statstutor

community project

encouraging academics to share statistics support resources

stcp-knox-kappaS

Measuring Agreement: Kappa

Cohen's kappa is a measure of the agreement between two raters who have recorded a categorical outcome for a number of individuals. Cohen's kappa factors out agreement due to chance and the two raters either agree or disagree on the category that each subject is assigned to (the level of agreement is not weighted). If you have more than two raters you need to consider an alternative approach which is detailed below.

To calculate a Cohen's Kappa the following assumptions need to be met:

1. The response being measured by the two raters is categorical (either a nominal or ordinal variable).
2. The responses are paired observations of the same measure and the two raters individually assess the measure for each individual.
3. The two raters remain fixed i.e. it is the same two raters assessing each individual.
4. The two raters are independent of each other.
5. For each assessment each rater makes an assessment based on the exact same number and definition of categories.

Example

Two pathologists read X-ray film from 60 individuals and grade tumours from 1 to 3 (Grade 3 is more severe). A portion of the data is shown opposite. Note that the two raters, the pathologists, are recorded in separate variables.

	Patologist_1	Patologist_2
1	Grade 1	Grade 1
2	Grade 1	Grade 1
3	Grade 1	Grade 2
4	Grade 3	Grade 3
5	Grade 3	Grade 3
6	Grade 2	Grade 1
7	Grade 1	Grade 1
8	Grade 2	Grade 2
9	Grade 1	Grade 1
10	Grade 1	Grade 1
11	Grade 1	Grade 1

Research question: Do the pathologists agree in their tumour assessment?

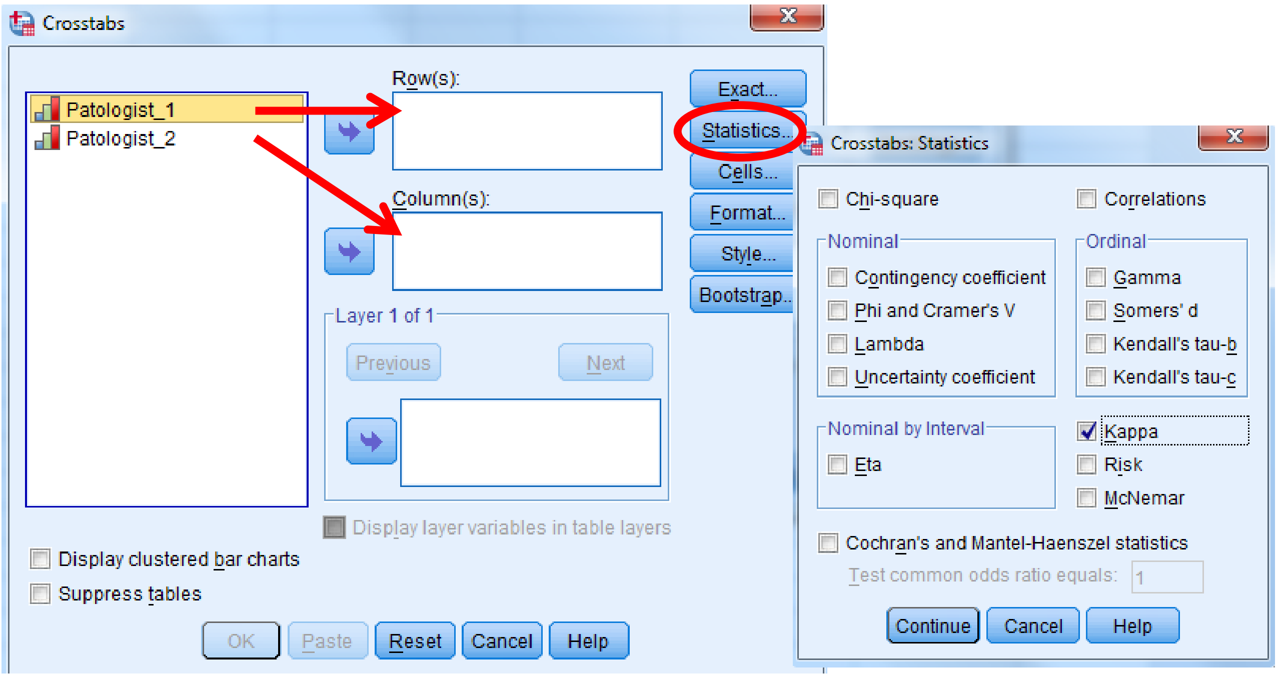
Steps in SPSS

Analyze → Descriptive Statistics → Crosstabs

Move the variable for each pathologist into the *Row(s):* and *Column(s):* box in either order. Select the *Statistics…* option and in the dialog box that opens select the *Kappa* checkbox. Select *Continue* to close this dialog box and then select *OK* to generate the output for the Cohen's Kappa.

© Chris Knox
University of Sheffield

Reviewer: Ellen Marshall
University of Sheffield



Output

The Symmetric Measures table in the output reports the Cohen's kappa (κ) which measures the agreement between the two raters accounting for chance.

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Measure of Agreement	Kappa	.595	.092	6.452	.000
N of Valid Cases		60			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

In this example the value for Cohen's kappa is **0.595.** This represents the proportion of agreement between the two pathologists beyond the agreement occurring due to chance. In this case, the p-value of p<0.001 suggests that the kappa value is statistically significant from zero. This does not imply statistically significant agreement and we can interpret the magnitude of agreement from the reported kappa value. To assess the level of agreement between two raters we can interpret the kappa value based on definitions outlined by Altman (1999).

Altman (1999) Guidelines for interpreting strength of agreement

Value of *K*	**Strength of agreement**
< 0.20	Poor
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Good
0.81 - 1.00	Very good

Reporting Kappa

Comparison of the assessment of tumours made by two pathologists produces a kappa value of **0.595** which suggests a moderate strength of agreement between the two pathologists. This value of kappa is significantly different from zero ($\kappa=0.595$, $p<0.001$).

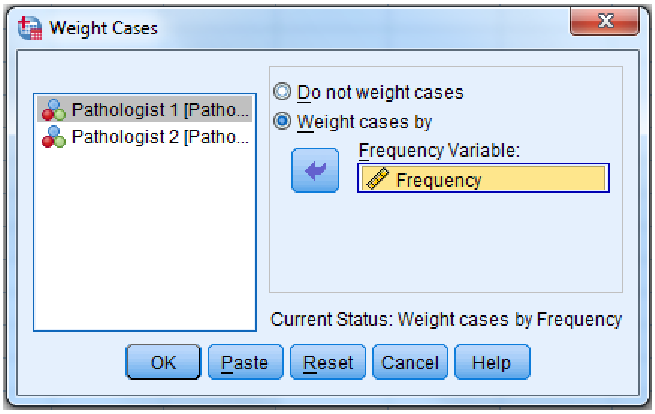
Analysing data already grouped into a table

It is also possible to analyse summary data (taken from a contingency table) in SPSS. The first step is to create a data set that reflects the cell counts in the table. In this example there are 9 possible combinations of grades that can be given by the two raters and so there should be 9 rows of data.

		Pathologist 2			Total
		Grade 1	Grade 2	Grade 3	
Pathologist 1	Grade 1	24	4	0	28
	Grade 2	10	10	1	21
	Grade 3	0	0	11	11
Total		34	14	12	60

	Pathologist...	Pathologist...	Frequency
1	Grade 1	Grade 1	24
2	Grade 1	Grade 2	4
3	Grade 1	Grade 3	0
4	Grade 2	Grade 1	10
5	Grade 2	Grade 2	10
6	Grade 2	Grade 3	1
7	Grade 3	Grade 1	0
8	Grade 3	Grade 2	0
9	Grade 3	Grade 3	11

Select *Data* → *Weight Cases…* then select *Weight cases by* and choose your *Frequency variable:* as Frequency. SPSS now treats the data according to the row and cell counts in the table e.g. as if there were 24 rows where both pathologists graded a tumour as 1.



Repeat the steps outlined above to calculate kappa which will give the same output as before.

Multiple Raters

For cases where there are more than two raters Cohen's kappa cannot be applied. In this instance Fleiss' kappa, an extension of Cohen's kappa for more than two raters, is required. Fleiss' kappa cannot be calculated in SPSS using the standard programme. The following website contains instructions and a downloadable program for calculating a Fleiss' kappa in Excel; http://www.real-statistics.com/reliability/fleiss-kappa/.

Weighted Kappa

The calculation of Cohen's kappa accounts for disagreement between two raters but does not take into account the level of the disagreement which is an important consideration when the ratings are in ordered categories. Consider the example below where individuals are rated on a scale of 1-5 by two raters.

	Rater 1	Rater 2	Difference between raters (*i*)
Individual 1	1	2	1
Individual 2	2	5	3
Individual 3	3	3	0

When calculating a Cohen's kappa the ratings for individual 1 & 2 are treated the same, they both represent a disagreement, even though there is clearly a bigger disagreement for individual 2 than individual 1. To account for this there is a need to calculate a weighted kappa. A predefined table of weights is used to calculate a kappa that accounts for the level of disagreement between raters so that different levels of disagreement are reflected in the contribution to kappa. Therefore in the example above Individual 2 would be assigned a higher disagreement weighting than Individual 1 and contribute less to the calculation of kappa which measures agreement.

Different weights w_i are assigned to individuals based on the difference in categories assigned by the raters denoted *i*. These are illustrated in the table above in the difference between raters column.

SPSS does not have an option to calculate a weighted kappa. To do so in SPSS you need to create a variable with the desired weights in and then select *Data* followed by *Weight Cases*…. before following the steps to calculate kappa.

Any chosen weights can be used; common weights are calculated as either linear or quadratic sets, where *k* denotes the total number of categories (5 in this example).

Linear weights set - the weights are calculated as follows: $w_i = 1 - \frac{i}{k-1}$

Quadratic weight set - the weights are calculated as follows: $w_i = 1 - \frac{i^2}{(k-1)^2}$

For this example based on 5 categories, the linear weights are 1 (total agreement), 0.75 (disagreement of 1), 0.50 (disagreement of 2), 0.25 (disagreement of 3) and 0 (total disagreement). In the quadratic set the weights are 1, 0.937, 0.750, 0.437 and 0 respectively.

If the difference between the first and second category is equally as important as a difference between the second and third category, etc., use linear weights. Use quadratic weights if the difference between the first and second category is less important than a difference between the second and third category, etc.

The following website contains instructions and a downloadable program for calculating a Fleiss' kappa in Excel; http://www.real-statistics.com/reliability/weighted-cohens-kappa/.

References

Altman, D. G. (1999). *Practical statistics for medical research*. New York, NY: Chapman & Hall/CRC Press.

Zaiontz, C. (2016). *Real Statistics Using Excel*. Retrieved 18 July, 2016, from http://www.real-statistics.com/