Computing Krippendorff's Alpha-Reliability

Klaus Krippendorff
kkrippendorff@asc.upenn.edu
2011.1.25, Literature updated 2013.9.13

Krippendorff's alpha (α) is a reliability coefficient developed to measure the agreement among observers, coders, judges, raters, annotators or measuring instruments drawing distinctions among typically unstructured phenomena or assign computable values to them. α emerged in content analysis but is widely applicable wherever two or more methods of generating data are applied to the same set of objects, predefined units of analysis or items and the question is how much the resulting data can be trusted to represent something worthy of analysis.

α's general form is:

$$\alpha = 1 - \frac{D_o}{D_e}$$

where D_o is the observed disagreement among values assigned to units of analysis:

$$D_o = \frac{1}{n}\sum_c\sum_k o_{ck\,\text{metric}}\,\delta^2_{ck}$$

and D_e is the disagreement one would expect when the coding of units is attributable to chance rather than to the properties of these units:

$$D_e = \frac{1}{n(n-1)}\sum_c n_c \sum_k n_{k\,\text{metric}}\,\delta^2_{ck}$$

The arguments in the two disagreement measures, o_{ck}, n_c, n_k and n, refer to the frequencies of values in coincidence matrices, to be defined below.

Algebraically, when observers agree perfectly, observed disagreement D_o=0 and α=1, which indicates perfect reliability. When observers agree as if chance had produced the results, D_o=D_e and α=0, which indicates the absence of reliability. α=0 occurs when observers are unable to distinguish among units or assign values to them drawn randomly from a collective estimate of the population of data. To rely on data generated by any method, α needs to be far from these two extreme conditions, ideally α=1. For reliability considerations, α's range is:

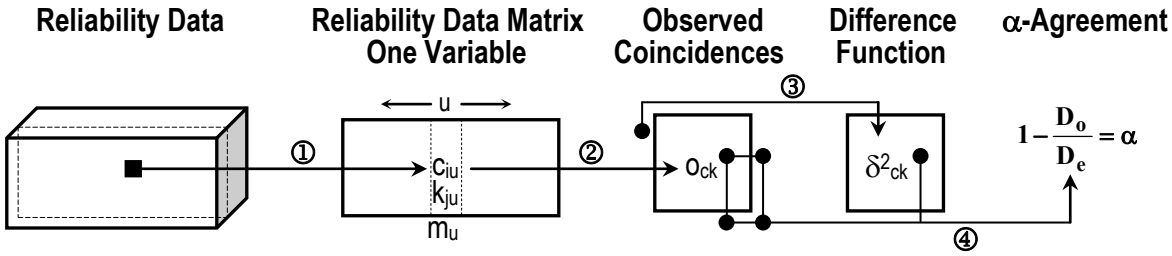
$$1 \;\geq\; \alpha \;\geq\; 0 \begin{cases} - \text{ Systematic disagreement} \\ \pm \text{ Sampling errors} \end{cases}$$

Unlike other specialized coefficients, α is a generalization of several known reliability indices. It enables researchers to judge a variety of data with the same reliability standard. α applies to:

- Any number of observers, not just two
- Any number of categories, scale values, or measures
- Any metric or level of measurement (nominal, ordinal, interval, ratio, and more)
- Incomplete or missing data
- Large and small sample sizes alike, not requiring a minimum

α evaluates reliability one variable at a time. It offers many analytical possibilities not presented here.

Reliability data duplicate the process of generating data whose reliability is in question. Given such data, α-reliability can be computed in four computational steps, graphed below.



These four computational steps will be defined and illustrated with four kinds of data of increasing complexity:

Finally,

A. Binary or dichotomous data, two observers, no missing data

① Construct a **reliability data matrix**; here, a **2 observers-by-*N* units** matrix, containing 2*N* values, **c** or **k**:

Units:		**1**	**2**	**…**	***u***	**…**	***N***
Observers:	***i*:**	c_{i1}	c_{i2}	…	c_{iu}	…	c_{iN}
	***j*:**	c_{j1}	c_{j2}	…	c_{ju}	…	c_{jN}

For example, when two observers judge *N*=10 units, the 2-by-10 data matrix contains 20 values:

Items judged:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Meg:	0	1	0	0	0	0	0	0	1	0
Owen:	1	1	1	0	0	1	0	0	0	0

② Tabulate **coincidences within units**. Coincidence matrices account for *all values* contained in a reliability data matrix. They differ from the familiar contingency matrices, which account for *units* in two dimensions, not values. The importance of this difference becomes apparent in C.

In a coincidence matrix, here a 2-by-2 matrix, units are entered twice, once as ***c*-*k*** pairs and once as ***k*-*c*** pairs. In the example, unit **1** is entered as a **0-1** pair of values *and* as a **1-0** pair of values. Unit **2** is entered as two **1-1** pairs of values, etc.:

Values:	**0**	**1**	
0	o_{00}	o_{01}	n_0
1	o_{10}	o_{11}	n_1
Number of Values:	n_0	n_1	$n = 2N$

	0	**1**	
0	10	4	14
1	4	2	6
	14	6	20

Accordingly, o_{00} represents the ten **0-0** pairs within units **4, 5, 7, 8**, and **10**. o_{01} represents the four **0-1** pairs in units **1, 3, 6**, and **9**, and o_{10} represents the four **1-0** pairs in the four same units. o_{11} represents the two **1-1** pairs found only in unit ***2***. n_0=14 is the number of **0**s in the reliability data matrix, n_1=6 is the number of **1**s, and n=2N=20 is the total number of values paired.

For these binary data, the frequencies of mismatching coincidences occur in two cells o_{01} and o_{10}.

③ skip

④ Compute **α-reliability** (its simplest form): $$_{\text{binary}}\alpha = 1 - \frac{D_o}{D_e} = 1 - (n-1)\frac{o_{01}}{n_0 \cdot n_1}$$

In the numerical example: $$_{\text{binary}}\alpha = 1 - (20-1)\frac{4}{14 \cdot 6} = 0.095$$

B. Nominal data, two observers, no missing data

① Construct a **reliability data matrix** – just as in A above. For a 2-by-12 example:

Items judged:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Ben:	*a*	*a*	*b*	*b*	*d*	*c*	*c*	*c*	*e*	*d*	*d*	*a*
Gerry:	*b*	*a*	*b*	*b*	*b*	*c*	*c*	*c*	*e*	*d*	*d*	*d*

② Tabulate **coincidences within units**. The general form of a coincidence matrix and with frequencies from the above example entered are:

Values:

	1	.	***k***	.	.	
1	o_{11}	.	o_{1k}	.	.	n_1
.	.		.			.
.	.		.			.
c	o_{c1}	.	o_{ck}	.	.	$n_c = \Sigma_k\, o_{ck}$
.	.		.			.
	n_1	.	n_k	.	.	$n = \Sigma_c \Sigma_k\, o_{ck}$

	a	***b***	***c***	***d***	***e***	
a	2	1	.	1	.	4
b	1	4	.	1	.	6
c	.	.	6	.	.	6
d	1	1	.	4	.	6
e	.	.	.	.	2	2
	4	6	6	6	2	24

Where $o_{ck} = \Sigma_u$ Number of ***c-k*** pairs in unit u specifically:

o_{ab} = 1 ***a-b*** pair in unit 1
o_{ba} = 1 ***b-a*** pair in unit 1
o_{aa} = 2 ***a-a*** pairs in unit 2
o_{bb} = 4 = 2 ***b-b*** pairs in unit 3
+ 2 ***b-b*** pairs in unit 4
and so forth.
n_a = 4 is the number of ***a***s
n_b = 6 is the number of ***b***s
and so forth.
n =24 is the total number of values.
For two observers: $n = 2N$

③ skip

④ Compute **α-reliability**:

$$_{\text{nominal}}\alpha = 1 - \frac{D_o}{D_e} = \frac{A_o - A_e}{1 - A_e} = \frac{\frac{\sum_c o_{cc}}{n} - \frac{\sum_c n_c(n_c-1)}{n(n-1)}}{1 - \frac{\sum_c n_c(n_c-1)}{n(n-1)}} = \frac{(n-1)\sum_c o_{cc} - \sum_c n_c(n_c-1)}{n(n-1) - \sum_c n_c(n_c-1)}$$

Wherein A_o is the percent of all observed matches within units u and A_e is the percent of matches obtainable by chance. The expression to the right simplifies the number of algebraic steps.

In the example:

$$_{\text{nominal}}\alpha = \frac{[(24-1)(2+4+6+4+2)]-[4(4-1)+6(6-1)+6(6-1)+6(6-1)+2(2-1)]}{[24(24-1)]-[4(4-1)+6(6-1)+6(6-1)+6(6-1)+2(2-1)]} = 0.692$$

C. Nominal data, any number of observers, missing data

① Construct a **reliability data matrix** – just as in A and in B above, but for m observers:

Units ***u:***	**1**	**2**	.	.	.	***u***	.	.	.	.	.	***N***
Observers: **1**	c_{1l}	c_{12}	.	.	.	c_{1u}	.	.	.	.	.	c_{1N}
i	c_{i1}	c_{i2}	.	.	.	c_{iu}	.	.	.	.	.	c_{iN}
j	c_{j1}	c_{j2}	.	.	.	c_{ju}	.	.	.	.	.	c_{jN}
.	.	.				.						.
m	c_{m1}	c_{m2}	.	.	.	c_{mu}	.	.	.	.	.	c_{mN}
Number of observers valuing ***u:***	m_1	m_2	.	.	.	m_u	.	.	.	.	.	m_N

When data are missing, m_u is variable and this matrix contains less than mN entries.

For example, a 4 observers-by-12 units reliability data matrix:

Units ***u*:**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	
Observer ***A*:**	1	2	3	3	2	1	4	1	2	**.**	**.**	**.**	
Observer ***B*:**	1	2	3	3	2	2	4	1	2	5	**.**	3	
Observer ***C*:**	**.**	3	3	3	2	3	4	2	2	5	1	**.**	
Observer ***D*:**	1	2	3	3	2	4	4	1	2	5	1	**.**	
Number m_u of values in unit u:	3	4	4	4	4	4	4	4	4	3	2	1	41

Note that 7 out of the 48 possible values in this matrix are missing. m_u varies from 1 to 4.

② Tabulate **coincidences within units**. The coincidence matrix appears as in B:

Values:	**1**	.	***k***	.	.	
1	o_{11}	.	o_{1k}	.	.	n_1
.	.		.			.
.	.		.			.
c	o_{c1}	.	o_{ck}	.	.	$n_c = \Sigma_k\, o_{ck}$
.	.		.			.
	n_1	.	n_k	.	.	$n = \Sigma_c \Sigma_k\, n_{ck}$

	1	**2**	**3**	**4**	**5**	
1	7	4/3	1/3	1/3	.	9
2	4/3	10	4/3	1/3	.	13
3	1/3	4/3	8	1/3	.	10
4	1/3	1/3	1/3	4	.	5
5	.	.	.	.	3	3
	9	13	10	5	3	40

Unlike in the two-observer case in B: $o_{ck} = \sum_u \dfrac{\text{Number of } c\text{-}k \text{ pairs in unit } u}{m_u - 1}$

Note that each unit contains $m_u(m_u–1)$ coincidences. A coincidence matrix accounts for all pairs of values found in u but adds each value only once. Unit **1** contains 3(3–1)=6 pairs of

matching **1**s. It contributes 6/(3-1)=3 to the o_{11} cell, one for each value. Unit ***2*** contains 4(4-1)=12 pairs, 6 matching **2-2** pairs, 3 mismatching **2-3** pairs, and 3 mismatching **3-2** pairs. It adds 6/(4-1)=2 to o_{22}, 3/(4-1)=1 to o_{23}, 1 to o_{32}, and 4 to the total n, thus fully accounting for its 4 values. Unit ***6*** contains 4·(4-1)=12 pairs of mismatching values, each adds 1/(4-1)=1/3 to a different cell. The lone value **3** in unit **12** affords no comparisons, 1·(1-1)=0 pairs and does not add to this account. Thus, the margins of coincidence matrices do not represent all values that occur in a reliability data matrix, only those that can be paired within units, here n=40 pairable values over all units.

③ Skip

④ Compute **α-reliability –** just as in B $_{\text{nominal}}\alpha = 1 - \frac{D_o}{D_e} = \frac{A_o - A_e}{1 - A_e} = \frac{(n-1)\sum_c o_{cc} - \sum_c n_c(n_c-1)}{n(n-1) - \sum_c n_c(n_c-1)}$

In the numerical example:

$$_{\text{nominal}}\alpha = \frac{(40-1)(7+10+8+4+3) - [9(9-1)+13(13-1)+10(10-1)+5(5-1)+3(3-1)]}{40(40-1) - [9(9-1)+13(13-1)+10(10-1)+5(5-1)+3(3-1)]} = 0.743$$

D. Any metric, any number of observers, missing data

① Construct a **reliability data matrix** – just as in C

② Tabulate **coincidences** within units – just as in C

③ Insert the **difference function** $_{metric}\delta^2_{ck}$ that is appropriate to the metric of the given data into the two disagreements D_o and D_e defined in the beginning of this document.

Note that α accounts for different metrics or levels of measurement by weighing the observed and expected coincidences by the squared difference between the coinciding values. These differences can be expressed as mathematical functions and in the form of a table. The latter makes their relative magnitudes transparent. Interval and ratio metric differences are functions of the values being paired. Ordinal differences depend on the frequencies of using values. And nominal differences are added here to generalize step ④.

- **Nominal metric differences –** Two values either match, or they do not:

$$_{\text{nominal}}\delta^2_{\text{ck}} = \begin{cases} 0 & \text{iff } c = k \\ 1 & \text{iff } c \neq k \end{cases}$$

Nominal categories, names:

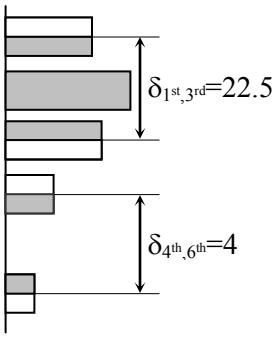
	a	***b***	***c***	***d***	***e***	***f***
a	0	1	1	1	1	1
b	1	0	1	1	1	1
c	1	1	0	1	1	1
d	1	1	1	0	1	1
e	1	1	1	1	0	1
f	1	1	1	1	1	0

- **Ordinal metric differences** – Values have the meaning of ranks and differences between ranks depend on how many ranks they are apart from each other. For example, with frequencies from data in C (and one unused rank added to show that it does not matter):

$$ {}_{\text{ordinal}}\delta^2_{ck} = \left(\sum_{g=c}^{g=k} n_g - \frac{n_c + n_k}{2} \right)^2 $$

Ranks:	**1st**	**2nd**	**3rd**	**4th**	**5th**	**6th**
1st	0	11^2	22.5^2	30^2	32.5^2	34^2
2nd	121	0	11.5^2	19^2	21.5^2	23^2
3rd	506	132	0	7.5^2	10^2	11.5^2
4th	900	361	56	0	2.5^2	4^2
5th	992	462	100	6.3	0	1.5^2
6th	1156	529	132	16	2.3	0

$n_{1st} = 9$
$n_{2nd} = 13$
$n_{3rd} = 10$
$n_{4th} = 5$
$n_{5th} = 0$
$n_{6th} = 3$



Ordinal metric differences may be standardized: $0 \le {}_{\text{ordinal}}\delta^2_{ck} \le 1$ by:

$$ {}_{\text{ordinal}}\delta^2_{ck} = \left(\frac{\sum_{g=c}^{g=k} n_g - \frac{n_c + n_k}{2}}{n - \frac{n_{c_{max}} + n_{c_{min}}}{2}} \right)^2 $$

where c_{max} is the largest and c_{min} the smallest rank among all ranks used. Standardization does not affect α, however.

- **Interval metric differences** – Values differ algebraically:

$$ {}_{\text{interval}}\delta^2_{ck} = (c-k)^2 $$

Interval values:	**-1**	**0**	**1**	**2**	**3**	**4**
-1	0	1^2	2^2	3^2	4^2	5^2
0	1	0	1^2	2^2	3^2	4^2
1	4	1	0	1^2	2^2	3^2
2	9	4	1	0	1^2	2^2
3	16	9	4	1	0	1^2
4	25	16	9	4	1	0

Interval metric differences may be standardized as well: $0 \le {}_{\text{interval}}\delta^2_{ck} \le 1$ by:

$$ {}_{\text{intervall}}\delta^2_{ck} = \left(\frac{c-k}{c_{max} - c_{min}} \right)^2 $$

where c_{max} is the largest and c_{min} is the smallest value occurring in the data. The use of standardized difference does not affect α either.

- **Ratio metric differences** – Algebraic differences between two values are expressed relative to an absolute zero point. They are proportional to the magnitude of their values:

$$_{\text{ratio}}\delta^2_{ck} = \left(\frac{c-k}{c+k}\right)^2$$

Ratio values:

	0	**1**	**2**	**3**	**4**	**5**
0	0	$(\frac{1}{1})^2$	$(\frac{2}{2})^2$	$(\frac{3}{3})^2$	$(\frac{4}{4})^2$	$(\frac{5}{5})^2$
1	1	0	$(\frac{1}{3})^2$	$(\frac{2}{4})^2$	$(\frac{3}{5})^2$	$(\frac{4}{6})^2$
2	1	.11	0	$(\frac{1}{5})^2$	$(\frac{2}{6})^2$	$(\frac{3}{7})^2$
3	1	.25	.04	0	$(\frac{1}{7})^2$	$(\frac{2}{8})^2$
4	1	.36	.11	.02	0	$(\frac{1}{9})^2$
5	1	.44	.18	.06	.01	0

- **Circular metric differences** – Shortest differences between any two values on a circular scale with arbitrary endpoints but a fixed circumference U = the number of equal intervals on a circle.

With the sine function expressed in degrees:

$$_{\text{circular}}\delta^2_{ck} = \left(sin\left\lceil 180\frac{c-k}{U}\right\rceil\right)^2$$

With the sine function expressed in radiance:

$$_{\text{circular}}\delta^2_{ck} = \left(sin\left[\pi\frac{c-k}{U}\right]\right)^2$$

Circular values:

	0	**1**	**2**	**3**	**4**	**5**
0	0	.25	.75	1	.75	.25
1	.25	0	.25	.75	1	.75
2	.75	.25	0	.25	.75	1
3	1	.75	.25	0	.25	.75
4	.75	1	.75	.25	0	.25
5	.25	.75	1	.75	.25	0

- **Bipolar metric differences** – Algebraic differences are expressed relative to the two endpoints, c_{min} and c_{max}, of the scale. Near the center, a bipolar metric behaves like an interval metric and near the poles it behaves like a ratio metric.

$$_{\text{polar}}\delta^2_{ck} = \frac{(c-k)^2}{(c+k-2c_{\min})(2c_{\max}-c-k)}$$

Bipolar values:

	-2	**-1**	**0**	**1**	**2**
-2	0	$\frac{1^2}{1\cdot7}$	$\frac{2^2}{2\cdot6}$	$\frac{3^2}{3\cdot5}$	$\frac{4^2}{4\cdot4}$
-1	.143	0	$\frac{1^2}{3\cdot5}$	$\frac{2^2}{4\cdot4}$	$\frac{3^2}{5\cdot3}$
0	.333	.067	0	$\frac{1^2}{5\cdot3}$	$\frac{2^2}{6\cdot2}$
1	.600	.250	.067	0	$\frac{1^2}{7\cdot1}$
2	1.00	.600	.333	.143	0

④ Compute **α-reliability** (the computationally most efficient form):

$$_{\text{metric}}\alpha = 1 - \frac{D_o}{D_e} = 1 - (n-1)\frac{\sum_c \sum_{k>c} o_{ck\ \text{metric}}\,\delta^2_{ck}}{\sum_c n_c \sum_{k>c} n_{k\ \text{metric}}\,\delta^2_{ck}}$$

Note that the sums in this form enumerate only one of the two symmetrical off-diagonal triangles of a coincidence matrix, containing all mismatching values. For nominal data the sums of these entries, o_{ck} and the products $n_c n_k$, suffice.

With data in C as **nominal data**: $$_{\text{nominal}}\alpha = 1 - (n-1)\frac{\sum_c \sum_{k>c} o_{ck}}{\sum_c n_c \sum_{k>c} n_k}$$

$$_{\text{nominal}}\alpha = 1 - (40-1)\frac{\frac{4}{3}+\frac{1}{3}+\frac{1}{3}+0+\frac{4}{3}+\frac{1}{3}+0+\frac{1}{3}+0+0}{9(13+10+5+3)+13(10+5+3)+10(5+3)+5\cdot 3} = 0.743$$

For other kinds of data, the frequencies of mismatching values must be weighted by the appropriate difference function $_{\text{metric}}\delta^2_{ck}$.

With data in C as **ordinal data**: $$_{\text{ordinal}}\alpha = 1 - (n-1)\frac{\sum_c \sum_{k>c} o_{ck\ \text{ordinal}}\,\delta^2_{ck}}{\sum_c n_c \sum_{k>c} n_{k\ \text{ordinal}}\,\delta^2_{ck}}$$

$$_{\text{ordinal}}\alpha = 1 - (40-1)\frac{\frac{4}{3}11^2+\frac{1}{3}22.5^2+\frac{1}{3}30^2+0+\frac{4}{3}11.5^2+\frac{1}{3}19^2+0+\frac{1}{3}7.5^2+0+0}{9(13\cdot 11^2+10\cdot 22.5^2+5\cdot 30^2+3\cdot 34^2)+13(10\cdot 11.5^2+5\cdot 19^2+3\cdot 23^2)+10(5\cdot 7.5^2+3\cdot 11.5^2)+5\cdot 3\cdot 4^2} = 0.815$$

With data in C as **interval data**: $$_{\text{interval}}\alpha = 1 - (n-1)\frac{\sum_c \sum_{k>c} o_{ck\ \text{interval}}\,\delta^2_{ck}}{\sum_c n_c \sum_{k>c} n_{k\ \text{interval}}\,\delta^2_{ck}}$$

$$_{\text{interval}}\alpha = 1 - (40-1)\frac{\frac{4}{3}1^2+\frac{1}{3}2^2+\frac{1}{3}3^2+0+\frac{4}{3}1^2+\frac{1}{3}2^2+0+\frac{1}{3}1^2+0+0}{9(13\cdot 1^2+10\cdot 2^2+5\cdot 3^2+3\cdot 4^2)+13(10\cdot 1^2+5\cdot 2^2+3\cdot 3^2)+10(5\cdot 1^2+3\cdot 2^2)+5\cdot 3\cdot 1^2} = 0.849$$

With data in C as **ratio data**: $$_{\text{ratio}}\alpha = 1 - (n-1)\frac{\sum_c \sum_{k>c} o_{ck\ \text{ratio}}\,\delta^2_{ck}}{\sum_c n_c \sum_{k>c} n_{k\ \text{ratio}}\,\delta^2_{ck}}$$

$$_{\text{ratio}}\alpha = 1 - (40-1)\frac{\frac{4}{3}\frac{1^2}{3^2}+\frac{1}{3}\frac{2^2}{4^2}+\frac{1}{3}\frac{3^2}{5^2}+0+\frac{4}{3}\frac{1^2}{5^2}+\frac{1}{3}\frac{2^2}{6^2}+0+\frac{1}{3}\frac{1^2}{7^2}+0+0}{9\left(13\frac{1^2}{3^2}+10\frac{2^2}{4^2}+5\frac{3^2}{5^2}+3\frac{4^2}{6^2}\right)+13\left(10\frac{1^2}{5^2}+5\frac{2^2}{6^2}+3\frac{3^2}{7^2}\right)+10\left(5\frac{1^2}{7^2}+3\frac{2^2}{8^2}\right)+5\cdot 3\frac{1^2}{9^2}} = 0.797$$

E. A general computational form, bypassing coincidence matrices:

① Start from a **reliability data matrix** as in C above:

Units **u:**	**1**	**2**	.	.	.	***u***	.	.	.	.	.	***N***
Observers: **1**	c_{11}	c_{12}	.	.	.	c_{1u}	.	.	.	.	.	c_{1N}
i	c_{i1}	c_{i2}	.	.	.	c_{iu}	.	.	.	.	.	c_{iN}
j	c_{j1}	c_{j2}	.	.	.	c_{ju}	.	.	.	.	.	c_{jN}
.	.	.				.						.
m	c_{m1}	c_{m2}	.	.	.	c_{mu}	.	.	.	.	.	c_{mN}

When data are missing, this matrix will contain fewer than mN values.

For the 4 observers-by-12 units example of reliability data used in C and D above:

Units ***u*:**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Observers: ***A*:**	1	2	3	3	2	1	4	1	2	**.**	**.**	**.**
***B*:**	1	2	3	3	2	2	4	1	2	5	**.**	3
***C*:**	**.**	3	3	3	2	3	4	2	2	5	1	**.**
***D*:**	1	2	3	3	2	4	4	1	2	5	1	**.**

Note that

- Out of the mN=4·12=48 possible values in this matrix, 7 are missing.
- The value 3, assigned by observer B to the 12th unit cannot be paired with other values in that unit, does contribute to observed agreements or disagreements, drops out when constructing a coincidence matrix, and has to be ignored.

Thus, this matrix contains a total of $n..$= 40 pairable values.

Instead of ② and ③,

enumerate the values found in units and create a **values-by-units matrix**:

Units:	**1**	**2**	.	.	.	***u***	.	.	.	.	.	.	
Values: **1**	n_{11}	n_{21}	.	.	.	n_{u1}	.	.	.	.	.	.	$n._1$
.	.	.				.							.
c	n_{1c}	n_{2c}	.	.	.	n_{uc}	.	.	.	.	.	.	$n._c$
k	n_{1k}	n_{2k}	.	.	.	n_{uk}	.	.	.	.	.	.	$n._k$
.	.	.				.							.
Totals:	$n_{1.}$	$n_{2.}$	.	.	.	$n_{u.}$	.	.	.	.	.	.	$n..$

Where n_{uc} = the number of values c assigned to unit u. $n_{uc} \le m$ observers. n_{uk} by analogy

$n_{u.} = \sum_c n_{uc}$ = the number of values assigned to unit u

$n._c = \sum_{u|n_{u.}\ge 2} n_{uc}$ = the number of pairable values c occurring in the reliability data (omitting all units with lone or no values: $n_{u.} \le 1$)

$n.. = \sum_{u|n_{u.}\ge 2} n_{u.}$ = the total number of all pairable values in the reliability data (omitting all units with lone or no values: $n_{u.} \le 1$); $n.. \le mN$

For the above example

Units u :	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	
Values c, k : ***1***	**3**	0	0	0	0	**1**	0	**3**	0	0	**2**	0	$9 = n_{\cdot 1}$
2	0	**3**	0	0	**4**	**1**	0	**1**	**4**	0	0	0	$13 = n_{\cdot 2}$
3	0	**1**	**4**	**4**	0	**1**	0	0	0	0	0	**1**	$10 = n_{\cdot 3}$
4	0	0	0	0	0	**1**	**4**	0	0	0	0	0	$5 = n_{\cdot 4}$
5	0	0	0	0	0	0	0	0	0	**3**	0	0	$3 = n_{\cdot 5}$
Totals $n_{u\cdot}$:	3	4	4	4	4	4	4	4	4	3	2	**1**	$40 = n_{\cdot\cdot}$

Note that the marginal sum $n_{\cdot 3}$ of pairable values 3 omits the only lone value in unit 12, $n_{12\cdot}$=1.

④ **Compute α** with the appropriate metric difference function as defined in D above:

$$_{\text{metric}}\alpha = 1 - \frac{D_o}{D_e} = 1 - (n_{\cdot\cdot} - 1)\frac{\sum_u \frac{1}{n_{u\cdot} - 1}\sum_c \sum_{k>c} n_{uc} n_{uk\ metric}\,\delta^2_{ck}}{\sum_c n_{\cdot c} \sum_{k>c} n_{\cdot k\ metric}\,\delta^2_{ck}}$$

If the above example consists of **nominal data** (perfect matches multiplied by $_{metric}\delta^2_{ck} = 0$):

$$_{\text{nominal}}\alpha = 1 - (40-1)\frac{0 + \frac{3\cdot 1}{4-1} + 0 + 0 + 0 + 6\frac{1\cdot 1}{4-1} + 0 + \frac{3\cdot 1}{4-1} + 0 + 0 + 0}{9(13+10+5+3) + 13(10+5+3) + 10(5+3) + 5\cdot 3} = 0.743$$

If the above example consists of **interval data**:

$$_{\text{interval}}\alpha = 1 - (40-1)\frac{0 + \frac{3\cdot 1\cdot 1^2}{4-1} + 0 + 0 + 0 + \frac{1\cdot 1\cdot 1^2 + 1\cdot 1\cdot 2^2 + 1\cdot 1\cdot 3^2 + \ldots + 1\cdot 1\cdot 1^2}{4-1} + 0 + \frac{3\cdot 1\cdot 1^2}{4-1} + 0 + 0 + 0}{9(13\cdot 1^2 + 10\cdot 2^2 + 5\cdot 3^2 + 3\cdot 4^2) + 13(10\cdot 1^2 + 5\cdot 2^2 + 3\cdot 3^2) + 10(5\cdot 1^2 + 3\cdot 2^2) + 5\cdot 3\cdot 1^2} = 0.849$$

Most computations of α can be performed with SPSS or SAS macros written by Andrew Hayes. Available at http://www.afhayes.com: Go to "SPSS and SAS Macros" then to "KALPHA." It also provides confidence limits and the probability of not exceeding the required α_{min}.

References:

Klaus Krippendorff (2013). *Content Analysis, an Introduction to Its Methodology, 3rd Edition*. Thousand Oaks, CA: Sage Publications – especially Chapter 12, pages 267-309.

Klaus Krippendorff (2004). Reliability in Content Analysis: Some common Misconceptions and Recommendations. *Human Communication Research 30,*3: 411-433. http://repository.upenn.edu/asc_papers/242 Last accessed 2013.9.13

Andrew F. Hayes, and Klaus Krippendorff (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures 1*: 77-89. http://www.unc.edu/courses/2007fall/jomc/801/001/HayesAndKrippendorff.pdf Last accessed 2013.9.13