

Macro for Pairwise Reliability Indicators

Carl M. Russell, Medical College of Georgia

Abstract

The purpose of this SAS macro is to estimate reliability indicators applicable to pairwise measurements on experimental units. The output consists of two pages with the following information: intraclass correlation, summary statistics, 1-way and 2-way ANOVA, paired t-test, test of equal variances, Bradley-Blackwood test, and an Altman-Bland plot. The macro is easily applied to data sets with experimental units as rows simply by specifying pairs of variables.

Introduction

The assessment of reliability or repeatability of measurements is important in many subject areas. Here we consider pairwise repeated measurements on experimental units. The repeated measurements may be by different methods, such as blood pressure measured by two different devices. The repeated measurements may be done by two evaluators randomly selected from a larger group of evaluators, such as two nurses each measuring blood pressure with the same device. Analysis of variance (ANOVA) models are often used for reliability assessment. In this framework, the difference that may exist because of different devices is a type of fixed effect. The difference in the nurses is a type of random effect. For fixed effects the investigator is usually interested in mean differences. Random effects usually indicate an interest in variance components.

The distinction between fixed and random effects is often ignored by investigators. In many biomedical sciences, reliability of pairwise measurements is routinely assessed using intraclass correlation. This approach has been criticized and additional approaches to reliability assessment suggested: paired t-test, confidence interval for mean difference, and plots of pairwise differences versus pairwise means (Altman and Bland, 1983; Bland and Altman, 1986; Bland and Altman, 1990). In addition to testing the mean difference, the variances of the individual methods or evaluators may be tested for equality (Pitman, 1937). A test for the simultaneous equality of means and variances has also been proposed (Bradley and Blackwood, 1990; Blackwood and Bradley, 1991).

The macro presented here was developed from a simulation that showed that intraclass correlation can be acceptable (>0.80) in cases in which mean or variance differences exist (Russell et al. 1994). The suggestion of that paper was that several reliability indicators should be examined to properly assess repeatability by methods or evaluators. The macro gives the investigator an opportunity to see many pertinent indicators of reliability in a concise 2-page format.

Macro

The macro is largely written in SAS/IML®. The syntax is:

%PAIRS(data set name, variable 1, variable 2);

The data set contains one line for each experimental unit. Variables 1 and 2 represent repeated measurements on each experimental unit. The output consists of two pages with the following information: intraclass correlation, summary statistics, 1-way and 2-way ANOVA, paired t-test, test of equal variances, Bradley-Blackwood test, and an Altman-Bland plot.

Example

The following data are taken from Blackwood and Bradley (1991).

<u>First</u> <u>Measurement</u>	<u>Second</u> <u>Measurement</u>
4.80	4.62
4.75	4.73
4.34	4.84
5.10	4.98
4.47	4.05
4.02	4.35
4.43	4.84
6.45	5.47
5.36	5.02
6.62	5.99
5.71	5.13
6.45	6.05
5.83	6.02
5.40	5.06
4.22	3.83

Each line represents an experimental unit. The macro was applied to variables FIRST and SECOND giving the output and plot shown on the next page. Artistic license has been used for the output to enhance printing for this article and to reduce it to one page. The macro invokes the print capability of IML and uses PROC PLOT for a lovely graph.

If the means of variables 1 and 2 are different, the Altman-Bland plot will show a shift up or down and the pairwise differences will be significantly different from zero. For FIRST and SECOND, the Altman-Bland plot is slightly pushed up but the mean difference of 0.1980 is not quite significantly different from zero ($p=0.0873$ -- it is a universal truth that 0.05 is the threshold for significance).

If the variances of variables 1 and 2 are different, the Altman-Bland plot will show a trend and the variance (slope) test of the Bradley-Blackwood test will be significant. For FIRST and SECOND, there is a slight increase in pairwise differences as pairwise means increase indicating the variance of FIRST is greater in magnitude than the variance of SECOND. However, the slope test of the Bradley-Blackwood test is not quite significant ($p=0.0743$).

The simultaneous equality of means and variances is assessed by the 2-df Bradley-Blackwood test. If this test rejects, but the variance (slope) test and paired t-test do not reject, there is a difference between variables 1 and 2 that is coming from both mean and variance differences (neither strong enough to show up alone). This is the case for these data. The Bradley-Blackwood test rejects ($p=0.0469$).

If you have historical information on acceptable intraclass correlation, compare it to the intraclass correlation derived from the 1-way ANOVA table. In most cases, investigators consider values in excess of 0.80 to be acceptable. Thus, this is a case of dubious repeatability with acceptable intraclass correlation.

Conclusion

This example was contrived by Blackwood and Bradley (1991) to demonstrate a situation in which the simultaneous test detects a difference but neither the mean nor variance test rejects. However, since this could happen in practice, it is prudent to examine reliability from several view points. This macro is a convenient tool for that purpose.

References

- Altman DG, Bland JM (1983) Measurement in medicine: The analysis of method comparison studies. *Statistician* 32:307-317.
- Blackwood LG, Bradley EL (1991) An omnibus test for comparing two measuring devices. *J. Qual. Technol.* 23:12-16.
- Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* Feb 8: 307-310.
- Bland JM, Altman DG (1990) A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput. Biol. Med.* 20:337-340.
- Bradley EL, Blackwood LG (1990) Comparing paired data: A simultaneous test for means and variances. *Am. Statist.* 43:234-235.
- Pitman EG (1937) A note on normal correlation. *Biometrika* 31:9-12.
- Russell CM, Williamson DF, Bartko JJ, Bradley EL (1994) Simulation study of a panel of reliability indicators applied to paired measurements. *Am. J. Human Biol.* 6:311-320.
- SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries.
® indicates USA registration.

Carl M. Russell
Office of Biostatistics
Medical College of Georgia
Augusta, Georgia 30912-4900
P: (706) 721-3785
F: (706) 721-6294
carlr@stat.mcg.edu

RELIABILITY ANALYSIS: FIRST AND SECOND MEASUREMENTS

N=15	MEAN	STD	T	PR(> T)
PAIRWISE DIFFERENCES (FIRST-SECOND)	0.1980	0.4171	1.8385	0.0873
PAIRWISE ABSOLUTE DIFFERENCES	0.3887	0.2327		
PAIRWISE MEANS	5.0977	0.7456		
FIRST	5.1967	0.8641		
SECOND	4.9987	0.6723		

ONE-WAY ANALYSIS OF VARIANCE (SUBJECTS)

	DF	SS	MS	F	PR(>F)
SUBJECTS	14	15.5643	1.1117	11.03	0.0000
ERROR	15	1.5119	0.1008		
TOTAL (CORR)	29	17.0761			

TWO-WAY ANALYSIS OF VARIANCE (SUBJECTS, MEASURERS)

	DF	SS	MS	F	PR(>F)
SUBJECTS	14	15.5643	1.1117	12.78	0.0000
MEASURERS	1	0.2940	0.2940	3.38	0.0873
ERROR	14	1.2178	0.0870		
TOTAL (CORR)	29	17.0761			

INTRAClass CORRELATION & SQUARE ROOTS OF VARIANCE COMPONENTS

	1-WAY ANOVA	2-WAY ANOVA
ICC	0.8338	0.8356
SIGMA (SUBJECTS)	0.7110	0.7158
SIGMA (MEASURERS)		0.1175
SIGMA	0.3175	0.2949

MODIFIED BRADLEY-BLACKWOOD TEST: PAIRWISE DIFFERENCES VS MEANS

	BETA	SE(BETA)	T	PR(> T)
INTERCEPT	-1.1537	0.7034	-1.6401	0.1249
SLOPE	0.2652	0.1366	1.9407	0.0743

	DF	SS	MS	F	PR(>F)
MODEL	2	1.1352	0.5676	3.90	0.0469
ERROR	13	1.8885	0.1453		
TOTAL (UNCORR)	15	3.0237			

