CONTINGENCY COEFFICIENT

Timo Törmäkangas, 2014

Procedure

The contingency coefficient can be calculated in the following way (see e.g. Blaikie).

1) Obtain a cross-table of the variables *X* and *Y*, where *X* has *r* categories and *Y* has *c* categories.
2) Calculate the value of the chi-square statistic.
3) The observed contingency coefficient is calculated as:

$$C_{obs} = \sqrt{\frac{\chi^2}{\chi^2+n}},$$

which varies between 0 and C_{max}. Note that C_{max} varies depending on the number of categories for *X* and *Y*.

4) If *X* and *Y* have the same number of categories (i.e. $r = c$), then the maximum value for the contingency coefficient is calculated as:

$$C_{max} = \sqrt{\frac{r-1}{r}}$$

where *r* is the number of rows (see step 1).

If *X* and *Y* have a differing number of categories (i.e. $r \neq c$), then the maximum value for the the contingency coefficient is calculated as:

$$C_{max} = \sqrt[4]{\frac{r-1}{r} \times \frac{c-1}{c}} = \left(\frac{r-1}{r} \times \frac{c-1}{c}\right)^{1/4}.$$

5) The standardized contingency coefficient is calculated as the ratio:

$$C_{stand} = \frac{C_{obs}}{C_{max}},$$

which varies between 0 and 1 with 0 indicating independence and 1 dependence.

Example

Here is a cross-table for gender (sex) and self-assessed economic situation (nc1049).

sex Sex * nc1049 ECON.SIT. Crosstabulation

Count

		nc1049 ECON.SIT.			
		1 GOOD	2 SATISFACTORY	3 BAD	Total
sex Sex	1 MALE	33	76	6	115
	2 FEMALE	47	153	25	225
Total		80	229	31	340

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	4,912[a]	2	,086
Likelihood Ratio	5,141	2	,076
Linear-by-Linear Association	4,659	1	,031
N of Valid Cases	340		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 10,49.

From the above table we find χ^2=4.912 and sample size n = 340. Although the chi-square statistic is not statistically significant (p = 0.086), we calculate the observed contingency coefficient as:

$$C_{obs} = \sqrt{\frac{\chi^2}{\chi^2+n}} = \sqrt{\frac{4.912}{4.912+340}} = 0.1193.$$

The table has two rows (r = 2) and three columns (c = 3), hence the maximum of contingency coefficient for this table is:

$$C_{max} = \left(\frac{r-1}{r} \times \frac{c-1}{c}\right)^{\frac{1}{4}} \left(\frac{2-1}{2} \times \frac{3-1}{3}\right)^{\frac{1}{4}} = 0.7598.$$

The standardized contingency coefficient is:

$$C_{stand} = \frac{C_{obs}}{C_{max}} = \frac{0.1193}{0.7598} = 0.157.$$

This indicates that the relationship is weak between the variables. We can report the result as showing no statistically significant dependence between gender and self-assessed economic situation (standardized C = 0.157, p = 0.086).

References

Blaikie, N. 2003. Analyzing Quantative Data. London: SAGE.