

## Sample Size Calculations for Agreement Coefficients

Posted: May 6, 2013

On June 28, 2012, I posted a note outlining an approach for calculating the required number of subjects necessary in an inter-rater reliability study to ensure a predetermined error margin for a chance-corrected agreement coefficient. The proposed approach requires the knowledge of some parameters that are generally unknown at the design stage. In this post, I am recommending a simpler and more practical approach for estimating the number of subjects as well as the number of raters in a multiple-rater study.

Many chance-corrected agreement coefficients are based on 2 components, which are the percent agreement and the percent chance agreement. While the percent chance agreement often differs from one coefficient to another, they generally share the same percent agreement. Because, researchers often compute and report 2 agreement coefficients or more, I recommend that the sample of subjects as well as the sample of raters for multiple-rater studies, be optimized on the percent agreement alone. The optimal numbers of subjects and raters will minimize a measure of the precision of percent agreement, and will apply to all coefficients that share the same percent agreement.

### ► Optimal Number of Subjects

The optimal number of subjects for a given inter-rater reliability represents the number of subjects that minimizes the standard error associated with the percent agreement between 2 arbitrary raters. The variance<sup>1</sup> of the percent agreement (denoted by  $p_a$ ) between 2 raters is given by:  $v = p_a(1 - p_a)/n$ , where  $n$  is the number of subjects. It can be shown that this variance is always smaller than  $1/(4 \times n)$ . Consequently, the op-

---

<sup>1</sup>More details about the variance associated with various agreement coefficients can be found in chapter 5 of Gwet (2012)

---

timal number of subjects is given by  $n = 1/E^2$ , where  $E = 2\sqrt{v}$  is the desired 95% error margin associated with the percent agreement.

Table 1 shows the minimum number of subjects required to achieve the desired 95% error margin. It follows that, if you want the estimated percent agreement to fall within 5% of its “true” error-free value, you will need to collect data on 400 subjects. This may appear to exceed by far what the budget allocated to many inter-rater reliability studies can accommodate. But the required number of subjects would decrease fast as the desired error margin goes up.

**Table 1:** Required Subject Sample Size by Desired Error Margin

Desired Error Margin	Required Subject Sample Size
5%	400
10%	100
15%	44
20%	25
25%	16
30%	11

### ► Optimal Number of Raters

A common issue that researchers often face when designing multiple-rater reliability studies, is about the number of raters that must participate to ensure an adequate precision in the results. Once again, I recommend to select a number of raters that will yield a predetermined desired precision of the percent agreement among raters. The multiple-rater version of the percent agreement and its variance are discussed in chapters 3, and 5 of Gwet (2012). For a given number of subjects, we may consider only the variation of the percent agreement due solely to the sampling of raters. One can prove that the percent agreement variance is smaller than  $4p_a^2/r^2$ , where  $p_a$  and  $r$  are

respectively the percent agreement, and the number of raters. Consequently, the ratio of the variance to the squared percent agreement is smaller than  $4/r^2$ . This proves that the coefficient of variation ( $cv$ ) of the percent agreement (i.e. the ratio of the standard error to the percent agreement) is smaller than  $2/r$ . The required number of raters is determined as follows:

$$r = 2/cv,$$

where  $cv$  is the anticipated coefficient of variation.

**Table 2:** Required Number of Raters by Desired Variation Coefficient

Desired Coefficient of Variation	Required Number Of Raters
5%	40
10%	20
15%	13
20%	10
25%	8
30%	7

## References

- 1 Gwet, K. (2012). *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Multiple Raters*, 3rd Edition. Advanced Analytics, LLC. Maryland, USA.
-