Chapter 13
Measuring Agreement (30 Patients)

General Purpose

Interaction matrices have myriad applications. In the Chap. 12 it can be observed that they perform well for assessing relative health risks, making predictions from nominal and ordinal clinical data (Chap. 9–11), and statistical testing of outcome scores. In this chapter we will assess, if they also can be applied to measure agreement. Agreement, otherwise called reproducibility or reliability, of duplicate observations is the fundament of diagnostic procedures, and, therefore, also the fundament of much of scientific research.

Primary Scientific Question

Can a 2×2 interaction matrix also be used to demonstrate the level of agreement between duplicate observations of the effect of antihypertensive treatment.

Example

In 30 patients with hypertension the effect of an antihypertensive treatment was measured with normotension as outcome. Each patients was tested twice in order to assess the reproducibility of the procedure. the example was used before (Chap. 19, Reliability assessment of qualitative diagnostic tests, in: SPSS for starters part 1, pp 69–70, Springer Heidelberg Germany, 2010, from the same authors as the current work).

T.J. Cleophas, A.H. Zwinderman, *Machine Learning in Medicine - a Complete Overview*, DOI 10.1007/978-3-319-15195-3_13

Variables	
1	2
1,00	1,00
1,00	1,00
1,00	1,00
1,00	1,00
1,00	1,00
1,00	1,00
1,00	1,00
1,00	1,00
1,00	1,00
1,00	1,00
1,00	,00

Variable 1 = responder after first test
(0 = non responder, 1 = responder)
Variable 2 = responder after second test

The above table shows the results of first 11 patients. The entire data file is entitled agreement, and is in extras.springer.com. We will start by opening the file in SPSS.

Command

Analyze....Descriptive Statistics....Crosstabs....Row(s): enter Variable 1....Column(s): enter Variable 2....click Statistics....Mark: kappa....click Continue....click Cells..... mark Observed....click continue....click OK.

VAR00001 * VAR00002 crosstabulation				
Count				
		VAR00002		
		,00	1,00	Total
VAR00001	,00	11	4	15
	1,00	5	10	15
Total		16	14	30

In the output sheets a interaction matrix of the data is shown. If agreement is 100 %, then the cells b and c would be empty, and the cells a and d would contain 30 patients.

		variable 2	
		0	1
variable 1	0	a	b
	1	c	d

However, the cells a and d contain only 21 patients.

If agreement would be 0 %, then the cells a and d would contain 15 patients. However, 21 is more than 15, and so, this may indicate that agreement is better than 0 %, although less than 100 %. Cohen's Kappa is computed by SPSS to estimate the exact level of agreement.

Symmetric measures					
		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Measure of agreement	Kappa	,400	,167	2,196	,028
N of valid cases		30			

[a]Not assuming the null hypothesis
[b]Using the asymptotic standard error assuming the null hypothesis

The above table shows that the kappa-value equals 0.400. A kappa-value of 0 means poor reproducibility or agreement, a kappa-value of 1 means excellent. This result of 0.400 is moderate. This result is significantly different from an agreement of 0 at $p = 0.028$.

Conclusion

In this chapter it is assessed if interaction matrices can be applied to measure agreement. Agreement, otherwise called reproducibility or reliability, of duplicate observations is the fundament of diagnostic procedures, and, therefore, also the fundament of much of scientific research.

A 2×2 interaction matrix can be used to demonstrate the level of agreement between duplicate observations of the effect of antihypertensive treatment. We should add that kappa-values can also be computed from larger interaction matrices, like 3×3, 4×4 contingency tables, etc.

Note

More background, theoretical and mathematical information of correct and incorrect methods for assessing reproducibility or agreement are given in Chap. 45, Testing reproducibility, pp 499–508, in: Statistics applied to clinical studies 5th edition, Springer Heidelberg Germany, 2012, from the same authors.