bachelor's Degree	229	276	64	12
Graduate Degree	130	147	32	2

A Goodness of Fit Test would only look at one of these. For example, suppose we want to see if the proportion for excellent health is related to education. In that case, the data would look like this.

	Excellent Health
Less than High School	72
High School Diploma	465
Some College/Associates Degree	80
Bachelor's Degree	229
Graduate Degree	130

The Categorical Association Test

So let us look at the categorical association test. This test determines if categories are related or not. The categories can have multiple options. The sample data for this test is either two raw categorical data sets or summary counts summarized in a contingency (two-way) table.

Null and Alternative Hypothesis

In the last section, we examined conditional proportions. We saw that we need to compare conditional proportions from the same variable. If the conditional proportions are equal or close in our groups, it indicates that the categorical variables are not related. If the conditional proportions are significantly different, then the categorical variables are related.



For example, we will want to compare the proportions for excellent health in all of our education groups. We will also want to compare the proportions for poor health in all of our education groups and so on. We will not want to compare the proportion of excellent health to poor health since they are not the same variable.

	Excellent Health	Good Health	Fair Health	Poor Health
Less than High School	72	202	199	62
High School Diploma	465	877	358	108
Some College/Associates Degree	80	138	49	11
Bachelor's Degree	229	276	64	12
Graduate Degree	130	147	32	2

If there is no relationship between the categories, we expect the conditional proportions for each variable to be equal in all the groups. If there is a relationship between the categories, we expect at least one or more of the conditional proportions for each variable to be different. It is difficult to specify all of the conditional proportions and groups, so to summarize, we often say that the “distribution of conditional proportions are the same” or the “distribution of conditional proportions are different”.

Note

- Saying that the categories are “related” can also be described as “associated” or “dependent”.
- Saying that the categories are “not related” can also be described as “not associated” or “independent”.

Categorical Association Test Null and Alternative Hypothesis

H_0 : The categories are not related (distribution of conditional proportions are equal)

H_A : The categories are related (distribution of conditional proportion are different)

The Chi-Squared Test Statistic (χ^2)

Since multiple proportions in multiple groups are involved, we will be using the Chi-Squared test statistic (χ^2) again. In our previous study of using the Chi-squared test statistic, we saw that this test statistic compares the observed sample counts to the expected counts based on the null hypothesis.

$$\text{Chi-Squared Test Statistic } (\chi^2) = \sum \frac{(O-E)^2}{E}$$

The expected counts are what we expect to happen if the null hypothesis is true. If the null hypothesis is true and there is no relationship between the categories, we expect the conditional proportions to be equal in the various groups. If we multiply the equal proportions by the size of the group, we can get our expected counts. Here is a formula that computer programs use to calculate the expected counts.

$$\text{Expected Counts (for contingency table)} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$$

How does this formula give us the expected counts and account for equal conditional proportions? To answer this, we will look at an example.



Example 1

A sample of 75 people were paid to participate in an experiment. The goal of the experiment was to determine if listening to music is related to a person's ability to memorize information. The people were randomly assigned into three groups. One group tried to memorize some information while listening to their favorite music. Another group tried to memorize some information while listening to music they hated. The third group tried to memorize some information in a silent room. All of the people attempted to memorize the same information and took a test to determine how much of the information they remembered. Confounding variables were controlled. For example, the volume of music was the same in the music groups. We will use a 10% significance level.

	Liked Music	Disliked Music	No Music	Total
High Retention	10	11	18	39
Low Retention	14	15	7	36
Total	24	26	25	Grand Total = 75

H_0 : Listening to music and memorizing information are not related (not associated, independent).

H_A : Listening to music and memorizing information are related (associated, dependent). CLAIM

Remember conditional proportions are important to explore when analyzing relationships between categorical variables. Here are two very important principles.

1. When conditional proportions were close or equal, it indicated that the variables were not related to each other.
2. When conditional proportions were significantly different, it indicated that the variables were related to each other.

Let us look at some conditional proportions. Remember we need to compare the same variable proportion in different groups.

Let us calculate the proportion of people in the liked music group that were able to memorize a lot of the information?

Let us compare that to the proportion of people in the hated music group that were able to memorize a lot of the information.

$$P(\text{high retention} \mid \text{liked music}) = 10/24 = 0.417 \text{ or } 41.7\%$$

$$P(\text{high retention} \mid \text{disliked music}) = 11/26 = 0.423 \text{ or } 42.3\%$$

These two conditional probabilities are close and so indicate that the music and high retention are not related or independent.

Here lies the fundamental problem. We are not really taking the entire contingency (two-way) table and all of the conditional probabilities into account. If we look at another conditional probability, we may come to a different conclusion. Look at these two.

Let us calculate the proportion of people in the liked music group that were able to memorize a lot of the information?

Let us compare that to the proportion of people in the no music group that were able to memorize a lot of the information.

$$P(\text{high retention} \mid \text{liked music}) = 10/24 = 0.417 \text{ or } 41.7\%$$

$$P(\text{high retention} \mid \text{no music}) = 18/25 = 72\%$$

These two probabilities are significantly different and so indicate that the music and high retention are related.



So it is difficult to determine if categorical variables are related or not by just looking at two conditional proportions. We need a better way to do this.

Calculating the Chi-Squared Test Statistic

A much better way to determine if the categories are related or not is by using the chi-squared test statistic. It takes into account all of the conditional probabilities possible instead of relying on only two. Remember to calculate the chi-squared test statistic, we need to compare the expected counts (expected frequencies) to the observed counts (observed frequencies).

Expected Counts

The “expected counts” or “expected frequencies” are what we expect to happen if the null hypothesis is true. For the Categorical Association Test, the null hypothesis is that the categories are not related (independent). This would imply that the distribution of conditional proportions are equal.

Let us work this out for the music and retention problem.

	Liked Music	Disliked Music	No Music	Total
High Retention	10	11	18	39
Low Retention	14	15	7	36
Total	24	26	25	Grand Total = 75

If the null hypothesis is true, we expect the proportion for high retention to be the same regardless of the music choice. If we disregard music, then the proportion of high retention would be the amount of high retention (39) divided by the grand total (75). So if the null hypothesis is true and music is not related to retention, then 52% of every group should memorize a significant amount of the information.

$$P(\text{high retention}) = 39/75 = 0.52$$

Remember the expected values are found by multiplying the proportion times the total number of people or objects in that group.

$$E = n \times p$$

Only the n is not the grand total, it is the total for each column (each music group).

If the null hypothesis is true, we expect the p for high retention to always be 0.52 and the expected values will be 0.52 x total students for each music choice.

$$E_{\text{liked music high retention}} = n \times p = 24 \times 0.52 = 12.48$$

$$E_{\text{hated music high retention}} = n \times p = 26 \times 0.52 = 13.52$$

$$E_{\text{no music high retention}} = n \times p = 25 \times 0.52 = 13.0$$

Similarly, we expect the proportion for low retention to be the same in all of the music groups. If we disregard music, then the proportion of low retention would be the amount of low retention (36) divided by the grand total (75). So if the null hypothesis is true and music is not related to retention, then 48% of every group should not be able to memorize much of the information.

$$P(\text{low retention}) = 36/75 = 0.48$$

So if the null is true we expect the p for low retention to always be 0.48 and the expected counts will be 0.48 x total people for each music group.



$$E_{\text{liked music low retention}} = n \times p = 24 \times 0.48 = 11.52$$

$$E_{\text{hated music low retention}} = n \times p = 26 \times 0.48 = 12.48$$

$$E_{\text{no music low retention}} = n \times p = 25 \times 0.48 = 12.0$$

Earlier we saw that computer programs often use this formula to calculate the expected counts.

$$\text{Expected Counts (for contingency table)} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$$

Notice that the column total is the total number of people in each music group. The row total divided by the grand total is the proportion that must be equal for all of the groups.

$$E = n \times p = \text{Column Total} \times \left(\frac{\text{Row Total}}{\text{Grand Total}} \right)$$

Now let us calculate the Chi-Squared Test Statistic

We learned that the Chi-Squared test statistic is a comparison of the observed sample values and the expected values from the null hypothesis. Here is the formula again.

$$\text{Chi-Squared Test Statistic } (\chi^2) = \sum \frac{(O-E)^2}{E}$$

So Chi-Squared subtracts the observed and expected values to find the difference. Since some differences are negative, it squares the differences. It also divides by E to make it a kind of average of squares and finally it adds up these values for every variable.

Here is the sentence to explain Chi-Squared again:

“The sum of the averages of the squares of the differences between the observed sample data and the expected values if the null hypothesis were true.”

	Liked Music	Disliked Music	No Music	Total
High Retention	10	11	18	39
Low Retention	14	15	7	36
Total	24	26	25	Grand Total = 75

In this example, the numbers in the two-way table are the observed counts. Note: The observed counts do not include the totals! This two-way table has two rows and three columns (not counting totals). This is often called a “two by three” (2x3) table. So we have six observed counts and six expected counts.

	Liked Music	Disliked Music	No Music
High Retention	10	11	18
Low Retention	14	15	7

Let us calculate the Chi-Squared test statistic for this problem. Here are the expected counts again. It is good to label so that you subtract the correct expected count from the correct observed count.

$$E_{\text{liked music high retention}} = n \times p = 24 \times 0.52 = 12.48$$

$$E_{\text{hated music high retention}} = n \times p = 26 \times 0.52 = 13.52$$

$$E_{\text{no music high retention}} = n \times p = 25 \times 0.52 = 13.0$$



$$E_{\text{liked music low retention}} = n \times p = 24 \times 0.48 = 11.52$$

$$E_{\text{hated music low retention}} = n \times p = 26 \times 0.48 = 12.48$$

$$E_{\text{no music low retention}} = n \times p = 25 \times 0.48 = 12.0$$

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(10-12.48)^2}{12.48} + \frac{(11-13.52)^2}{13.52} + \frac{(18-13)^2}{13} + \frac{(14-11.52)^2}{11.52} + \frac{(15-12.48)^2}{12.48} + \frac{(7-12)^2}{12}$$
$$\approx 0.49282 + 0.46970 + 1.92308 + 0.53388 + 0.50885 + 2.08333 \approx 6.012$$

The numbers that were added to get the Chi-Squared test statistic are called the “Contributions to Chi-Squared”. Notice that the largest contributions to Chi-squared were 1.92308 and 2.08333. These calculations came from the low and high retention from the no music group.

Is this Chi-squared test statistic significant?

In a contingency table the degrees of freedom is the number of rows minus one times the number of columns minus one. Note: Do not include the totals when you count the number of rows and columns.

Degrees of Freedom (for Contingency Table Data) = $(r - 1) \times (c - 1)$
where “r” is the number of rows and “c” is the number of columns.

In the music and retention data, there were two rows and three columns so the degrees of freedom will be two.

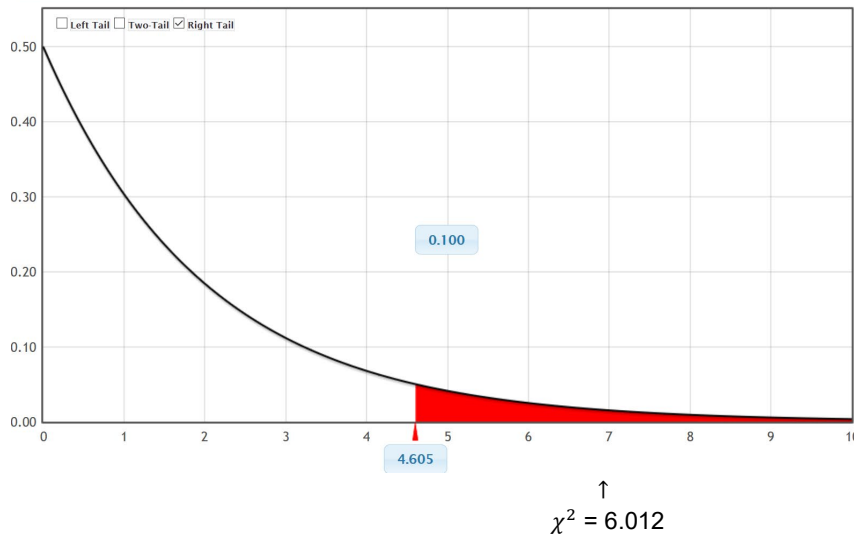
$$\text{Degrees of Freedom (for Contingency Table Data)} = (r - 1) \times (c - 1) = (2 - 1) \times (3 - 1) = 1 \times 2 = 2$$

The Categorical Association Test is a right tailed test. We can use the degrees of freedom to look up the critical value using the Chi-Squared theoretical distribution calculator in StatKey. Go to www.lock5stat.com and click on StatKey. Under the “Theoretical Distributions” menu, click on “ χ^2 ”. Put in the degrees of freedom and click “right-tail”. Since we are using a 10% significance level, we will enter 0.10 in the proportion for the right tail.



StatKey Theoretical Distribution

Chi-squared Distribution



Chi-squared Distribution

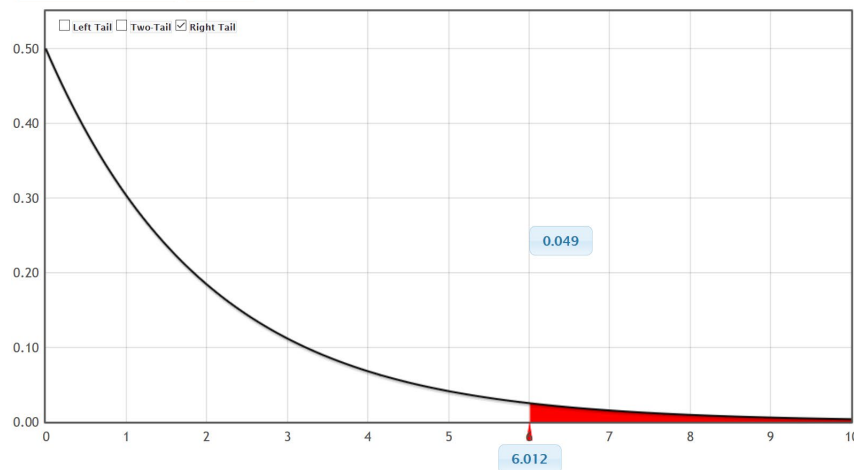
df 2

Notice that the critical value was 4.605. So our test statistic must be 4.605 or greater to be considered significant. Our Chi-squared test statistic was 6.012, which is in the right tail. This tells us that the sample data significantly disagrees with the null hypothesis. It also tells us that our observed counts are significantly different from our expected counts.

We can also use the theoretical Chi-squared curve to calculate the P-value. Just put the test statistic of 6.012 in the bottom box. Notice the computer calculated a P-value of 0.049 or 4.9%. This is less than our 10% significance level.

StatKey Theoretical Distribution

Chi-squared Distribution



Chi-squared Distribution

df 2

We can also use a randomized simulation in StatKey. Under the “More Advanced Randomization Tests” menu, click on “ χ^2 Test for Association”. We will need to type in our observed counts into StatKey. Do not include the totals. The computer will calculate the totals automatically.



	Liked Music	Disliked Music	No Music
High Retention	10	11	18
Low Retention	14	15	7

Under the “Edit Data” menu, type in the contingency table with commas. Notice that “[blank]” must be in the top left corner.

[blank], Liked Music, Disliked Music, No Music

High Retention, 10, 11, 18

Low Retention, 14, 15, 7

Edit data ✕

```
[blank], Liked Music, Disliked Music, No Music
High Retention, 10, 11, 18
Low Retention, 14, 15, 7
```

Raw Data
 Data has header row

Manually edit the values above or paste a tab or comma separated file into the box and click Ok. The file must have only two columns if it contains raw data. Summary counts tables require both row and column headers.

Ok

Under “Original Sample”, we see that StatKey has calculated the Chi-Squared test statistic of 6.012. If you click on “Show Details”, you can see the expected counts and the contributions to chi-squared.

Original Sample Show Details

$n = 75, \chi^2 = 6.012$

	Liked Music	Disliked Music	No Music	Total
High Retention	10	11	18	39
Low Retention	14	15	7	36
Total	24	26	25	75

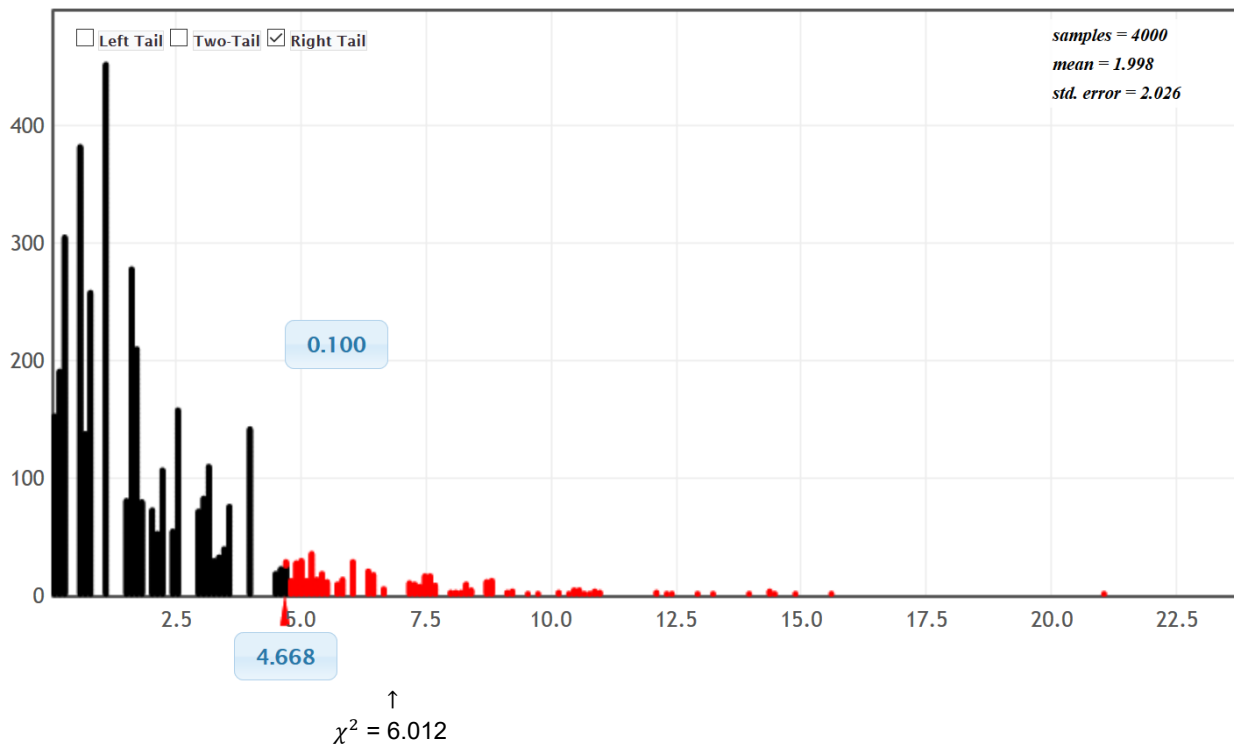


	Liked Music	Disliked Music	No Music	Total
High Retention	10 12.5 0.493	11 13.5 0.47	18 13 1.923	39
Low Retention	14 11.5 0.534	15 12.5 0.509	7 12 2.083	36
Total	24	26	25	75

Observed, Expected, Contribution to χ^2

If we click "Generate 1000 Samples" a few times, we get our randomized simulation. Notice the null hypothesis is "No Association" (not related). Putting our 10% significance level in the tail proportion, gives us an approximate critical value of 4.668. This is close to what we got with the theoretical curve. Notice that our Chi-squared test statistic of 6.012 does fall in the tail of the simulation.

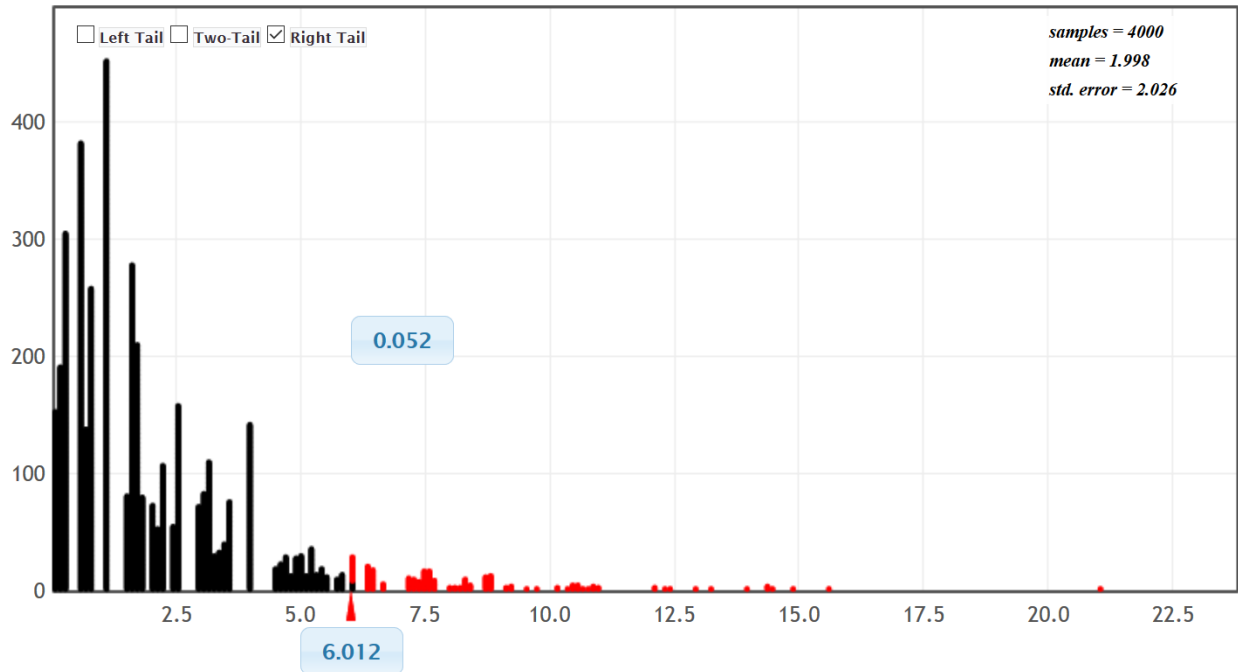
Randomization Dotplot of χ^2 , Null hypothesis: No Association



We can also use the simulation to calculate the approximate P-value. Just put the test statistic of 6.012 into the bottom box. Notice the P-value came out to be 0.052. This is almost the same P-value as we calculated with the theoretical Chi-squared distribution.



Randomization Dotplot of χ^2 , Null hypothesis: No Association



Calculating the Chi-Squared test statistic and P-value with Statcato

- First type in the contingency (two-way) table exactly as you see it into Statcato. The column titles will be in the grey cells (VAR) at the top, but the row titles will be in the regular cells. Do not type the totals. Statcato will automatically calculate them.

	C1	C2	C3	C4
Var		Liked Music	Disliked Music	No Music
1	High Retention	10	11	18
2	Low Retention	14	15	7

- Go to the “statistics” menu and click on “multinomial experiment” then “chi-square contingency table”. Hold the control key down and click on the columns that have your observed frequencies (C2, C3 & C4). Do not include the row labels (C1). They can be added later.

Chi-Square Test: Contingency Table:

	C2 Liked Music	C3 Disliked Music	C4 No Music	Total
High Retention	10.0 (12.48) [0.49]	11.0 (13.52) [0.47]	18.0 (13.0) [1.92]	39.0
Low Retention	14.0 (11.52) [0.53]	15.0 (12.48) [0.51]	7.0 (12.0) [2.08]	36.0
Total	24.0	26.0	25.0	75.0

(expected frequency), [test statistic contribution]

Significance Level	DOF	χ^2	Critical value	p-Value
0.05	2	6.0117	5.9915	0.0495



Notice the test statistic, critical value and P-value in Statcato are about the same as we got in the simulation.

The expected counts (expected frequencies) are in parenthesis in the Statcato printout and the contributions to chi-squared are in brackets.

Assumptions

The assumptions for the Categorical Association Test are as follows:

Categorical Association Test Assumptions

- The categorical sample or samples should be collected randomly or be representative of the population.
- Data values within each sample should be independent of each other.
- If multiple samples are collected, then the data values between the samples should be independent of each other.
- The expected counts from the null hypothesis should be at least five.

Did the music and retention data meet the assumptions?

Random? Yes. This was an experiment and the groups were randomly assigned.

Were the expected counts at least 5? Yes. The expected counts were 12.4, 13.52, 13.0, 11.52, 12.48, and 12.0. All the expected frequencies were at least five.

Independence? The individuals in the experiment were randomly assigned, so we can probably assume the data met the independence requirement.

Writing the conclusion

Should we reject the null hypothesis or fail to reject the null hypothesis? Since our P-value is lower than our significance level, we reject the null hypothesis.

Remember a conclusion must address the evidence and the claim. Since our P-value is low, we have significant evidence. We rejected the null hypothesis, but the claim was the alternative hypothesis. We will therefore support the claim.

H_0 : Listening to music and memorizing information are not related (not associated, independent).

H_A : Listening to music and memorizing information are related (associated, dependent). CLAIM

Conclusion: We have significant sample evidence to support the claim that listening to music and retaining information are related.

In fact, since confounding variables were controlled, the experiment proves cause and effect. The “no music” group did significantly better than either of the music groups. This experiment appears to indicate that when a person listens to either music they like or music they hate, they have a harder time memorizing information.

Note about “Independence” and “Homogeneity”

Sometimes we obtain multiple categorical data from one random sample of people or objects. We ask multiple categorical questions from the same group of people. Statisticians sometimes refer to that situation as an “independence” test.

If we collect the same categorical data from multiple random samples, then that is sometimes referred to as a “homogeneity” test. We may ask the same categorical question from different groups of people.



Whether you collected the data from one sample or multiple samples, the data still summarizes into a contingency table. Look at the following sample data.

	Business	English	History	Music	Biology	Math
Female	89	71	62	48	56	9
Male	112	58	59	53	62	13

Suppose we took one random sample of college students and asked them two categorical questions. What gender do you most identify with? What is your major? This would be referred to as an “independence” test.

We could collect this data in another way. Suppose we took a random sample of female college students and asked them what their major was. Later we took a random sample of male college students and asked them the same question. This would now be referred to as a “homogeneity” test.

Notes about the Categorical Association Test

- The Categorical Association Test is used to determine if categories with multiple options are related or not.
 - The categorical association test is always a right-tailed test.
 - The degrees of freedom for the categorical association test is the number of rows minus one times the number of columns minus one. $df = (r - 1) \times (c - 1)$
 - The categorical association test uses the Chi-squared test statistic (χ^2), which compares the observed sample counts to the expected counts if the null hypothesis was true.
 - Always use a computer to calculate the test statistic. Focus on being able to interpret and judge significance.
-

