

14.1: Categories and Frequency Tables

Our data for the χ^2 test are categorical, specifically nominal, variables. Recall from unit 1 that nominal variables have no specified order and can only be described by their names and the frequencies with which they occur in the dataset. Thus, 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 14.1.1: Pet Preferences

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

Table 14.1.1 gives an example of a frequency 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 our computations long and our 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. 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:

$$E = \frac{N}{C} \quad (14.1.1)$$

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). 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 [14.1: Categories and Frequency Tables](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Foster et al. \(University of Missouri's Affordable and Open Access Educational Resources Initiative\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.