Educational Measurement
ISSUES AND PRACTICE
NCME national council on measurement in education

Educational Measurement: Issues and Practice
Fall 2015, Vol. 34, No. 3, p. 9

How Much Is Enough? A Reply to Sinharay, Haberman, and Boughton

Richard A. Feinberg and Howard Wainer, *National Board of Medical Examiners*

How much is enough? In addition to being the most important existential question of the modern age, this is also an important practical question. In the current manifestation it specializes to how much accuracy is enough. To begin the discussion it is important to remember that the standard error of a statistic is proportional to $\frac{1}{\sqrt{n}}$.

$$\text{standard error} \sim \frac{1}{\sqrt{n}}. \tag{1}$$

And so if we wish to report a statistic, say a correlation (in which the proportionality shifts to a rough equality) to two decimal places, say .36, we should have sufficient precision in the second digit so that it is more likely to be a 6 rather than a 5 or a 7. To be confident of this we would want the standard error to be .005 or smaller. Substituting the criterion of .005 into the equation version of expression (1) we get:

$$\text{standard error} = .005 = \frac{1}{\sqrt{n}},$$

or

$$\sqrt{n} = \frac{1}{.005} = 200,$$

or

$$n = (200)^2 = 40,000.$$

And so, if we have sample sizes that are in the range of 40,000 we can justify presenting results to two decimal places. But if our samples are of more modest sizes,say between 400 and 20,000, one decimal place is about all we can justify.

But, even if you have a sample size large enough to make a difference in the second decimal place statistically significant, is such a difference worthy of our attention? In medicine we often ask whether a difference is "clinically significant."

In 2014, Walmart reported corporate revenues as $476,294,366,412. It is said that accountants present such figures to the nearest dollar to prove they have a sense of humor. If it is rounded to $476 billion the resulting error is smaller than .1%.

Let us consider the results that Sinharay, Haberman, and Boughton (this issue) provided in this light. Table 1 shows their results rounded to one decimal place. We note with delight that there is 100% agreement on the decision (whether or not to report the subscore) between the two methods. We find it hard to imagine a situation in which a difference between VAR and PVAR in the second decimal place would be important to know about or would lead to any practical differences in subscore interpretation.

We leave it to the users to determine whether or not the ease of our simple approximation makes it a useful tool to keep in your pocket. We would like to thank Sinharay, Haberman, and Boughton for their providing yet another validation of the accuracy and applicability of our approximation.[1]

Table 1. Sinharay, Haberman, and Boughton Table Rounded to One Decimal Place

Test	Quantity	Subscore 1	Subscore 2	Subscore 3	Subscore 4	Subscore 5
Inview	VAR	1.0	1.0	1.1	1.1	1.1
	PVAR	1.0	1.0	1.0	1.0	1.0
	Agree?	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
TerraNova	VAR	1.0	1.1	1.0	1.0	1.0
	PVAR	1.0	1.0	1.0	1.0	1.0
	Agree?	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
SweSAT	VAR	1.2	1.3	.9	1.1	1.0
	PVAR	1.1	1.1	.9	1.0	1.0
	Agree?	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Test B1	VAR	1.0	1.0			
	PVAR	1.0	1.0			
	Agree?	**Yes**	**Yes**			
Test B2	VAR	1.2	1.0			
	PVAR	1.1	1.1			
	Agree?	**Yes**	**Yes**			

Richard A. Feinberg and Howard Wainer, National Board of Medical Examiners, 3750 Market Street, Philadelphia, PA 19104-3102; rfeinberg@nbme.org, hwainer@nbme.org.

Note

[1]This work was collaborative in all respects and the order of authorship is alphabetical.