Step out of the past: Stop using coefficient alpha; there are better ways to calculate reliability.

Jon Starkweather, PhD

Jon Starkweather, PhD
`jonathan.starkweather@unt.edu`
Consultant
Research and **S**tatistical **S**upport

`http://www.unt.edu`

`http://www.unt.edu/rss`

RSS hosts a number of "Short Courses".
A list of them is available at:
`http://www.unt.edu/rss/Instructional.htm`

The programming scripts similar to those in this article can also be found at:
`http://www.unt.edu/rss/class/Jon/R_SC`

Step out of the past: Stop using coefficient alpha; there are better ways to calculate reliability.

First, let me apologize up front. This will be a short article, in part because the topic is so clearly documented in the literature and in part because Research and Statistical Support[1] has just moved from its home in Sycamore Hall (formerly the Information Sciences Building) to its new home on the 3rd floor of Sage Hall (formerly the Business Administration Building). So, I apologize for the brevity of this article, but we have got a great deal of unpacking still left to do.

Coefficient Alpha

Coefficient alpha (Guttman, 1945; Cronbach, 1951), from here on referred to as just *alpha*, is one of those statistics which has built up a tremendous amount of momentum since its appearance and acceptance some 60 odd years ago. Momentum, in the previous sentence, refers to the popularity and subsequent intergenerational transfer of this statistic (from advisors to graduate students) which have resulted in the stubborn reliance upon it as a required standard measure of reliability or internal consistency. Unfortunately, over the decades of established use of alpha its assumptions and limitations have been overlooked or swept under the rug so to speak.

Alpha has three core assumptions. The first is the classical test theory assumption which indicates each item's observed score is the result of adding the item's true score and error. Second, alpha assumes Tau equivalency which indicates that all items carry equal loadings (i.e. the same true score contributes equally to the observed scores of all items) and all items have the same amount of variance. Third, alpha assumes uncorrelated error scores. "All three of these assumptions are likely to be violated to some degree in practice, and, therefore, the accuracy of coefficient alpha as an estimate of reliability is problematic" (Yang & Green, 2011, p. 379). The most commonly occurring flaw with alpha is that most social science survey instruments do not fulfill the second assumption. There are three likely reasons for violation of the second assumption. First, more than one latent factor contributes to the observed score of an item (or items) and second; items often do not have equivalent loadings to a single latent factor; and third, items do not have the same variance. In essence, alpha considers all items interchangeable with respect to how they measure a single latent factor. Violation of any of these assumptions leads to a biased estimate of reliability (Shevlin, Miles, Davies, & Walker, 2000). Furthermore, it has been documented (Hattie, 1985; Barchard, & Hakstian, 1997; Raykov, 1997) that alpha is inflated in the following situations: as the number of items increases, as the number of latent factors related to each item increases, as repetitive item content increases, and as item communalities increase. It is relatively common to refer to alpha as a lower bound estimate of reliability, however; given the strict assumptions and the likelihood of violating those assumptions, as well as the ease with which alpha can be inflated, it should greatly unnerve researchers that alpha has remained a standard in social science for so long.

Some researchers have pointed out the limitations of alpha and recognized more appropriate statistics for estimating reliability (Zinbarg, Revelle, Yovel, & Li, 2005; Sijtsma, 2009; Yang & Green, 2011; Cheng, Yuan, & Liu, 2012), including Cronbach toward the end of his life (Cronbach, & Shavelson, 2004). There are many alternatives to alpha available to the conscientious research. These choices can reasonably be classified as falling into two categories: traditional and contemporary. Traditional procedures of assessing reliability are likely known to anyone reading this, however; these procedures are not widely adopted. These procedures are often covered in a typical first year research methods and/or applied statistics class. Examples of these procedures include test-retest, equivalent forms, and split-half coefficients. Unfortu-

---

nately, these traditional procedures carry with them their own biases (e.g., memory effects, sample bias) and dilemmas (e.g., how do you decide to split a sample?). For these reasons, it is recommended that researchers adopt the more contemporary estimates of reliability, as discussed below.

Composite Reliability

Composite reliability is a rather general term which refers to a variety of robust reliability estimates; such as omega (McDonald, 1999) and intra-class correlation coefficient (ICC; Bartko, 1976; Shrout, & Fleiss, 1979). These estimates take account of the individual contribution of each latent factor to each item and each item's error; they are based on proportions of variance, and can be used in situations where hierarchical structure exists in the data. They provide a much less biased estimate of reliability than alpha. Fortunately with the advent of relatively cheap computing resources and open source software (**R**[2]), the calculation of these estimates is easy. The psych[3] package (Revelle, 2012) in **R**[4] provides easy to use functions (e.g., function 'alpha' and function 'omega') for calculating these estimates under a variety of psychometric conditions. The psych package also contains a wide variety of useful functions for applying factor analytic models. Revelle has provided lengthy vignettes (overview[5], input for SEM[6]) and a standard package manual[7] for explaining how to use the functions of the psych package and what they do.

Until next time; *put it all on at a hundred to one...*

---

[2]http://cran.r-project.org/
[3]http://cran.r-project.org/web/packages/psych/index.html
[4]http://cran.r-project.org/
[5]http://cran.r-project.org/web/packages/psych/vignettes/overview.pdf
[6]http://cran.r-project.org/web/packages/psych/vignettes/psych_for_sem.pdf
[7]http://cran.r-project.org/web/packages/psych/psych.pdf

References & Resources

Barchard, K. A., & Hakstian, A. R. (1997). The effects of sampling model on inference with coefficient alpha. *Educational and Pscyhological Measurement, 57*, 893 - 905. DOI: 10.1177/0013164497057006001

Bartko, J. J. (1976). On various intraclass correlation reliability coefficients. *Psychological Bulletin, 83*, 762 - 765.

Cheng, Y., Yuan, K., & Liu, C. (2012). Comparison of reliability measures under factor analysis and item response theory. *Educational and Psychological Measurement, 72*, 52 - 67. DOI: 10.1177/0013164411407315

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297 - 334.

Cronbach, L. J., & Shavelson, R. (2004). My [Cronbach] current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64*(3), 391 - 418.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika, 10*(4), 255 - 282.

Hatttie, J. (1985). Methodology review: Assessing unidimensionality of tests and item. *Applied Psychological Measurement, 9*(2), 139 - 164. DOI: 10.1177/014662168500900204

McDonald, R. P. (1999). *Test Theory: A Unified Treatment.* Mahwah, NJ: Erlbaum.

Raykov, T. (1997). Scal reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research, 32*(4), 329 - 353. DOI: 10.1207/s15327906mbr3204 2

Revelle, W. (2012). Package 'psych'. Package manual and explanatory vignettes available at: `http://cran.r-project.org/web/packages/psych/index.html`

Revelle, W., & Zinbarg, R. E. (submitted 2008). Coefficients alpha, beta, omega, and the glb: comments on Sijtsma. *Psychometrika*. Available at: `http://personality-project.org/revelle/publications/revelle.zinbarg.08.pdf`

Shevlin, M., Miles, J. N. V., Davies, M. N. O., & Walker, S. (2000). Coefficient alpha: A useful indicator of reliability? *Personality and Individual Differences, 28*, 229 - 237.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*, 107 - 120. DOI: 10.1007/s11336-008-9101-0

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420 - 428.

Yang, Y., & Green, S. B. (2011). Coefficient alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment, 29*(4), 377 - 392. DOI: 10.1177/0734282911406668

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's alpha, Revelle's beta, and McDonals's omega: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*, 123 - 133.

Zinbarg, R. E., Yovel, I., Revelle, W., & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for omega. *Applied Psychological Measurement, 30*, 121 - 144. DOI: 10.1177/0146621605278814

This article was last updated on June 19, 2012.

This document was created using LaTeX