Some Chi-Square Business for Contingency Tables

Assumption 1: Observations are independent.

This is generally met when each person in the table is only in the table once – they are not counted twice or more.

Assumption 2: The test statistic is approximately distributed Chi-Square for relatively large samples.

This is generally met when expected frequencies in each cell of the contingency table are greater than or equal to 5 (there has to be the potential to observe 5 cases in each cell).

Effect Sizes

Phi, Φ, is a special case of the Pearson product-moment correlation coefficient for dichotomous items (0/1) – or can be thought of as a correlation in a 2 × 2 table.

Φ is a function of the Pearson chi-square statistic, χ^2:

$$\Phi = \sqrt{\frac{\chi^2}{n}}$$

This ranges from -1 to 1, like a correlation. If both the rows and columns of the contingency table exceed 2 levels, Φ can exceed 1.0. There is an adjustment made to Φ for contingency tables larger than 2×3 or 3×2 called Cramér's Phi (SPSS calls this Cramer's V).

$$\text{Cramér's } \Phi = \sqrt{\frac{\Phi^2}{(the\, smaller\, \#\, of\ rows\, or\, columns) - 1}}$$

For tables that are 2×2, 2×3, or 3×2, Phi and Cramér's Phi are equal.

Consider the following question:

Do males and females equally support building a new football stadium?

Female * Support building a football stadium Crosstabulation

			Support building a football stadium		Total
			No	Yes	
Female	Male	Count	22	58	80
		% within Gender	27.5%	72.5%	100.0%
	Female	Count	71	59	130
		% within Gender	54.6%	45.4%	100.0%
Total		Count	93	117	210
		% within Gender	44.3%	55.7%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	14.758(b)	1	.000		
N of Valid Cases	210				

a Computed only for a 2x2 table
b 0 cells (.0%) have expected count less than 5. The minimum expected count is 35.43.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	-.265	.000
	Cramer's V	.265	.000
N of Valid Cases		210	

a Not assuming the null hypothesis.
b Using the asymptotic standard error assuming the null hypothesis.

Based on our results, 73% of Males and 45% of Females (±5%) support building a stadium. There is a statistically significant difference in level of support between males and females, where $\chi^2(1, n=210)=14.8$, $p<.001$. This is a small, but statistically significant, relationship where Phi=.264.

STEP 1: Analyze → Descriptives → Frequencies

Check the frequency distribution to see if the values are "plausible"
That no strange values outside the possible range

STEP 2: Analyze → Descriptives → Crosstabs

Rows: put first question
Columns: put second question
Check your "Statistics" and "Cells" options and get "Percents" for either rows or columns – whichever you are more interested in

STEP 3: Interpret results