Chi-Square Goodness of Fit Test

Suppose that you manage an academic support service for high school students. The district leadership has been skeptical about adopting your model because they claim that your students (we'll call them tutored students) do not represent the general population of students in the district high schools. To challenge their claim, you need to compare your students with the general characteristics of students in the district. You could conduct a number of independent samples *t* tests on performance measures (e.g., CAHSEE scores), but how do you compare characteristics, such as gender, SES, ethnicity, and mobility?

Let's start with gender. The district high school student gender composition is 65% female and 35% male. You have been helping 14 female students and 10 male students. How similar or different are these two groups regarding gender composition? To answer this question, you should conduct a chi-square goodness of fit test (or one-sample chi-square test). The specifics for this test are as follows. First, you are interested in a nominal variable (i.e., gender). Second, you have both a current group and a comparison group. In this example, the current group is the 24 students using your service (14 girls, 10 boys), and the comparison group is the total district high school student population, which you know to be 65% female and 35% male. Third, based on these two groups, you compare the two sets of frequencies. Two very important concepts are involved: observed frequency and expected frequency. Observed frequency is the count of observations for the current group - in this case, the 24 students. Expected frequency is the count of frequencies in the comparison group - in this case, all high school students. The comparison of observed frequencies and expected frequencies is called chi-square analysis. This type of analysis is often displayed using a cross-tabulation (or crosstab) table. Here is the start of the table for this analysis.

	Female	Male	Total
Tutored Students	14	10	24
High School Students			

The numbers in the row for Tutored Students are the **observed** frequencies. What are the expected frequencies? Using the percentages 65% and 35% we can construct expected frequencies to use for comparison. For the girls, 65% of 24 is 15.6 and for the boys 35% of 24 is 8.4. These numbers are used in the row for High School Students.

	Female	Male	Total
Tutored Students	14	10	24
High School Students	15.6	8.4	24

Obviously, the frequencies are different, but how important is the difference? To answer this question, we use a familiar technique - we find the difference between the Observed and Expected frequencies (called a residual), square each of those differences, divided each squared difference by the Expected frequency, and then add up the results. This technique should sound vaguely similar to the calculation of the variance. Residual is another synonym for deviation.

For this example, the residuals are shown below:

	Female	Male
Residuals	-1.6	1.6

Squaring these residuals and dividing by 15.6 and 8.4, respectively, results in the following:

	Female	Male	Total
Squared Residuals/ Expected Frequency	.164	.305	.469

The value in the **Total** column is called X^2 (chi-square, pronounced ki as in kite). This is the test statistic for the comparison of the observed and expected frequencies. As with other test statistics, we compare the obtained value with the critical value to determine whether to reject or retain the null hypothesis.

For a chi-square test, the null hypothesis is that the two sets of frequencies (i.e., observed and expected) are equal. The alternative hypothesis is that they are unequal.

The closer the obtained chi-square is to zero, the more similar the two sets of frequencies are - or, stated another way, the better the observed data fit the expected pattern. This interpretation is where the term "goodness of fit" originates.

Simply put, higher the value of chi-square is, higher is the difference between two data sets.