Chapter 20
Kappas for Reliability Assessment of Binary Data

The reproducibility of continuous data can be estimated with duplicate standard deviations (Chap. 19). With binary data Cohen's kappas are used for the purpose. Reliability assessment of diagnostic procedures is an important part of the validity assessment of scientific research.

Example

Positive (pos) or negative (neg) laboratory tests of 30 patients are assessed. All patiënts are tested a second time in order to estimate the level of reproducibility of the test.

		1st time		
		pos	neg	
2nd time	pos	10	5	15
	neg	4	11	15
		14	16	30

If the test is not reproducible at all, then we will find twice the same result in 50% of the patients, and a different result the second time in the other 50% of the patients.

Overall	30 tests have been carried out twice.
We observe	10 times 2 × positive and
	11 times 2 × negative.

And thus, twice the same is found in

21 patients which is considerable more than in half of the cases, which should have been 15 times.

Minimal indicates the number of duplicate observations if reproducibility were zero, maximal indicates the number of duplicate observations if the reproducibility were 100%.

$$\text{Kappa} = \frac{\text{observed} - \text{minimal}}{\text{maximal} - \text{minimal}}$$

$$= \frac{21 - 15}{30 - 15}$$

$$= 0.4$$

A kappa-value of 0.0 means that reproducibility is very poor.

A kappa of 1.0 would have meant excellent reproducibility.

In our example we observed a kappa of 0.4, which means reproducibility is very moderate.