

Users' guide

to software for calculating the α_{nominal} reliability coefficient for multi-valued nominal data

Reference: Klaus Krippendorff and Richard Craggs (2016). The reliability of multi-valued data. *Communication Methods and Measures* – in press

The software is implemented to run using Java 8. Users with Java versions released prior to 2015 may need to update theirs.

The running time of the calculations increases exponentially with number w of values from which the multi-valued data are composed. Depending on the speed of a user's computer $w > 10$ can become problematic.

The software accepts data files in comma-separated or tab-delimited formats, which could be created in Excel.

Reliability data contain the multi-valued sets of attributes that two or more observers, coders, raters, or measuring instruments assigned to two or more predefined units of analysis. The cells of an observers-by-units reliability data matrix may contain:

- Sets of any number of values, separated by a separator, for example “|”.
- Single values (without a separator)
- Empty sets, indicated by lone separators. Empty sets record such judgements as “no attribute,” “cannot code,” or “NA.”
- Cells left blank to indicate that an observer did not code a particular unit. They represent so-called missing data (not to be confused with empty sets).

Our paper relied on the following trivial example:

As Excel file:

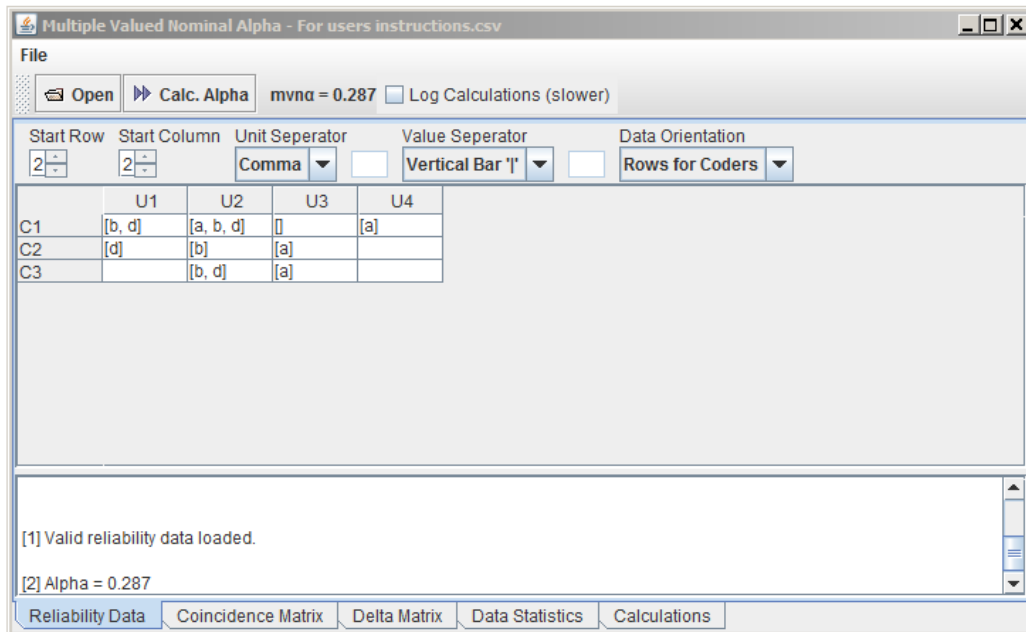
| | | | | |
|-------------|-----|-------|---|---|
| Units u : | 1 | 2 | 3 | 4 |
| Observer F | b d | a b d | | a |
| Observer G | d | b | a | |
| Observer H | | b d | a | |

| | A | B | C | D | E |
|---|------------|--------|--------|--------|--------|
| 1 | | Unit 1 | Unit 2 | Unit 3 | Unit 4 |
| 2 | Observer F | b d | a b d | | a |
| 3 | Observer G | d | b | a | |
| 4 | Observer H | | b d | a | |
| 5 | | | | | |

Steps to take (with reference to the interface on the following page):

1. Import the data file from Open
2. Observe the Reliability Data as entered
3. Indicate where the cell entries Start Row, here 2, and Start Column, here 2
4. Choose the Unit Separator – here Comma-delineated
5. Define the Value Separator used, here the Vertical Bar |. (Don't use Comma with *.csv data).
When a value separator is not defined, all cell contents are taken to be single categories.
6. Change the Data Orientation as needed, here: Rows for Coders (not Rows for Units)
7. Calculate Alpha

The interface of the software exhibits all of the above-mentioned terms:



8. Coincidence Matrix reproduces the matrix of observed coincidences

| | {} | {a} | {b} | {d} | {b, d} | {a, b, d} |
|-----------|-----|-----|-----|-----|--------|-----------|
| {} | 0 | 1.0 | 0 | 0 | 0 | 0 |
| {a} | 1.0 | 1.0 | 0 | 0 | 0 | 0 |
| {b} | 0 | 0 | 0 | 0 | 0.5 | 0.5 |
| {d} | 0 | 0 | 0 | 0 | 1.0 | 0 |
| {b, d} | 0 | 0 | 0.5 | 1.0 | 0 | 0.5 |
| {a, b, d} | 0 | 0 | 0.5 | 0 | 0.5 | 0 |

9. Delta Matrix reproduces the multi-valued-nominal differences

| | {} | {a} | {b} | {d} | {b, d} | {a, b, d} |
|-----------|-----|-----|-------|-------|--------|-----------|
| {} | 0.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| {a} | 1.0 | 0.0 | 1.0 | 1.0 | 1.0 | 0.5 |
| {b} | 1.0 | 1.0 | 0.0 | 1.0 | 0.333 | 0.5 |
| {d} | 1.0 | 1.0 | 1.0 | 0.0 | 0.333 | 0.5 |
| {b, d} | 1.0 | 1.0 | 0.333 | 0.333 | 0.0 | 0.199 |
| {a, b, d} | 1.0 | 0.5 | 0.5 | 0.5 | 0.199 | 0.0 |

10. Data Statistics offers accounts of the data entered (shown here only in part)

| |
|---|
| - Reliability data statistics - |
| Number of different individual values :3 |
| All individual values :{a, b, d} |
| Total number of individual values :11 |
| Number of uses of each value :{a=3, b=4, d=4} |
| Number of (comparable) occupied cells :8 |
| All multi values :{[], [a, b, d], [a], [b, d], [b], [d]} |
| All used sizes of multi values :{0, 1, 2, 3} |
| Number of multi values of each different size :{0=1, 1=4, 2=2, 3=1} |
| Proportions multi values for each different size :{0=0.125, 1=0.5, 2=0.25, 3=0.125} |

11. ☒ Log Calculations (slower) produces the Calculations that can be traced in detail