The Chi-square Goodness-of-Fit Test -- Analysis of a Single Qualitative Variable

Application: To test a hypothesis about the distribution of subjects across the categories of a single qualitative variable.

H0: The population represented by the sample has the expected distribution across the categories of the variable.

To reject H0: is to say that the distribution of scores is somehow different from the expected distribution.

The data: The researcher wanted to investigate the distribution of the sample across the types of stores. The variable for this analysis is **chain** (1=chain store, 2=privately owned and 3=coop store The values for each of the 12 stores is shown below.

3 3 3 3 1 1 2 1 2 1 2 1

Research Hypothesis (it is often the case that the researcher's hypothesis is the statistical null, something more common for this statistical model than for most others): The researcher expected that 3/4 (75%) of the stores would be private and 1/8 (12.5%) each chain and coop.

H0: for this analysis : Pet stores will be distributed as 12.5% chain stores, 75% private stores and 12.5% coop stores.

Step 1 Rearrange the data in to a table having as many cells are there are categories of the variable being analyzed. Each cell should contain the frequency of that category. These are called the obtained frequency (**of**) of each category.

Chain Stores	Private Stores	Coop Stores
5	3	4

Step 2 Determine the number of subjects in the sample

N = 12

Step 3 Compute the expected frequency (**ef**) for each category, using the following formula. Be sure to use the expected **proportion** of cases for each category, not the percentage (e.g., 75% = a proportion of .75) For example, Chain stores are expected to make up 12.5% of the sample.

ef = N * expected % for that category = 12 * .125 = 1.5

Chain Stores		Private Stores		Coop Stores	
of	ef	of	ef	of	ef
5	1.5	3	9	4	1.5

Step 4 Compute Chi-square

$$X^2 = \sum \frac{(of - ef)^2}{ef} = \frac{(5 - 1.5)^2}{1.5} + \frac{(3 - 9)^2}{9} + \frac{(4 - 1.5)^2}{1.5} = 8.17 + 4 + 4.17 = 16.34$$

Step 5 Compute the degrees of freedom (df)

df = (number of columns - 1) = (3-1) = 2

Step 7 Use the Chi-square table to determine the critical Chi-square value for df = 2 and $\alpha = .05$

$$X^2(df=2, \alpha = .05) = 5.991$$

Step 8 Compare the obtained X^2 and critical X^2, and determine whether or not there is a statistically significant relationship between the two categorical variables.

-- if the obtained X^2 is less than the critical X^2, then retain the null hypothesis -- conclude that the distribution of scores across the conditions of the variable was as expected.

-- if the obtained X^2 is greater than the critical X^2, then reject the null hypothesis -- conclude that the distribution of scores across the conditions of the variable was different than was expected.

For the example data, we would decide to reject the null hypothesis, because the obtained Chi-square value of 16.23 is larger than the critical Chi-square value of 5.991.

By the way: This test should only be applied when at least 80% of the cells have expected frequencies (**ef**) of five or larger. Applying the test when there are fewer cells with this minimum expected frequency can lead to inaccurate results.

Step 9 Determine whether the obtained distribution, completely supports, partially supports, or does not support the research hypothesis. It is possible to find only partial support for the expected distribution of scores even when the H0: has been rejected.

Step 10 Reporting the results

It is important to describe the obtained distribution and the expected distribution, as well as the results of the significance test.

There were five chain, three private, and four coop stores. This distribution was significantly different from the hypothesized distribution, which was 3/4 private, and 1/8 each coop and chain, $\underline{X}^2(2) = 16.34$, $\underline{p} < .05$.

Critical values of C^2

df	α = .05	α = .01
1	3.84	6.63
2	5.99	9.21
3	7.81	11.34
4	9.49	13.28
5	11.07	15.09
6	12.59	16.81
7	14.07	18.48
8	15.51	20.09
9	16.92	21.67
10	18.31	23.21
11	19.68	24.72
12	21.03	26.22
13	22.36	27.69
14	23.68	29.14
15	25.00	30.58
16	26.30	32.00
17	27.59	33.41
18	28.87	34.81
19	30.14	36.19
20	31.41	37.57
21	32.67	38.93
22	33.92	40.29
23	35.17	41.64
24	36.42	42.98
25	37.65	44.31
26	38.89	45.64
27	40.11	46.96
28	41.34	48.28
29	42.56	49.59
¥	43.77	50.89