

Improper MCS Techniques

0.0.1 Paired sample T-test

? discusses the use of the well known paired sample t test to test for inter-method bias; $H : \mu_D = 0$. The test statistic is distributed a t random variable with $n - 1$ degrees of freedom and is calculated as follows;

$$t^* = \bar{D} / \frac{S_D}{\sqrt{n}} \quad (1)$$

where \bar{D} and S_D is the average of the differences of the n observations.

- Paired t tests test only whether the mean responses are the same. Certainly, we want the means to be the same, but this is only a small part of the story. The means can be equal while the (random) differences between measurements can be huge.
- The correlation coefficient measures linear agreement—whether the measurements go up-and-down together. Certainly, we want the measures to go up-and-down together, but the correlation coefficient itself is deficient in at least three ways as a measure of agreement. The correlation coefficient can be close to 1 (or equal to 1!) even when there is considerable bias between the two methods. For example, if one method gives measurements that are always 10 units higher than the other method, the correlation will be 1 exactly, but the measurements will always be 10 units apart.
- The magnitude of the correlation coefficient is affected by the range of subjects/units studied.
- The correlation coefficient can be made smaller by measuring samples that are similar to each other and larger by measuring samples that are very different from each other. The magnitude of the correlation says nothing about the magnitude of the differences between the paired measurements which, when you get right down to it, is all that really matters.
- The usual significance test involving a correlation coefficient—whether the population value is 0—is irrelevant to the comparability problem. What is important is not merely that the correlation coefficient be different from 0. Rather, it should be close to (ideally, equal to) 1!

intra-class correlation coefficient

- The intra-class correlation coefficient has a name guaranteed to cause the eyes to glaze over and shut the mouth of anyone who isn't an analyst. The ICC, which takes on values between 0 and 1, is based on analysis of variance techniques. It is close to 1 when the differences between paired measurements is very small compared to the differences between subjects. Of these three procedures—t test, correlation coefficient, intra-class correlation coefficient—the ICC is best because it can be large only if there is no bias and the paired measurements are in good agreement, but it suffers from the same faults ii and iii as ordinary correlation coefficients. The magnitude of the ICC can be manipulated by the choice of samples to split and says nothing about the magnitude of the paired differences.

Regression Methods

- Regression analysis is typically misused by regressing one measurement on the other and declare them equivalent if and only if the confidence interval for the regression coefficient includes 1. Some simple mathematics shows that if the measurements are comparable, the population value of the regression coefficient will be equal to the correlation coefficient between the two methods. The population correlation coefficient may be close to 1, but is never 1 in practice. Thus, the only things that can be indicated by the presence of 1 in the confidence interval for the regression coefficient is (1) that the measurements are comparable but there weren't enough observations to distinguish between 1 and the population regression coefficient, or (2) the population regression coefficient is 1 and therefore, the measurements aren't comparable.
- There is a line whose slope will be 1 if the measurements are comparable. It is known as a structural equation and is the method advanced by Kelly (1985). Altman and Bland (1987) criticize it for a reason that should come as no surprise: Knowing the data are consistent with a structural equation with a slope of 1 says something about the absence of bias but *nothing* about the variability about $Y = X$ (the difference between the measurements), which, as has already been stated, is all that really matters.