



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

Meeting assumptions in the estimation of reliability

Brian P. Shaw
Indiana University
Bloomington, IN
bpshaw@indiana.edu

Abstract. Researchers and psychometricians have long used Cronbach’s α as a measure of reliability. However, there have been growing calls to replace Cronbach’s α with measures that have more defensible assumptions. One of the most common and straightforward recommended reliability estimates is ω . After a review of reliability and its estimation in Stata, I introduce the community-contributed command `omegacoeft`. This command reports McDonald’s ω in a format similar to the base `alpha` command. `omegacoeft` provides Stata users the ability to easily compute estimates of reliability with the confidence that the necessary statistical assumptions are met.

Keywords: st0660, `omegacoeft`, McDonald’s ω , Raykov’s ρ , alpha, reliability, measurement

1 Introduction: Reliability and alpha

Quantitative researchers and test developers frequently use multiple measures to increase the construct validity of an instrument. One issue facing both psychometricians and applied researchers is the extent to which the instrument’s measurement is consistent across different cases. This is known as the *reliability* of the measure: the proportion of variance in the scale score that is explained by a common source (nominally the construct of interest) rather than unique variation attributable only to individual items. In classical test theory parlance (1), reliability is defined as the ratio of the latent “true score” (T) to the observed score (Y), which is the sum of the “true” score (T) + error (E).

$$\rho_Y = \frac{\text{Var}(T)}{\text{Var}(Y)} = \frac{\text{Var}(T)}{\text{Var}(T) + \text{Var}(E)} \quad (1)$$

Reliability is also commonly described as the precision of an instrument, or the degree to which it consistently measures the construct it purports to measure. Low reliability indicates that a scale suffers from substantial measurement error. Even though the two terms are often conflated, reliability (consistent measurement) is not the same as *validity*, or the extent to which the measure faithfully represents the construct.

The most common measure of reliability that is currently in use is Cronbach's α . α is the expected score correlation between two random samples of test items (Cronbach 1951). Several algebraically equivalent formulas for α exist. In the first, (2), N is the number of items in the scale, \overline{C} is the average covariance between items, and \overline{v} is the average total variance.

$$\alpha = \frac{N}{N-1} \times \frac{\overline{C}}{\overline{v}} \quad (2)$$

One can rewrite (2) into (3), where V_i is the variance of each of the items in the scale and V_t is the total variance of the scale. This formula for α more closely resembles the conception of reliability in (1).

$$\alpha = \frac{N}{N-1} \left(1 - \frac{\sum_i V_i}{\sum V_t} \right) \quad (3)$$

α is widely reported and understood and is among the default capabilities of almost all software packages. However, it has come under increasing criticism from methodologists (for example, Green and Yang [2009]; Raykov [1997]; Sijtsma [2009]). The primary critique is that estimation of α requires a number of restrictive statistical assumptions that are almost never met in practice. Chief among these unmet assumptions is “essential tau-equivalence”, or the condition in which all the items in the scale contribute equally to its reliability. Calculation of α is also predicated upon all the variables being continuous and having normal distributions, among other assumptions. Furthermore, α is sensitive to scale length and increases as the number of items k increases even with the average covariance \overline{C} held constant. A commonly held view is that the reason for α 's predominance in published literature is its incorporation into the base functions of statistical software packages and not its statistical properties (Hayes and Coutts 2020; McNeish 2018).

2 Omega

A related measure of scale reliability that has relaxed assumptions relative to α is ω . Because ω is interpreted similarly to α but does not depend on essential tau-equivalence, methodologists frequently propose it as an alternative (for example, Graham [2006]; Hayes and Coutts [2020]; McNeish [2018]). The statistic is often referred to as “McDonald's ω ” (McDonald 1999), although the formula has appeared elsewhere in the methodological literature (for example, Bollen [1989]; Raykov [2001]). The formula for ω in (4) shows its closer resemblance to the general reliability formula in (1). Here b_i is each item's individual contribution to the true score, which has been described as the “construct loading” (Raykov 1997) of each of the k items.

$$\omega = \frac{\left(\sum_{i=1}^k b_i\right)^2 \text{Var}(T)}{\left(\sum_{i=1}^k b_i\right)^2 \text{Var}(T) + \left(\sum_{i=1}^k b_i\right)^2 \text{Var}(E)} \quad (4)$$

In practice, ω is most often based on loadings from confirmatory factor analysis, with all the scale items loading onto a single factor representing the construct that the scale is attempting to quantify. Accordingly, (5) represents ω using λ_i to represent each of the k items' factor loadings on the single latent factor (see Hayes and Coutts [2020]). The formula in (5) is equivalent to the one found in (4) but relies on the notation used for factor analytic models. As with many latent variable models, ω is typically estimated using maximum likelihood.

$$\omega = \frac{\left(\sum_{i=1}^k \lambda_i\right)^2}{\left(\sum_{i=1}^k \lambda_i\right)^2 + \sum_{i=1}^k \text{Var}(E_i)} \quad (5)$$

Cronbach's α can be regarded as a special case of the more general ω , and the two measures will be equivalent if the assumptions for α are met. The estimated reliability is often close to α but has been shown to differ under certain conditions (see McNeish [2018]). Because ω is more robust and because it is equivalent to α under the condition of essential tau-equivalence (but not the reverse), it is a desirable reliability coefficient. The only situation where α is preferable is when the scale consists of only two measures, where estimates of ω can be unstable even when the factor model is estimable.

3 Reliability estimation in Stata

Like all major statistical software packages, Stata easily estimates α with its `alpha` command. There is also a community-contributed command, `relicoeff` (Mehmetoglu 2015), that computes ω after estimation of latent factor models. However, to date Stata does not have a stand-alone command to calculate ω when the user is not already using confirmatory factor analysis.

4 The `omegacoeef` command

The `omegacoeef` command calculates the ω reliability coefficient for scales with three or more continuous variables. It is designed to function similarly to the `alpha` command.

4.1 Syntax

```
omegacoeef varlist [ , iterate(#) usemissing reverse(varlist)  
  noreverse(varlist) ]
```

4.2 Options

`iterate(#)` specifies the number of iterations the confirmatory factor analysis model will use to attempt to converge. In most applications, the model should converge with just a few iterations. If it does not converge after the default 1,000 iterations, it is most likely a sign that there is a specification issue, which is often the inclusion of a categorical variable. It is unlikely to be resolved with more iterations except in rare instances, but the option is available if the user desires it.

`usemissing` requests that the estimation incorporate information from variables in the scale with missing values if the missingness can be assumed to be missing at random (that is, the reason that the data are missing is adequately explained by the other variables in the estimation) or missing completely at random (that is, the reason that the data are missing is not related to the scale score or the other variables; see Schafer and Graham [2002]). In technical terms, the underlying factor model is fit using full-information maximum likelihood (FIML), which is a sophisticated method for predicting what the missing values would have been if they had been observed.

`reverse(varlist)` specifies variables to be treated as reverse coded (that is, their values are predicted to be negatively correlated with the majority of the other items in the scale). By default, the program will use the absolute value of the loadings of any variables that appear to be reverse coded (that is, those with negative factor loadings) to estimate the scale reliability. Thus, this option should rarely be necessary.

`noreverse(varlist)` could be used in rare cases to force the program not to use the absolute value of a factor loading. However, this option is provided only for advanced users who have a specific reason to desire the possibility of negative factor loadings in the calculation of ω .

4.3 Stored results

`omegacoe` stores the following in `r()`:

Scalars

<code>r(omega)</code>	estimated reliability coefficient
<code>r(k)</code>	number of items in the scale used to estimate reliability

4.4 Examples

The first example is a dataset containing physicians' attitudes toward the influence of cost on the treatments they prescribed. The goal is to estimate the reliability of the scale, which had six items.

```
. webuse bg2
(Physician-cost data)
. omeagacoe bg2cost1-bg2cost6
Calculating McDonald's omega/Raykov's rho...

NOTE: some of your items were treated as reverse coded.
Reversed items: bg2cost2 bg2cost3 bg2cost4
Number of items in the scale:      6
Scale reliability coefficient:      0.4717
```

The results from `omegacoe` are similar to the results obtained from the `alpha` command with the same items:

```
. alpha bg2cost1-bg2cost6
Test scale = mean(unstandardized items)
Reversed items:  bg2cost2 bg2cost3 bg2cost4
Average interitem covariance:      .134797
Number of items in the scale:      6
Scale reliability coefficient:      0.4831
```

The next example demonstrates an application of `omegacoe` with fictional data for four related tests that have missing values. In the first estimation, the `usemissing` option is not specified, and ω is estimated using only complete cases. In the second procedure, `usemissing` is specified, and Stata uses FIML to approximate complete cases and no observations are discarded.

```
. webuse cfa_missing
(CFA MAR data)

. * Without using information from the cases with missing values
. omeagacoe test1-test4
Calculating McDonald's omega/Raykov's rho...
Number of items in the scale:      4
Scale reliability coefficient:      0.7390

. * Using all information with FIML
. omeagacoe test1-test4, usemissing
Calculating McDonald's omega/Raykov's rho...
Number of items in the scale:      4
Scale reliability coefficient:      0.7960
```

These results are again analogous to those obtained using `alpha`. Using the `casewise` option, `alpha` discards all cases with missing values. Without this option, the command calculates the interitem covariances pairwise, so no cases are deleted, but no missing values are imputed.

```
. * Deleting cases with missing values
. alpha test1-test4, casewise
Test scale = mean(unstandardized items)
Average interitem covariance:      56.08045
Number of items in the scale:      4
Scale reliability coefficient:      0.7311

. * Retaining all cases
. alpha test1-test4
Test scale = mean(unstandardized items)
Average interitem covariance:      87.34249
Number of items in the scale:      4
Scale reliability coefficient:      0.7688
```

5 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 21-4
. net install st0660      (to install program files, if available)
. net get st0660          (to install ancillary files, if available)
```

6 References

- Bollen, K. A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.
- Cronbach, L. J. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* 16: 297–334. <https://doi.org/10.1007/BF02310555>.
- Graham, J. M. 2006. Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. *Educational and Psychological Measurement* 66: 930–944. <https://doi.org/10.1177/0013164406288165>.
- Green, S. B., and Y. Yang. 2009. Commentary on coefficient alpha: A cautionary tale. *Psychometrika* 74: 121–135. <https://doi.org/10.1007/s11336-008-9098-4>.
- Hayes, A. F., and J. J. Coutts. 2020. Use omega rather than Cronbach’s alpha for estimating reliability. But *Communication Methods and Measures* 14: 1–24. <https://doi.org/10.1080/19312458.2020.1718629>.
- McDonald, R. P. 1999. *Test Theory: A Unified Treatment*. Mahwah, NJ: Lawrence Erlbaum.

- McNeish, D. 2018. Thanks coefficient alpha, we'll take it from here. *Psychological Methods* 23: 412–433. <https://doi.org/10.1037/met0000144>.
- Mehmetoglu, M. 2015. relicoef: Stata module to compute Raykov's factor reliability coefficient. Statistical Software Components S458000, Department of Economics, Boston College. <https://ideas.repec.org/c/boc/bocode/s458000.html>.
- Raykov, T. 1997. Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research* 32: 329–353. https://doi.org/10.1207/s15327906mbr3204_2.
- . 2001. Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. *British Journal of Mathematical and Statistical Psychology* 54: 315–323. <https://doi.org/10.1348/000711001159582>.
- Schafer, J. L., and J. W. Graham. 2002. Missing data: Our view of the state of the art. *Psychological Methods* 7: 147–177. <https://doi.org/10.1037/1082-989X.7.2.147>.
- Sijtsma, K. 2009. On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika* 74: 107–120. <https://doi.org/10.1007/s11336-008-9101-0>.

About the author

Brian P. Shaw is a faculty member at Indiana University–Bloomington.