SPSS: Goodness-of-fit Chi-Square

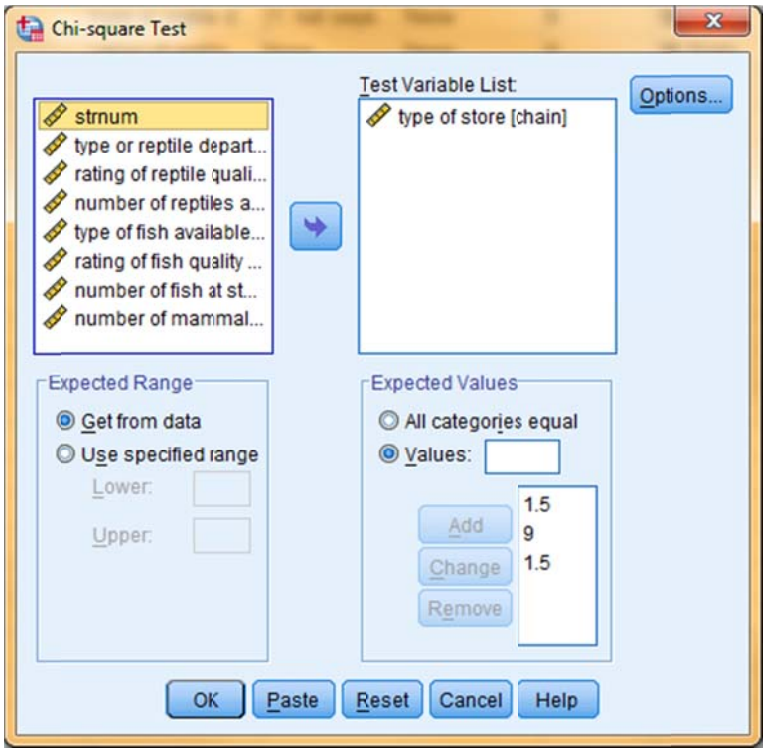
Application: To test a hypothesis about the distribution of subjects across the categories of a single qualitative/categorical variable.

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 3/4 of the population of stores represented by this sample would be Private, and Chain and Coop stores would each make up 1/8 of the stores.

H0: 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.

Statistics ➔ Nonparametric Tests ➔ Legacy Dialogs ➔ Chi-Square

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



You can enter "expected" values that are percentages (be sure the 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
- 12.5% for Coop

The expected proportions are:

- .125 for Chain
- .75 for Private
- .125 for Coop

The expected frequencies are:

- 12.5% of 12 stores = 1.5 Chain stores expected
- 75% of 12 stores = 9 Private stores expected
- 12.5% of 12 stores = 1.5 Coop stores expected

SPSS Syntax

```
NPAR TESTS
  /CHISQUARE=chain            ← qualitative variable
  /EXPECTED=1.5 9 1.5.        ←expected frequencies, proportions or percentages
```

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	16.333[a]
df	2
Asymp. Sig.	.000

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

The p-value estimates the probability that this result is a Type I error.

Remember, even if the printout shows it, never report p = .000, because that would suggest there is no possibility of a Type 1 error. Instead, report “p < .001”

Chi-square results are “suspicious” if more than 15% of the cells have expected frequencies less than 5.

Reporting the Results:

There were five Chain, three Private and four Coop stores. This distribution was significantly different from the hypothesized distribution that 75% of th stores would be private, and 12.5% each would be coop and chain ($X^2(2)=16.33$, $p < .001$).

It is important to show the category frequencies before presenting the Chi-square results.

As in the example, be sure to communicate:

- The research hypothesis (if there is one)
- The statistical results
- Whether or not those results support the research hypothesis