Chi-Squared-test

The chi-square test, also written as chi-squared test (χ^2 test), is a fundamental tool in statistics used to assess the relationship between two categorical variables

There are two types of Pearson's chi-square tests:

- The chi-square goodness of fit test is used to test whether the frequency distribution of a categorical variable is different from your expectations.
- The chi-square test of independence is used to test whether two categorical variables are related to each other.

The chi-square formula

Both of Pearson's chi-square tests use the same formula to calculate the test statistic, chi-square (X2):

$$\chi^2 = \Sigma[(O - E)^2 / E]$$

Where:

- X is the chi-square test statistic

  2

- Σ is the summation operator (it means "take the sum of")
- *O* is the observed frequency
- *E* is the expected frequency

When to use a chi-square test

A Pearson's chi-square test may be an appropriate option for your data if **all** of the following are true:

1. You want to test a hypothesis about **one or more categorical variables**.
2. The **sample was randomly/**
3. There are a **minimum of five observations expected** in each group or combination of groups.

Types of chi-square tests

The two types of Pearson's chi-square tests are:

1. Chi-square goodness of fit test

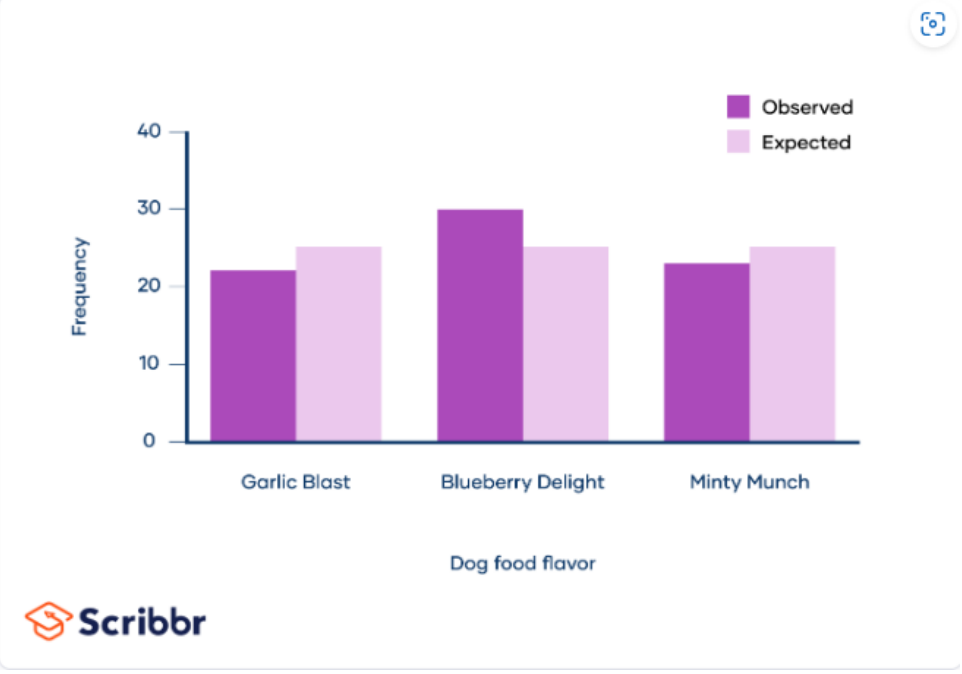
You can use it to test whether the observed distribution of a one categorical variable differs from your expectations.

Example:

You're hired by a dog food company to help them test three new dog food flavors.

You recruit a random sample of 75 dogs and offer each dog a choice between the three flavors by placing bowls in front of them. You expect that the flavors will be equally popular among the dogs, with about 25 dogs choosing each flavor.

- **Null hypothesis (*H*0):** The dog population chooses the three flavors in equal proportions (*p1* = *p2* = *p3*).
- **Alternative hypothesis (*Ha*):** The dog population does not choose the three flavors in equal proportions.



2. Chi-square test of independence

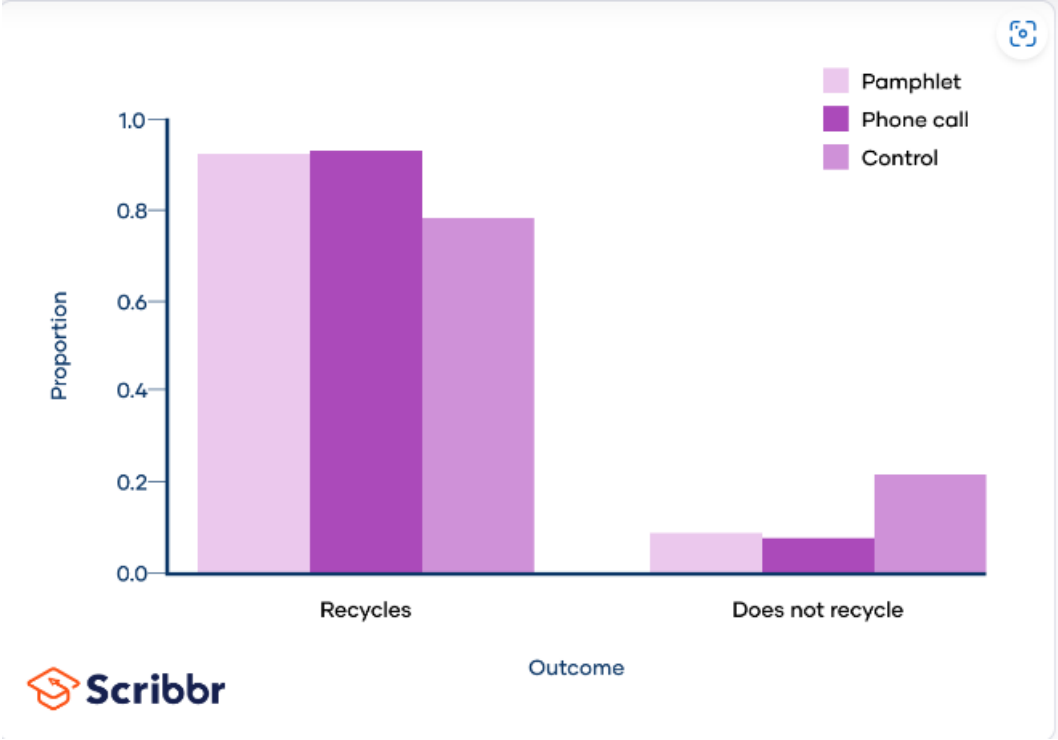
You can use it to test whether two categorical variables are related to each other.

Example:

Imagine a city wants to encourage more of its residents to recycle their household waste.

The city decides to test two interventions: an educational flyer(pamphlet) or a phone call. They randomly select 300 households and randomly assign them to the flyer, phone call, or control group (no intervention). They'll use the results of their experiment to decide which intervention to use for the whole city.

The city plans to use a chi-square test of independence to test whether the proportion of households who recycle differs between the interventions.



Chi-distribution

When you square a value from a standard normal distribution, you get its squared distance from the mean (0) If you take multiple independent standard normal variables (let's say k of them) and square each one, then sum those squares, you get a chi-square distribution with k degrees of freedom.