Chi-Square

χ^2

Chi-Square

Is a Test Statistic which is very versatile in the types of data it can handle: Discrete, Continuous, nonNormal, Categorical

Like z, *t*, and *F, the Chi-square is a* Test Statistic which has a Distribution or Distributions with known Probabilities for every value of the Test Statistic

Chi-Square

Chi-Square Test	df
for Goodness of Fit	$n - 1$
for Independence	$(r - 1)(c - 1)$
for Variance	$n - 1$

Chi-Square

Chi-Square Test for the Variance

- This test compares the Sample Variance to a value of Variance which we specify
- The test tells us whether there is a Statistically Significant difference between the Sample Variance and the specified Variance. It is analogous to the 1-Sample t-test for Means.

Chi-Square

$$\chi^2 = (n-1)\frac{\sigma^2}{s^2}$$

Chi-Square

- **Chi-Square Test for Goodness of Fit**
  - The Chi-Square (χ^2) Test for Goodness of Fit is a 1-way test of a Categorical Variable
  - The Test determines whether there is a Good Fit between Expected (E) Frequencies which we have specified or a Distribution and Observed (O) Counts from data

Chi-Square

- Null Hypothesis (Ho): There is no Statistically Significant difference between the Observed Counts and the Expected Frequencies. Therefore, there is a Good Fit.

Chi-Square

The Chi-Square Test for Goodness of Fit has certain minimum size requirements (test Assumptions)

- Every Expected Frequency must be 1 or greater
- and no more than 20% of the Expected Frequencies can be below 5
- ***If either of these Assumptions are not met, increasing the Sample Size will often help***

Chi-Square

Test Statistic: $\chi^2 = \sum \frac{(O-E)^2}{E}$

Critical Value: determined by α, and the Degrees of Freedom, df.

df = n – 1, where n is the number of categories.

Chi-Square

Results of the Test

- If $\chi^2 \geq \chi^{2\text{-critical}}$, or equivalently, $p \leq \alpha$,
  - There is a Statistically Significant difference
  - There is not a Good Fit.
  - Reject Ho, the Null Hypothesis.

Chi-Square

- If $\chi^2 < \chi^{2\text{-critical}}$, or equivalently, $p > \alpha$,
  - There is not a Statistically Significant difference.
  - There is a Good Fit.
  - Fail to Reject (Accept) Ho, the Null Hypothesis

Chi-Square

- **CHI-SQUARE TEST FOR INDEPENDENCE**
  - The Test Statistic, Chi-Square (χ^2), can be used to test whether two Categorical Variables are Independent or Associated

Chi-Square

- Independence is determined between Variables not between values (category names) of the Variables

- For example, “Is Ice Cream flavor preference dependent on Gender?”

Chi-Square

- If two Categorical Variables are Independent, then the Observed Frequencies (Counts) of the different values of the Variables should be Proportional

- **Important: the Expected Frequencies in each cell must be 5 or larger**

Chi-Square

Juice Study: Proportions are the same, so the Variables, Gender and Juice are Independent

	female		male	
	Count	Proportion	Count	Proportion
apple	28	0.35	14	0.35
grape	12	0.15	6	0.15
orange	40	0.50	20	0.50
Total	80	1.00	40	1.00

Ice Cream Study: Proportions are very different, so the Variables, Gender and Ice Cream are Associated (<u>not</u> Independent).

	female		male	
	Count	Proportion	Count	Proportion
chocolate	48	0.48	16	0.20
strawberry	28	0.28	40	0.50
vanilla	24	0.24	24	0.30
Total	100	1.00	80	1.00

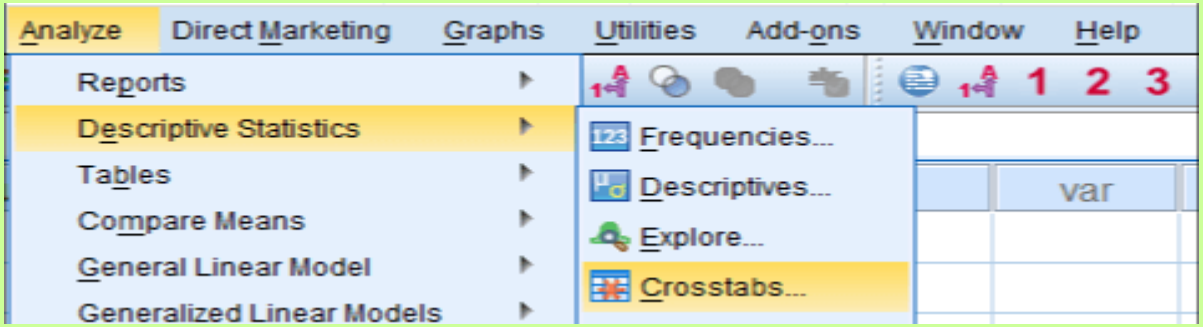
Chi-Square

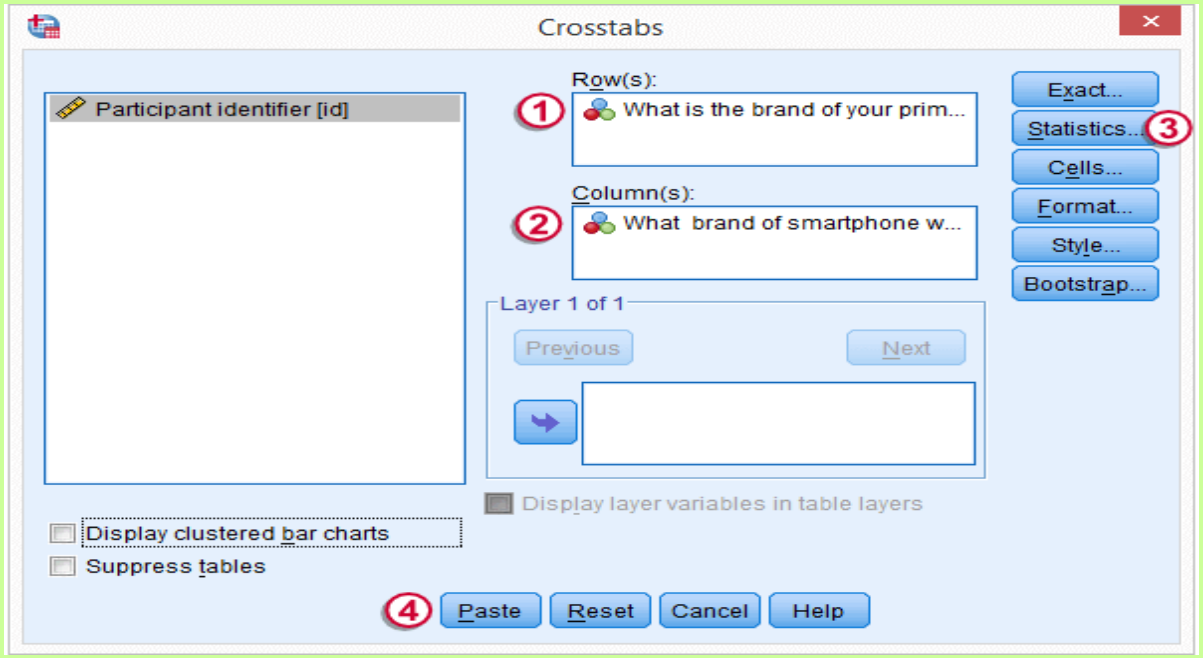
$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Chi-Square

- **Null Hypothesis**: there is no Association, that is, the Variables are Independent

Chi-Square


Analyze
Direct Marketing
Graphs
Utilities
Add-ons
Window
Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Frequencies...
Descriptives...
Explore...
Crosstabs...
var


Crosstabs
Participant identifier [id]
Row(s):
What is the brand of your prim...
Column(s):
What brand of smartphone w...
Layer 1 of 1
Previous
Next
Exact...
Statistics...
Cells...
Format...
Style...
Bootstrap...
Display layer variables in table layers
Display clustered bar charts
Suppress tables
Paste
Reset
Cancel
Help

Chi-Square

What is the brand of your primary smartphone? * What brand of smartphone would you like have? Crosstabulation

Count

		What brand of smartphone would you like have?				
		Samsung	HTC	Apple	Other	Total
What is the brand of your primary smartphone?	Samsung	27	11	4	8	50
	HTC	2	25	0	3	30
	Apple	3	2	17	8	30
	Other	8	2	4	26	40
Total		40	40	25	45	150

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	131.206[a]	9	.000
Likelihood Ratio	115.888	9	.000
Linear-by-Linear Association	35.269	1	.000
N of Valid Cases	150		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.00.

1

Chi-Square

- **Fisher's exact test** is used when you have two nominal variables. A data set like this is often called an "R×C table," where R is the number of rows and C is the number of columns. Fisher's exact test is more accurate than the chi-squared test or G-test of independence when the expected numbers are small

Chi-Square

- The null hypothesis is that the relative proportions of one variable are independent of the second variable

Chi-Square

- **McNemar's test** is used when the two samples are not independent, but instead are two sets of observations on the same individuals (equivalent to the paired t-Test but for nominal variables)