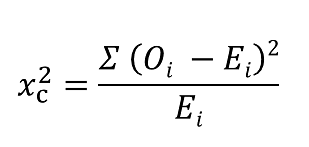
**CHI-SQUARE TEST**

The Chi-Square test is a statistical procedure for determining the **difference between observed and expected data**. This test can also be used to decide whether it correlates to our data's categorical variables. It helps to determine whether a difference between two categorical variables is due to chance or a relationship between them.

A chi-square test or comparable nonparametric test is required to test a hypothesis regarding the distribution of a **categorical variable.** Categorical variables, which indicate categories such as animals or countries, can be nominal or ordinal. They cannot have a normal distribution since they only have a few particular values.

Chi-Square Test Formula



Where

c = Degrees of freedom

O = Observed Value

E = Expected Value

# Types of Chi-Square Tests

1. **Independence**
2. **Goodness-of-Fit**

# Independence

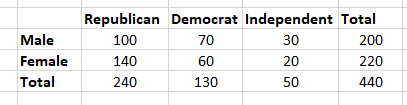
The Chi-Square Test of Independence is a derivable ( also known as inferential ) statistical test which examines whether the two sets of variables are likely to be related with each other or not. This test is used when we have counts of values for two nominal or categorical variables and is considered as non-parametric test.

# Goodness-Of-Fit

In statistical hypothesis testing, the Chi-Square Goodness-of-Fit test determines whether a variable is likely to come from a given distribution or not. We must have a set of data values and the idea of the distribution of this data.

# **How to Perform a Chi-Square Test?**

Let's say you want to know if gender has anything to do with political party preference. You poll 440 voters in a simple random sample to find out which political party they prefer. The results of the survey are shown in the table below:



To see if gender is linked to political party preference, perform a Chi-Square test of independence using the steps below.

Step 1: Define the Hypothesis

H0: There is no link between gender and political party preference.

H1: There is a link between gender and political party preference.

Step 2: Calculate the Expected Values

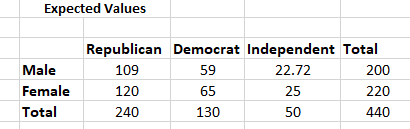
Now you will calculate the expected frequency.

Chi_Sq_formula_1.

For example, the expected value for Male Republicans is:

Chi_Sq_formula_2

Similarly, you can calculate the expected value for each of the cells.



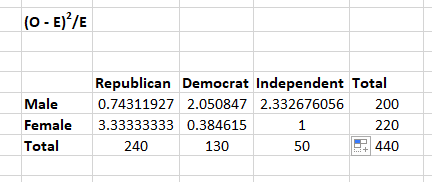
Step 3: Calculate (O-E)2 / E for Each Cell in the Table

Now you will calculate the (O - E)2 / E for each cell in the table.

Where

O = Observed Value

E = Expected Value



Step 4: Calculate the Test Statistic X2

X2  is the sum of all the values in the last table

 =  0.743 + 2.05 + 2.33 + 3.33 + 0.384 + 1

 = 9.837

Before you can conclude, you must first determine the critical statistic, which requires determining our degrees of freedom. The degrees of freedom in this case are equal to the table's number of columns minus one multiplied by the table's number of rows minus one, or (r-1) (c-1). We have (3-1)(2-1) = 2.

Finally, you compare our obtained statistic to the critical statistic found in the chi-square table. As you can see, for an alpha level of 0.05 and two degrees of freedom, the critical statistic is 5.991, which is less than our obtained statistic of 9.83. You can reject our null hypothesis because the critical statistic is higher than your obtained statistic.

**This means you have sufficient evidence to say that there is an association between gender and political party preference.**

