

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*, 77-89.

Please cite this paper rather than this web page if you use the macro in research you are publishing.

The KALPHA macro computes Krippendorff's alpha reliability estimate for judgments made at any level of measurement, any number of observers, with or without missing data. The macro assumes the data file is set up in an SPSS data file with judges as the variables and the units being judged in the rows. The image below represents how the data in Krippendorff (2004), section 11.3.3 (page 230), would look as an SPSS data file. As can be seen, the entries in the data matrix should be the coding (quantified or numerically coded for nominal judgments) given to the unit in that row by the judge in that column.

	obsa	obsb	obsc	obsd
1	1.00	1.00	.	1.00
2	2.00	2.00	3.00	2.00
3	3.00	3.00	3.00	3.00
4	3.00	3.00	3.00	3.00
5	2.00	2.00	2.00	2.00
6	1.00	2.00	3.00	4.00
7	4.00	4.00	4.00	4.00
8	1.00	1.00	2.00	1.00
9	2.00	2.00	2.00	2.00
10	.	5.00	5.00	5.00
11	.	.	1.00	1.00
12	.	.	3.00	.

Once the macro command set is executed **exactly as downloaded**, the syntax for the macro is

`KALPHA judges = judgelist/level = lev/detail = det/boot = z.`

where *judgelist* is a list of variable names holding the names of the judges, *lev* is the level of measurement (1 = nominal, 2 = ordinal, 3 = interval, 4 = ratio), *det* is set to 1 if you desire SPSS to print the coincidence and delta matrices (as defined by Krippendorff, 2004), and *z* is the number of bootstrap samples desired for inference. The '/level', '/detail' and '/boot' boot subcommands are optional and default to 1, 0, and 0, respectively, if omitted. The number of bootstrap samples, if requested, must be at least 1000 and should be a multiple of 1000. Bootstrapping can take considerable time, so I recommend leaving off the /boot subcommand (or setting *z* to 0) until you are ready to generate confidence intervals and other inferential results.

The SPSS version of the macro recognizes both string and numeric data. Make sure when entering data that you specify the same variable type for all variables. (For example, if you mix string and numeric variable types, a numeric '1' will be perceived by SPSS as different from a string '1'.)

Missing data can be represented as a period ('.'). Units that are not coded by at least two judges are excluded from the analysis.

Here is example output applied to the data file above, requesting an estimate for nominal data, detailed output:

kalpha judges = obsa obsb obsc obsd/level = 1/detail = 1/boot = 5000.

Run MATRIX procedure:

Krippendorff's Alpha Reliability Estimate

	Alpha	LL95%CI	UL95%CI	Units	Observrs	Pairs
Nominal	.7434	.6035	.8600	11.0000	4.0000	55.0000

Probability (q) of failure to achieve an alpha of at least alphamin:

alphamin	q
.9000	.9932
.8000	.7936
.7000	.2940
.6700	.1176
.6000	.0154
.5000	.0000

Number of bootstrap samples:

5000

Judges used in these computations:

obsa obsb obsc obsd

=====
Observed Coincidence Matrix

7.00	1.33	.33	.33	.00
1.33	10.00	1.33	.33	.00
.33	1.33	8.00	.33	.00
.33	.33	.33	4.00	.00
.00	.00	.00	.00	3.00

Expected Coincidence Matrix

1.85	3.00	2.31	1.15	.69
3.00	4.00	3.33	1.67	1.00
2.31	3.33	2.31	1.28	.77
1.15	1.67	1.28	.51	.38
.69	1.00	.77	.38	.15

Delta Matrix

.00	1.00	1.00	1.00	1.00
1.00	.00	1.00	1.00	1.00
1.00	1.00	.00	1.00	1.00
1.00	1.00	1.00	.00	1.00
1.00	1.00	1.00	1.00	.00

Rows and columns correspond to following unit values

1.00	2.00	3.00	4.00	5.00
------	------	------	------	------

----- END MATRIX -----

(Note: Krippendorff reports on page 232 of his book that $\alpha = 0.734$. This is a typographical error in the book, as the macro is computing alpha correctly. On a document on his web page, Krippendorff uses the same data and correctly reports alpha as 0.743. The macro has been checked numerous times using many different data files in Krippendorff and Neuendorf (2002) and different levels of measurement, and it has always worked and computed alpha correctly. Nevertheless, if you find an error in the computations produced by the macro using your own data, let me know).

In order to be included in the computations, at least two observers must code the unit. In the example above, “units” is 11 because unit 12 is excluded (because only a single observer coded that unit). The number of pairs is equal to the number of pairs of judgments that contribute to the computation of alpha.

Problems can occur during bootstrapping that will invalidate the confidence intervals that are produced. When problems are detected during bootstrapping, the macro will let you know and then not produce any output that is based on bootstrap results. You should always check the output file (or, in SAS, the log file) for evidence of errors, and don't interpret the output if any errors are found, as the macro may still produce output when errors occur.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*, 77-89.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage.

Neuendorf, K. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.