

Goodness-of-fit Chi-square -- Analysis of a Single Qualitative Variable

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

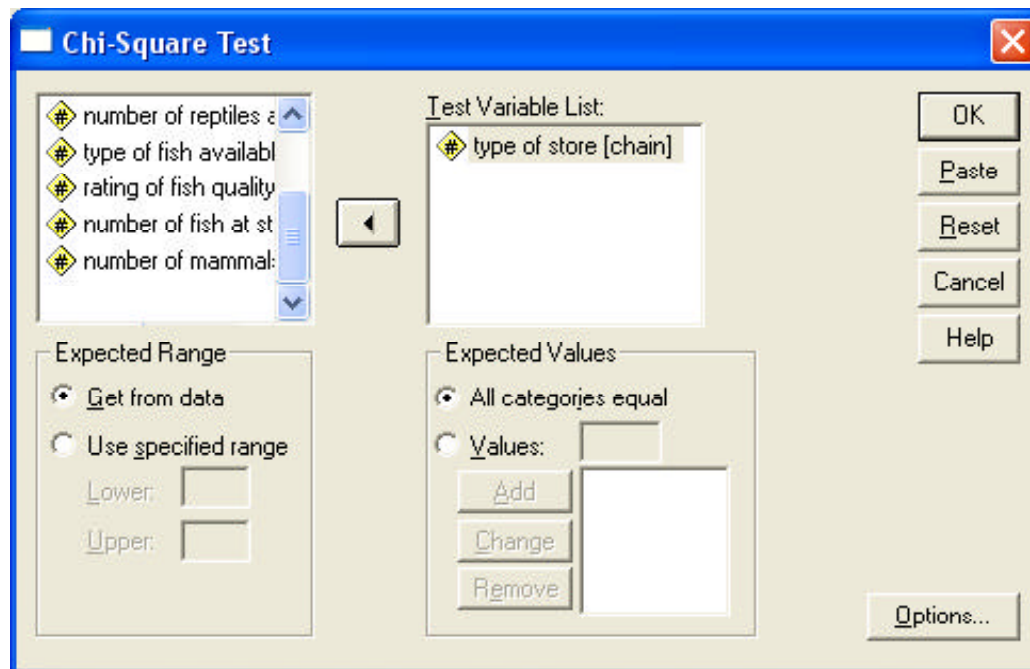
Research Hypothesis (it is often the case that the researcher's hypothesis is the statistical null, something more common for this statistical model than for most others):
The researcher hypothesized that the pet stores would be equally distributed across the three categories (chain stores, private stores and coop stores).

H₀: The sample represents a population of pet stores that is equidistributed across the store categories.

Analyze/Statistics → Nonparametric Tests → ChiSquare

- highlight the variable (be sure it is qualitative) and press the arrow to put it into the "Test Variable List" box
- under "Expected Values" be sure "all categories equal" is checked

If you only specify the variable and category values, SPSS assumes you want to test a Null Hypothesis of equiprobability (equidistribution).



Output:

type of store

	Observed N	Expected N	Residual
chain store	5	4.0	1.0
privately owned	3	4.0	-1.0
coop	4	4.0	.0
Total	12		

Test Statistics

	type of store
Chi-Square ^a	.500
df	2
Asymp. Sig.	.779

a. 3 cells (100.0%) have expected frequencies less than 5. The minimum expected cell frequency is 4.0.

This is the p-value.

Retaining Ho: means that the obtained frequencies are similar to the expected frequencies (equiprobability).

Chi-square significance tests produce more reliable results when all of the cells have expected frequencies of 5 or greater. Analyses are "allowed" when at least 80% of the cells have expected frequencies of 5 or more.

Reporting the Results:

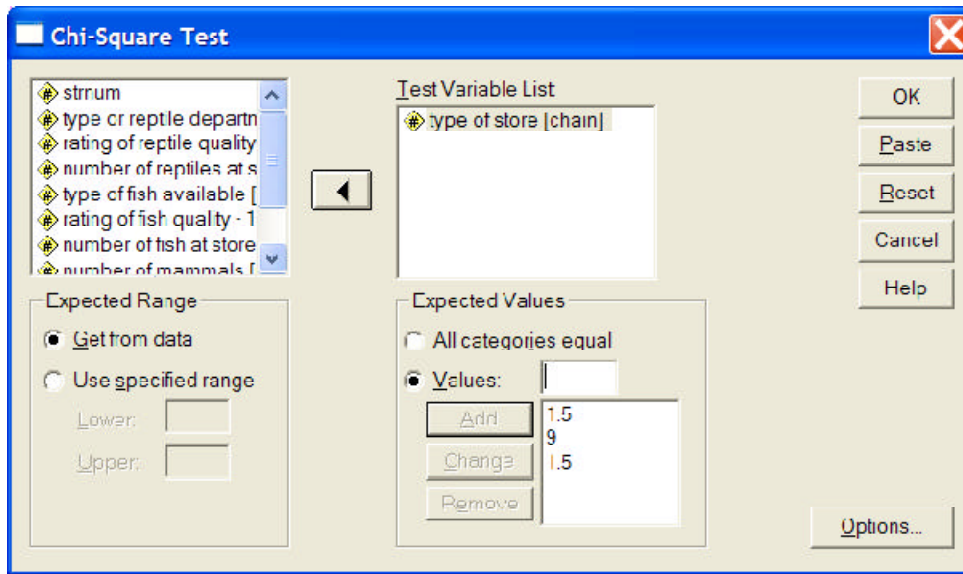
There were five Chain, three Private and four Coop stores. This distribution was not significantly different from an equiprobability model ($X^2(2) = .50$, $p = .779$).

Research Hypothesis: The researcher hypothesized that 3/4 of the stores would be Private, and Chain and Coop stores would each make up 1/8 of the stores.

H0: The sample represents a population that has the hypothesized distribution of frequencies across the categories of the qualitative variable.

Statistics → Nonparametric Tests → ChiSquare

- highlight the variable (be sure it is qualitative) and press the arrow to put it into the "Test Variable List" box
- under "Expected Values" click the "Values" button
 - enter each expected frequency into the window and press "add"
 - be sure to enter the expected values in the same order as the condition codes



type of store

	Observed N	Expected N	Residual
chain store	5	1.5	3.5
privately owned	3	9.0	-6.0
coop	4	1.5	2.5
Total	12		

Test Statistics

	type of store
Chi-Square ^a	16.333
df	2
Asymp. Sig.	.000

a. 2 cells (66.7%) have expected frequencies less than 5.
The minimum expected cell frequency is 1.5.

When the hypothesis is not of equiprobability, you can enter "expected" values that are percentages (be sure they add up to 100), proportions (be sure they add up to 1.0) or frequencies.

For this example:

The expected percentages are 12.5% for Chain, 75% for Private and 12.5% for Coop

The expected proportions are .125 for Chain, .75 for Private and .125 for Coop

The expected frequencies are

12.5% were hypothesized to be chain--
12.5% of 12 stores = 1.5 Chain stores expected

75% were hypothesized to be Private --
75% of 12 stores = 9 Private stores expected

12.5% were hypothesized to be Coop --
12.5% of 12 stores = 1.5 Coop stores expected

This is the p-value.

Rejecting H0: indicates that the distribution of cases across the categories differed from the expected frequencies.

Chi-square significance tests produce more reliable results when all of the cells have expected frequencies of 5 or greater. Analyses are "allowed" when at least 80% of the cells have expected frequencies of 5 or more.

Reporting Results:

There were five Chain, three Private and four Coop stores. This distribution was significantly different from the hypothesized distribution that 75% of the stores would be private, and 12.5% each would be coop and chain ($X^2(2)=16.33, p < .01$).