

Categories comparison tool

This tool can be used to:

- Perform statistical analysis with categorical data.
- Perform one of the following statistical tests: chi-squared, chi-squared with Yates's correction, Log likelihood and Fisher exact test.
- Visualize relationships in categorical data using the mosaic plot.

Instructions:

1) Copy-paste data in the text-box in the following format directly from a spreadsheet.

The data can be in either of these two formats: i) Cross-tabulated data and ii) Data set. The examples below show identical data presented in these two formats.

The cross-tabulated data

	A	B	C	D
1	Variety	must	have_to	need_to
2	American	352	355	201
3	British	448	405	190

Linguistic variable categories

Explanatory variable categories

Absolute frequencies (counts)

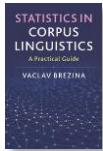
Data set

	A	B	C
1	ID	Variety	Outcome
2	1 BR		have_to
3	2 BR		must
4	3 AM		have_to
5	4 BR		have_to
6	5 AM		have_to
7	6 BR		need_to
8	7 BR		have_to
9	8 AM		need_to
10	9 BR		have_to
11	10 BR		must

Categorical explanatory variable

Categorical linguistic variable

Categories of variables (no spaces)



Brezina, V. (2018). *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge University Press.

2) Select options

2. Select options.

Input format of the data: Cross-tab Data set

Test: Chi-squared Chi-squared (Yates's correction) Log likelihood Fisher exact test

Visualize relationship

3) Click on 'Compare'.

1. Paste tab delimited data including header row and id column. For help click [here](#).

```
414 BR have_to
415 BR must
416 BR have_to
417 BR have_to
418 BR must
419 BR have_to
420 BR have_to
421 BR have_to
422 BR have_to
423 BR have_to
424 BR have_to
425 BR must
426 BR must
427 BR must
```

2. Select options.

Input format of the data: Cross-tab Data set

Test: Chi-squared Chi-squared (Yates's correction) Log likelihood Fisher exact test

Visualize relationship

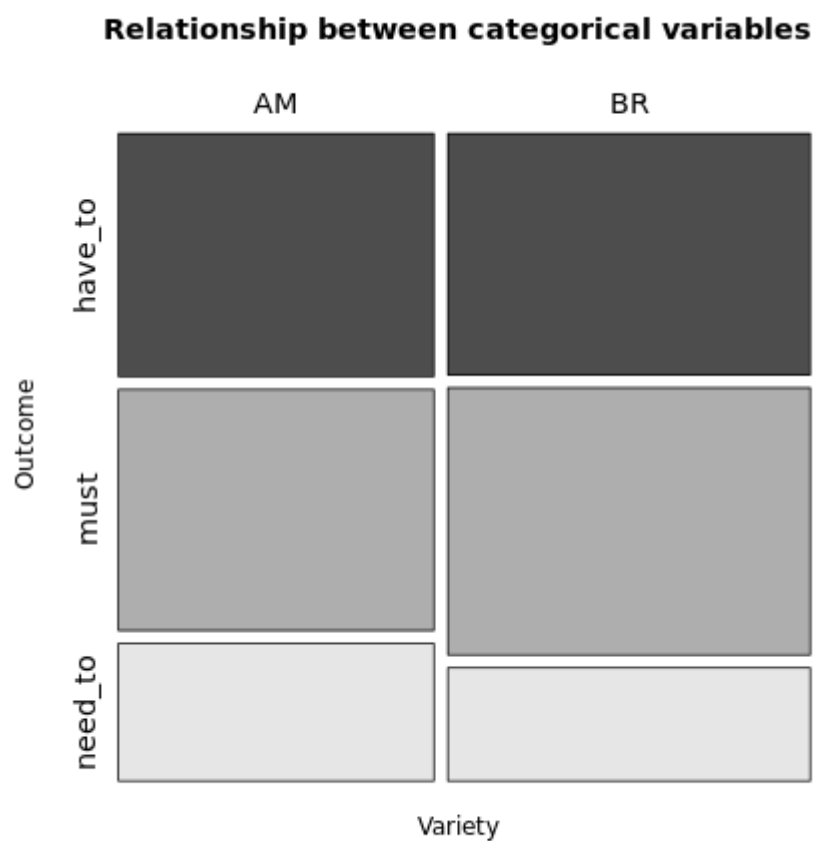
4) The output


The output shows the results of the chosen statistical test, a relevant effect size measure, 95% confidence interval for this effect size and the mosaic plot (optional).

test statistic degrees of freedom (df) p-value effect size

Pearson's Chi-squared test: 5.81 (2), p = 0.0548757, Cramer's V = 0.055 , 95% CI [0, 0.095], NEGLIGIBLE EFFECT

95% confidence interval interpretation of the effect size



 R code that performs the analysis can be viewed and copied when going with the mouse pointer to [R code](#)