Chi-Square Test Notes

Overview

The chi-squared (χ^2) test is a statistical method used to determine whether there is a significant difference between two categorical variables or if a sample distribution fits an expected distribution. It compares the observed frequencies in each category to the expected frequencies, assuming no relationship between the variables.

The test is commonly applied in hypothesis testing, where a low p-value indicates that the observed data significantly deviates from the expected values. The chi-squared test is widely used in fields like social sciences, genetics, and market research to analyze contingency tables and goodness-of-fit models.

In a Chi-square test, the null and alternative hypotheses are formulated based on the type of Chi-square test you're conducting—either a *goodness-of-fit test* or a *test of independence*.

Chi-Square Goodness-of-Fit Test (a.k.a. 1-way chi-squared test)

Null Hypothesis (H_0): The observed data fits the expected distribution. In other words, there is no significant difference between the observed and expected frequencies.

Alternative Hypothesis (H_a): The observed data does not fit the expected distribution, implying a significant difference between observed and expected frequencies.

Notes:

(1) The “goodness of fit” term emphasizes the idea of comparing the observed data to a theoretical or expected distribution (i.e. observed frequencies vs. expected frequencies).
(2) The “1-way” term indicates that there is only one categorical variable being analyzed.

Chi-Square Test of Independence (a.k.a. 2-way chi-squared test)

Null Hypothesis (H_0): The two categorical variables are independent of each other. There is no association between the variables.

Alternative Hypothesis (H_a): The two categorical variables are dependent, meaning there is a significant association between them.

Notes:

(1) The “test of independence” or “2-way” chi-squared test is applied to a contingency table.
(2) “Independence” term emphasizes the idea of testing whether the two variables are independent of each other.

(3) “2-way” term indicates that there are two categorical variables being analyzed simultaneously.

In both cases, the goal is to assess whether the observed data aligns with the null hypothesis. If the Chi-square test statistic exceeds a critical value, the null hypothesis is rejected in favor of the alternative.

The chi-square test is used when a parametric test (that assumes normality in the population) can not be used. Data will be categorical in nature (i.e. nominal or ordinal (ranked)). Chi-squared is one of many nonparametric tests of significance that does not require normality.

—----------

1-way Chi-squared test (a.k.a. “Goodness-of-fit” test) Example

A	12
B	14
C	9
D	5
E	10

These are the observed frequencies from a test key – the number of times each option was the correct answer on the test. **“Observed Frequencies”** is the set of frequencies that are actually obtained via research.

With a quick view, you might conclude that the professor who made the test favors putting the answer near the top as A or B. Also, the professor seems to shy away from D.

Can we generalize about the tendencies of this professor observed from one exam key?

To answer this, we need a test for frequencies, not means like parametric tests (t-test, ANOVA).

The Chi-Square test is the one for this!

The 1-way Chi-Square test can be used to determine whether the frequencies we observed previously differ significantly from an even distribution (or any other distribution we might hypothesize).

Null Hypothesis: *The professor shows no tendency to assign any particular correct response from A to E.*

Alternative Hypothesis: *The professor shows a tendency to assign particular correct responses from A to E.*

What would the frequency distribution of correct responses look like if the null hypothesis were true? Since there are 5 categories, 20% of the correct responses should fall in each. So, with 50 question, the **"Expected Frequencies"** for the hypothesized even distribution should look like this:

A 10
B 10
C 10
D 10
E 10

The Chi-square test allows us to test the significance between a set of observed frequencies (*fo*) and expected frequencies (*fe*).

$$\chi^2 = \Sigma \frac{(f_o - f_e)^2}{f_e}$$

The following table shows the computations:

Category	fo	fe	fo - fe	(fo - fe)^2	[(fo - fe)^2]/fe
A	12	10	2	4	0.4
B	14	10	4	16	1.6
C	9	10	-1	1	0.1
D	5	10	-5	25	2.5
E	10	10	0	0	0
				CHI-SQUARE	**4.6**

To interpret this chi-square value of 4.6 we need to find the degrees of freedom.

df = k - 1 where k = number of categories in the observed frequency distribution

df = 5 - 1 = 4

The next step is to look up this df = 4 in a chi-square table. Here are a few sites where you can input the df and α = .05 to determine the critical chi-square value.

https://www.socscistatistics.com/tests/criticalvalues/default.aspx

https://www.danielsoper.com/statcalc/calculator.aspx?id=12

https://www.omnicalculator.com/statistics/critical-value

At α = .05, the critical value of chi-square with df = 4 is 9.488.

This is the value that we must exceed before we can reject the null hypothesis.

Since chi-square calculated (4.6) is less than chi-square critical (9.488), we will fail to reject the null hypothesis. The observed frequencies are not different enough from the expected frequencies. As such, the professor shows no tendency to assign any particular correct response from A to E.

Two-way Chi-Square (Test for Independence) Example

Just like the 1-way chi-square, the 2-way chi-square test of significance is concerned with the distinction between expected frequencies (fe) and observed frequencies (fo). Observed frequencies (fo) are the results we actually get from the survey. The expected frequencies refers to the terms of the null hypothesis, that the relative frequencies are supposed to be the same.

Example: Is there a difference between male and female attitudes towards recycling?

Null Hypothesis: *The relative frequency of males attitudes about recycling is the same as the relative frequency of female attitudes towards recycling.*

Alternative Hypothesis: *The relative frequency of males attitudes about recycling is not the same as the relative frequency of female attitudes towards recycling.*

In this study, we surveyed 20 men and 20 women, the results are shown below.

	Female	Male	Total
Recycling Important	15	10	25
Recycling Not Important	5	10	15
Total	20	20	N = 40

Here is the “contingency table” or “cross-tabulation table.” It contains the observed frequencies as well as the expected frequencies for each outcome.

	Female	Male	Total
Recycling Important	15 (12.5)	10 (12.5)	25
Recycling Not Important	5 (7.5)	10 (7.5)	15
Total	20	20	N = 40

In this table, the expected frequencies are in ( ). The expected frequencies are calculated as follows:

fe = (row marginal total)*(column marginal total)/N

So, for the expected frequency of “important recycling” females …

fe = (25)(20)/40 = 12.5

Recall that the null hypothesis assumes the expected frequencies are the same, so this would be the expected frequency for females that regard recycling as important and as not important.

Here’s the fe calculation for “not important recycling” males:

fe = (15)(20)/40 = 7.5

The chi-square statistic is found the same way as we did for the 1-way chi-square test.

$$\chi^2 = \Sigma \frac{(f_o - f_e)^2}{f_e}$$

Cell	fo	fe	fo - fe	(fo - fe)^2	[(fo - fe)^2]/fe
Female - Recycling Important	15	12.5	2.5	6.25	0.5
Female - Recycling Not Important	5	7.5	-2.5	6.25	0.83
Male - Recycling Important	10	12.5	-2.5	6.25	0.50
Male - Recycling Not Important	10	7.5	2.5	6.25	0.83
				CHI-SQUARE	2.67

To interpret this chi-square results, we need to know the degrees of freedom. The df and our level of significance will be used to determine the chi-square critical value from a chi-square table.

For a 2-way chi-square test, the degrees of freedom are calculated as follows:

$df = (r - 1)(k - 1)$

Where r = number or rows, and k = number of columns.

There are 2 rows and 2 columns so the degrees of freedom will be: $df = (2 -1)(2 -1) = 1$.

With df = 1 and $\alpha = .05$, our chi-square critical value is 3.84 (found via chi-square table).

Since our chi-square calculated value of 2.67 is less than the chi-square critical value of 3.84, we will fail to reject the null hypothesis. *The relative frequency of males attitudes about recycling are the same as the relative frequency of female attitudes towards recycling.*