

Chapter 33

Reliability Assessment of Quantitative Diagnostic Tests (17 Patients)

1 General Purpose

In statistics the term reliability is synonymous to reproducibility, like validity to accuracy, and precision to robustness (small-errors). For testing the reproducibility of quantitative diagnostic tests incorrect methods are often applied, like small mean differences between the first and second assessment, or a strong linear correlation between the first and second test but no direction coefficient of 45°. Correct methods include duplicate standard deviations, repeatability coefficients, and large intraclass correlations. In this chapter the intraclass correlation procedure is explained.

2 Schematic Overview of Type of Data File

Outcome	predictor
.	.
.	.
.	.
.	.
.	.
.	.
.	.

3 Primary Scientific Question

Are the first and second assessment of an experimental sample reproducible. Is intraclass correlation an adequate procedure to answer this question.

4 Data Example

In 17 patients quality of life scores were assessed twice. The primary scientific question: is the underneath quantitative diagnostic test adequately reproducible. The entire data file is entitled “chapter33reliabilityquantitative”, and is in extras.springer.com.

Quality of life score at first assessment	Quality of life at second assessment
10,00	10,00
9,00	10,00
7,00	6,00
5,00	6,00
3,00	7,00
8,00	8,00
7,00	7,00
8,00	7,00
7,00	8,00
8,00	8,00
7,00	9,00
10,00	11,00

5 Intraclass Correlation

For analysis the statistical model Reliability Analysis in the module Scale is required.

Command:

Analyze....Scale....Reliability Analysis....Items: enter quality of life first, quality of life second....Statistics.....mark: Intraclass Correlation Coefficient....Model: Two-way Mixed....Type: Consistency....Test value: 0....click Continue....click OK.

Reliability statistics	Crobach's Alpha	N of Items
	,832	2

Intraclass correlation coefficient

	Intraclass correlation ^a	95 % confidence interval		F test with true value 0			
		Low bound	Upper bound	Value	df1	df2	Sig
Single measures	,712 ^b	,263	,908	5,952	11	11	,003
Average measures	,832 ^c	,416	,952	5,952	11	11	,003

Two-way mixed effects model where people effects are random and measures effects are fixed

^aType C intraclass correlation coefficients using a consistency definition-the between-measure variance is excluded from the denominator variance

^bThe estimator is the same, whether the interaction effect is present or not

^cThis estimate is computed assuming the interaction effects is absent, because it is not estimable otherwise

The above tables show that the intraclass correlation (= SS between subjects/ (SS between subjects + SS within subjects), SS = sum of squares), otherwise called Cronbach's alpha, equals 0,832 (=83 %),if interaction is not taken into account, and 0,712 (=71 %), if interaction is accounted. An intraclass correlation of 0 means, that the reproducibility/agreement between the two assessments in the same subject is 0, 1 indicates 100 % reproducibility / agreement. An agreement of 40 % is moderate and of 80 % is excellent. In the above example there is, thus, a very good agreement with a p-value much smaller than 0,05, namely 0,003. The agreement is, thus, significantly better than an agreement of 0 %.

6 Conclusion

Intraclass correlations otherwise called Cronbach's alphas are used for estimating reproducibilities of novel quantitative diagnostic tests. An intraclass correlation of 0 means, that the reproducibility/agreement between the two assessments in the same subject is as poor as 0, 1 indicates 100 % reproducibility / agreement.

7 Note

More background, theoretical, and mathematical information about reliabilities of quantitative diagnostic tests is given in Statistics applied to clinical studies 5th edition, Chap. 45, Springer Heidelberg Germany, 2012, from the same authors.