

16.1: Introduction to Chi-Square

I don't know about you, but I am TIRED. We've learned SO MUCH.

Can you stay with me for one more chapter, though? See, we've covered the appropriate analyses when we have means for different groups and when we have two different quantitative variables. We've also briefly covered when we have ranks or medians for different groups, and when we have two binary or ranked variables. But what we haven't talked about yet is when we only have qualitative variables. When we have things with names, and all that we can do is count them. For those types of situations, the Chi-Square (χ^2) analysis steps in! It's pronounced like "kite" not like "Chicago" or "chai tea".

Let's practice a little to remind ourselves about qualitative and quantitative variables; it's been a minute since we first introduced these [types of variables](#) (and [scales of measurement](#))!

? Exercise 16.1.1

What type is each of the following? Qualitative or Quantitative?

1. Hair color
2. Ounces of vodka
3. Type of computer (PC or Mac)
4. MPG (miles per gallon)
5. Type of music

Answer

1. Hair color: Qualitative (it's a quality, a name, not a number)
2. Ounces of vodka: Quantitative (it's a number that measures something)
3. Type of computer (PC or Mac): Qualitative
4. MPG: Quantitative
5. Type of music: Qualitative

? Exercise 16.1.2

Do you use means to find the average of qualitative or quantitative variables?

Answer

Quantitative. Means are mathematical averages, so the variable has to be a number that measures something.

Instead of means, you use counts with qualitative variables.

Frequency counts: Counts of how many things are in each level of the categories.

Introducing Chi-Square

Our data for the χ^2 test (the chi is a weird-looking X) are quantitative, (also known as nominal) variables. Recall from our discussion of scales of measurement that nominal variables have no specified order (no ranks) and can only be described by their names and the frequencies with which they occur in the dataset. Thus, we can only count how many "things" are in each category. Unlike our other variables that we have tested, we cannot describe our data for the χ^2 test using means and standard deviations. Instead, we will use frequencies tables.

Table 16.1.1: Pet Preferences

	Cat	Dog	Other	Total
Observed	14	17	5	36
Expected	12	12	12	36

Table 16.1.1 gives an example of a contingency table used for a χ^2 test. The columns represent the different categories within our single variable, which in this example is pet preference. The χ^2 test can assess as few as two categories, and there is no technical upper limit on how many categories can be included in our variable, although, as with ANOVA, having too many categories makes interpretation difficult. The final column in the table is the total number of observations, or N . The χ^2 test assumes that each observation comes from only one person and that each person will provide only one observation, so our total observations will always equal our sample size.

There are two rows in this table. The first row gives the observed frequencies of each category from our dataset; in this example, 14 people reported liking preferring cats as pets, 17 people reported preferring dogs, and 5 people reported a different animal. This is our actual data. The second row gives expected values; expected values are what *would* be found if each category had *equal* representation. The calculation for an expected value is:

$$E = \frac{N}{C}$$

Where N is the total number of people in our sample and C is the number of categories in our variable (also the number of columns in our table). Thank the Higher Power of Statistics, formulas with symbols that finally mean something! The expected values correspond to the null hypothesis for χ^2 tests: equal representation of categories. Our first of two χ^2 tests, the Goodness-of-Fit test, will assess how well our data lines up with, or deviates from, this assumption.

This page titled [16.1: Introduction to Chi-Square](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Michelle Oja](#).

- [14.1: Categories and Frequency Tables](#) by [Foster et al.](#) is licensed [CC BY-NC-SA 4.0](#). Original source: <https://irl.umsl.edu/oer/4>.