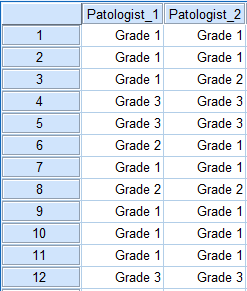
stcp-knox-kappaS

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**

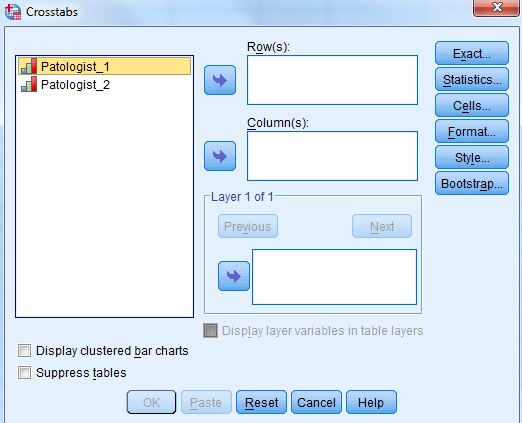
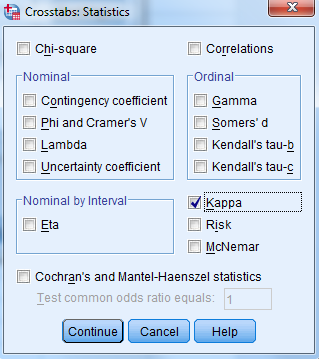
Two pathologists read X-ray film from 60 individuals and grade tumours from 1 to 3 (Grade 3 is more severe). A portion of the data is shown opposite. Note that the two raters, the pathologists, are recorded in separate variables.

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

**Steps in SPSS**

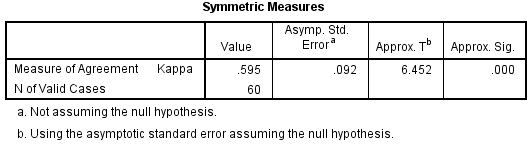
*Analyze → Descriptive Statistics → Crosstabs*

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.



**Output**

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



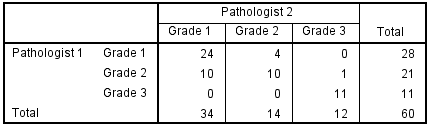
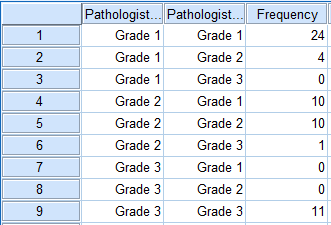
In this example the value for Cohen’s kappa is **0.595.** This represents the proportion of agreement between the two pathologists beyond the agreement occurring due to chance. In this case, the p-value of p<0.001 suggests that the kappa value is statistically significant from zero. This does not imply 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 (ᴋ=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.

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/>.

**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/