

CHI-SQUARE TEST FOR INDEPENDENCE

Dr. Bibha Verma
PG Department of Zoology
Maharaja College Ara

PG Sem II, CC-6, Unit -III

Chi-Square Test for Independence

This lesson explains how to conduct a **chi-square test for independence**. The test is applied when you have two **categorical variables** from a single population. It is used to determine whether there is a significant association between the two variables.

For example, in an election survey, voters might be classified by gender (male or female) and voting preference (Democrat, Republican, or Independent). We could use a chi-square test for independence to determine whether gender is related to voting preference. The **sample problem** at the end of the lesson considers this example.

When to Use Chi-Square Test for Independence

The test procedure described in this lesson is appropriate when the following conditions are met:

- The sampling method is **simple random sampling**.
- The variables under study are each **categorical**.
- If sample data are displayed in a **contingency table**, the expected frequency count for each cell of the table is at least 5.

This approach consists of four steps: (1) state the hypotheses, (2) formulate an analysis plan, (3) analyze sample data, and (4) interpret results.

State the Hypotheses

Suppose that Variable A has r levels, and Variable B has c levels. The [null hypothesis](#) states that knowing the level of Variable A does not help you predict the level of Variable B. That is, the variables are independent.

H_0 : Variable A and Variable B are independent.

H_a : Variable A and Variable B are not independent.

The [alternative hypothesis](#) is that knowing the level of Variable A *can* help you predict the level of Variable B.

Note: Support for the alternative hypothesis suggests that the variables are related; but the relationship is not necessarily causal, in the sense that one variable "causes" the other.

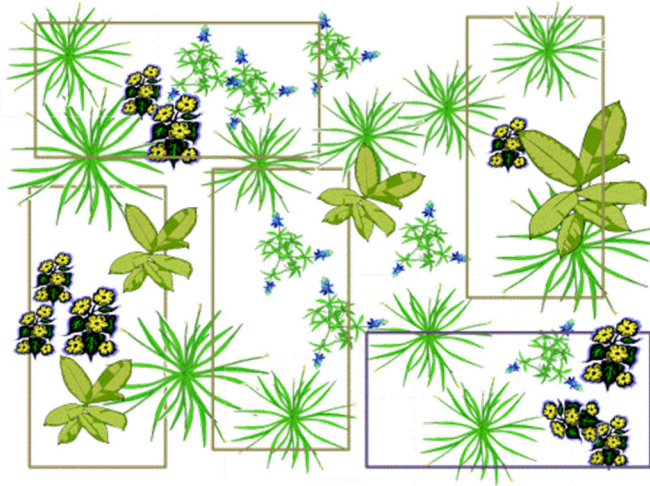
Formulate an Analysis Plan

The analysis plan describes how to use sample data to accept or reject the null hypothesis. The plan should specify the following elements.

- Significance level. Often, researchers choose [significance levels](#) equal to 0.01, 0.05, or 0.10; but any value between 0 and 1 can be used.
- Test method. Use the [chi-square test for independence](#) to determine whether there is a significant relationship between two categorical variables.

Chi-Square (X^2) Test for Independence

Chi-square Test for Independence is a statistical test commonly used to determine if there is a significant association between two variables. For example, a biologist might want to determine if two species of organisms associate (are found together) in a community.



Does Species A associate with Species B?



Species A



Species B

Just like other statistical tests, the Chi-Square Test for Independence tests two hypotheses:

Null Hypothesis:

"There is not a significant association between variables, the variables are independent of each other; any association between variables is likely due to chance and sampling error."

For example:

- There is no significant association between Species A and Species B; the species are independent of each other. The location of Species A has no effect on the location of Species B.

Alternative Hypothesis:

"There is a significant (positive or negative) association between variables; the association between variables is likely not due to chance or sampling error."

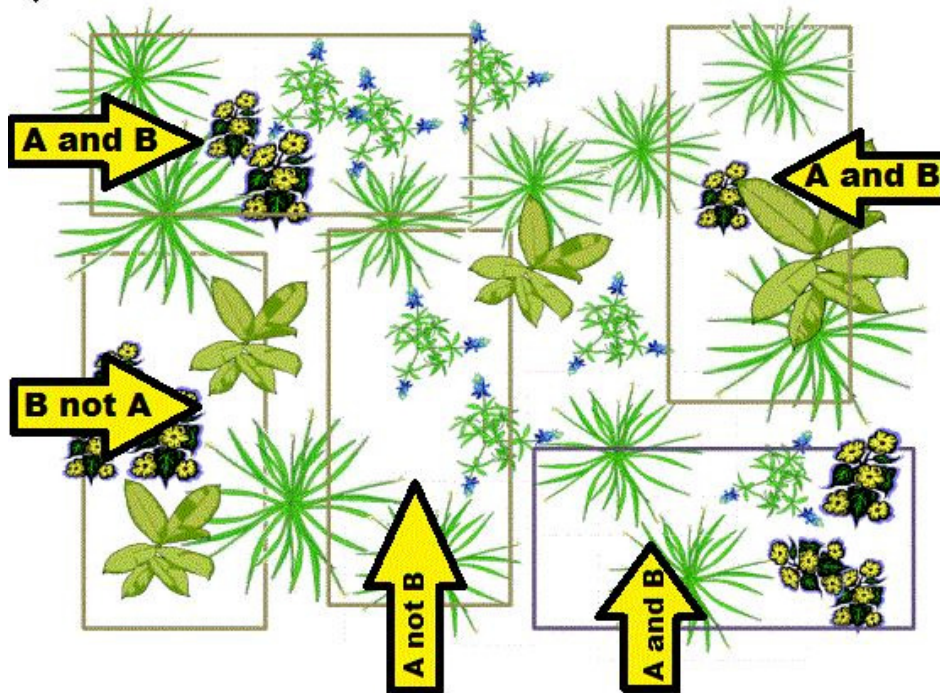
For example:

- There is a significant association between Species A and Species B; the species are dependent. Either Species A significantly associates with Species B **or** Species A does not significantly associate with Species B.

How to Calculate a Chi-Square Test of Independence

The first step is to collect raw data for the occurrence of each variable. This is often done via random sampling using a quadrant. In our example, there are five quadrants. Determine:

- The number of quadrants with both species present
- The number of quadrants with Species A but not Species B
- The number of quadrants with Species B but not Species A
- The number of quadrants with neither species



Then create a "contingency table" to display your results. In the Chi-Square test, these are your OBSERVED values.

OBSERVED Number of Quadrants with Species

		Species B		
		Present	Absent	Total
Species A	Present	3	1	4
	Absent	1	0	1
	Total	4	1	5

Next you need to determine what would be EXPECTED assuming the species are randomly distributed with respect to each other. **Expected frequencies = (row total X column total) / grand total**

EXPECTED Number of Quadrants with Species

		Species B		
		Present	Absent	Total
Species A	Present	= (4 X 4) / 5	= (4 X 1) / 5	4
	Absent	= (4 X 1) / 5	= (1 X 1) / 5	1
	Total	4	1	5

EXPECTED Number of Quadrants with Species

		Species B		
		Present	Absent	Total
Species A	Present	3.2	0.8	4
	Absent	0.8	0.2	1
	Total	4	1	5

Now that you have OBSERVED and EXPECTED values, apply the Chi-Square formula in each part of the contingency table by determining **(O-E)² / E** for each box.

CHI-SQUARE CALCULATIONS

		Species B		
		Present	Absent	Total
		$= (O - E)^2 / E$		
Species A	Present	= (3 - 3.2) ² / 3.2	= (1 - 0.8) ² / 0.8	
	Absent	= (1 - 0.8) ² / 0.8	= (0 - 0.2) ² / 0.2	
	Total			

		CHI-SQUARE CALCULATIONS		
		<i>Rounded to tenths spot</i>		
		Species B		
		Present	Absent	Total
Species A	Present	0.0	0.1	
	Absent	0.1	0.2	

$$X^2 = \sum \frac{(o - e)^2}{e}$$

The final calculated chi-square value is determined by summing the values:

- $X^2 = 0.0 + 0.1 + 0.1 + 0.2 = \mathbf{0.4}$

References:

<https://stattrek.com/chi-square-test/independence.aspx>

<https://www.biologyforlife.com/x2-test-for-independence.html>