# Descriptions of Cronbach's alpha

Matthijs J Warrens

M&S colloquium
24 feb 2014

# Cronbach's alpha

Reliability of a test score

- Ratio of true score variance and observed score variance

Reliability must be estimated

- Often only one test administration
- Spit-half method, internal consistency method

Coefficient alpha
Guttman (1945), Cronbach (1951)

Most commonly used internal consistency coefficient

# Criticism

Alpha is most commonly used internal consistency coefficient

Criticism against use of alpha

- Not a measure of one-dimensionality
- Lower bound to reliability
  $\rightarrow$ Better lower bounds available

Cortina (1993), Sijtsma (2009):

Alpha is likely to be a standard tool in the future

# Definition

Common definition of alpha is

$$\alpha = \frac{n}{n-1} \cdot \frac{\sum_{i \neq i'} \sigma_{ii'}}{\sigma_X^2}$$

where

- $n \geq 2$ is the number of items
- $\sigma_{ii'}$ the covariance between items $i$ and $i'$
- $\sigma_X^2$ the variance of the test score

# Assumptions of alpha

Alpha estimates reliability of a test score

Two major assumptions

- Items are essentially tau-equivalent
- Uncorrelated errors

Essential tau-equivalency fails in practice

If assumptions do not hold, alpha underestimate reliability (lower bound)

Are there alternative descriptions of alpha?

(i.e. interpretations that are valid if assumptions do not hold)

# Outline talk

Alternative descriptions of alpha

Mean of all split-half reliabilities

- Cronbach (1951): Split into two groups of equal size
- Raju (1977): Mean of any split with groups of equal size
- What if groups have **unequal sizes**?

Relationship between alpha and Spearman-Brown formula

- Alpha in S-B formula $\rightarrow$ stepped down alpha
- Stepped down alpha is **weighted average of subtest alphas**

# Split-half reliability

Cronbach (1951): Mean of all split-half reliabilities

Split-half reliability

- Split test into two halves
- Correlation between half scores is estimate of reliability
- Correct estimate for half test length

Limitations of result by Cronbach (1951)

- Split-half reliability of Flanagan (1937) and Rulon (1937)
- Two halves must have equal size
  number of items must be even

# Split-half reliability

Split $n$ into two halves $n_1$ and $n_2$ with $n_1 + n_2 = n$

$$p_1 = \frac{n_1}{n} \qquad p_2 = \frac{n_2}{n}$$

Flanagan (1937) and Rulon (1937) proposed

$$\alpha_2 = \frac{4\sigma_{12}}{\sigma_X^2}$$

where

- $\sigma_{12}$ is the covariance between the sum scores of the two halves

Cronbach (1951): $\alpha = E(\alpha_2)$ if $p_1 = p_2$

# Alpha of $k$-split

$$\alpha_k = \frac{k}{k-1} \cdot \frac{\sum_{j \neq j'} \sigma_{jj'}}{\sigma_X^2}$$

where

- $k = 2, \ldots, n$ is the number of parts
- $\sigma_{jj'}$ the covariance between sum scores of parts $j$ and $j'$
- $\sigma_X^2$ the variance of the total score
- $\alpha_n = \alpha$

# Perfect split

Raju (1977)

- Perfect split: If split is such that $k$ parts have equal size, then alpha is mean of alphas of all possible $k$-splits
  $\alpha = E(\alpha_k)$
  $\alpha = E(\alpha_2)$ (Cronbach 1951)

Example 12 items

- into (6)(6) (2 parts of size 6)
- into (4)(4)(4)
- into (3)(3)(3)(3)
- into (1)(1)(1)(1)(1)(1)(1)(1)(1)(1)(1)(1)

# Research question

Raju (1977)

- If split is such that parts do not have equal sizes, then alpha exceeds mean of all possible splits
  $\alpha > E(\alpha_k)$

How close are $\alpha$ and $E(\alpha_k)$ in this case?

# Result 1: Formula for $E(\alpha_k)$

Using tools from Raju (1977)

$$E(\alpha_k) = \frac{n}{n-1} \cdot \frac{k}{k-1} \cdot \frac{\sum_{j \neq j'} p_j p_{j'} \sum_{i \neq i'} \sigma_{ii'}}{\sigma_X^2} = \alpha \cdot \frac{k}{k-1} \sum_{j \neq j'} p_j p_{j'}$$

Non-negative difference

$$\alpha - E(\alpha_k) = \alpha \left( 1 - \frac{k}{k-1} \sum_{j \neq j'} p_j p_{j'} \right) \leq 1 - \frac{k}{k-1} \sum_{j \neq j'} p_j p_{j'}$$

# Non-negative difference

$\alpha$ and $E(\alpha_k)$ 'equal' if $\alpha - E(\alpha_k) < 0.01$

Using previous inequality

$$\alpha - E(\alpha_k) \leq 1 - \frac{k}{k-1} \sum_{j \neq j'} p_j p_{j'} < 0.01$$

or

$$\frac{k}{k-1} \sum_{j \neq j'} p_j p_{j'} > 0.99$$

# 2-split

$2m+1$ items
$k=2$ parts
'Best' split is

$$p_1 = \frac{m}{2m+1} \qquad p_2 = \frac{m+1}{2m+1}$$

Check inequality

$$4p_1p_2 > 0.99$$

With $\geq 11$ items alpha 'equal' to mean of all split-half reliabilities

| $m$ | $n$ | $4p_1p_2$ |
|---|---|---|
| 1 | 3 | 0.889 |
| 2 | 5 | 0.960 |
| 3 | 7 | 0.980 |
| 4 | 9 | 0.988 |
| 5 | **11** | 0.992 |
| 6 | 13 | 0.994 |
| 7 | 15 | 0.996 |
| 8 | 17 | 0.997 |

# 3-split

$3m + 1$ items
$k = 3$ parts
'Best' split is

$$p_1 = p_2 = \frac{m}{3m+1} \qquad p_3 = \frac{m+1}{3m+1}$$

Check inequality

$$3(p_1 p_2 + p_1 p_3 + p_2 p_3) > 0.99$$

| $m$ | $n$ | |
|---|---|---|
| 1 | 4 | 0.938 |
| 2 | 7 | 0.980 |
| 3 | **10** | 0.990 |
| 4 | 13 | 0.994 |
| 5 | 16 | 0.996 |
| 6 | 19 | 0.997 |
| 7 | 22 | 0.998 |
| 8 | 25 | 0.998 |

# Another 3-split

$3m + 2$ items
$k = 3$ parts
'Best' split is

$$p_1 = \frac{m}{3m+2} \qquad p_2 = p_3 = \frac{m+1}{3m+2}$$

Check inequality

$$3(p_1 p_2 + p_1 p_3 + p_2 p_3) > 0.99$$

With $\geq 10$ items $\alpha \approx E(\alpha_3)$

| $m$ | $n$ | |
|---|---|---|
| 1 | 5 | 0.960 |
| 2 | 8 | 0.984 |
| 3 | **11** | 0.992 |
| 4 | 14 | 0.995 |
| 5 | 17 | 0.997 |
| 6 | 20 | 0.998 |
| 7 | 23 | 0.998 |
| 8 | 26 | 0.999 |

# Worst 2-split

With 'best' splits we have for sufficiently large $n$

$$\frac{k}{k-1}\sum_{j\neq j'} p_j p_{j'} > 0.99 \qquad \text{and thus} \qquad \alpha - E(\alpha_k) < 0.01$$

Does this hold for any split?

No. Suppose $n$ items and 2-split

$$p_1 = \frac{1}{n} \qquad p_2 = \frac{n-1}{n}$$

Worst possible 2-split. We have

$$4p_1p_2 = \frac{4(n-1)}{n^2} \to 0 \quad \text{as} \quad n \to \infty$$

# Shortened tests

Shortened test

- Test that measures same construct with fewer items
- Available time and resources usually limited
  $\rightarrow$ short tests more efficient
- Literature review in Kruyen et al. (2013)
- Old psychometric wisdom:
  many items are needed for reliable and valid measurement

Examples

- Beck Depression Inventory: 21 $\rightarrow$ 13 items
- Marlowe-Crowne Social Desirability Scale: 33 $\rightarrow$ 10 items

# Definition alpha

Alpha can also be defined as

$$\alpha_n = \frac{n\overline{\text{cov}}_n}{\overline{\text{var}}_n + (n-1)\overline{\text{cov}}_n}$$

where

- $n \geq 2$ is the number of items
- $\overline{\text{cov}}_n$ is the average covariance
- $\overline{\text{var}}_n$ is the average variance
- subscript $n$: $\alpha_n$ defined on $n$ items

# Subtests

Shortened test is a subtest of the full test

A $k$-item subtest with where $2 \leq k < n$ (! new use of $k$)
is obtained by removing $n - k$ items from the original $n$-item test

Alpha of a $k$-item subtest is defined as

$$\alpha_k = \frac{k\overline{\text{cov}}_k}{\overline{\text{var}}_k + (k-1)\overline{\text{cov}}_k}$$

where

- $k$ is the number of items
- $\overline{\text{cov}}_k$ is the average covariance between the $k$ items
- $\overline{\text{var}}_k$ is the average variance of the $k$ items

# How many subtest alphas?

How many $\alpha_k$'s?
How many $k$-item subtests?

Subtest is obtained by removing $n-k$ items from the original $n$-item test

Binomial coefficient

$$\binom{n}{n-k} = \frac{n!}{k!(n-k)!} = \binom{n}{k} \qquad \text{e.g.} \quad \binom{5}{2} = \frac{5!}{2!3!} = \frac{5 \cdot 4}{2} = 10$$

We have $\binom{n}{k}$ $k$-item subtests and as many versions of $\alpha_k$

Let $\binom{n}{k} = m$ (! new use of $m$)

# Spearman-Brown formula

To predict reliability of a similar test of different length we may use the Spearman-Brown formula

$$\rho^* = \frac{N\rho}{1 + (N-1)\rho}$$

where

- $\rho$ is the old reliability
- $\rho^*$ is the new reliability
- $N$ is the extension factor, e.g. $N = 2$ double length
- assumption: items are parallel
  (stronger requirement than essential tau-equivalency)

# Stepped down alpha

Suppose we want to predict the reliability of a $k$-item test

Using

$$\rho = \alpha_n = \frac{n\overline{\text{cov}}_n}{\overline{\text{var}}_n + (n-1)\overline{\text{cov}}_n}$$

and extension factor $N = k/n$ (contraction $N < 1$) in

$$\rho^* = \frac{N\rho}{1 + (N-1)\rho}$$

we obtain

$$\alpha_n^* = \frac{k\overline{\text{cov}}_n}{\overline{\text{var}}_n + (k-1)\overline{\text{cov}}_n}$$

Coefficient $\alpha_n^*$ is called stepped down alpha

# Result 2: stepped down alpha = weighted average

How is the stepped down alpha

$$\alpha_n^* = \frac{k\overline{\text{cov}}_n}{\overline{\text{var}}_n + (k-1)\overline{\text{cov}}_n}$$

related to the subtest alphas $\alpha_k(1)$, $\alpha_k(2)$, $\alpha_k(3)$, ... where

$$\alpha_k = \frac{k\overline{\text{cov}}_k}{\overline{\text{var}}_k + (k-1)\overline{\text{cov}}_k}$$

Stepped down alpha = weighted average of subtest alphas

$$\alpha_n^* = \frac{w_1\alpha_k(1) + w_2\alpha_k(2) + \cdots + w_m\alpha_k(m)}{w_1 + w_2 + \cdots + w_m}$$

# Result 2: stepped down alpha = weighted average

Stepped down alpha = weighted average of subtest alphas

$$\alpha_n^* = \frac{w_1\alpha_k(1) + w_2\alpha_k(2) + \cdots + w_m\alpha_k(m)}{w_1 + w_2 + \cdots + w_m}$$

where number of subtest alphas is

$$m = \binom{n}{k}$$

and the weights are the denominators

$$\overline{\text{var}}_k + (k-1)\overline{\text{cov}}_k$$

of the subtest alphas

# Proof

Weighted average is a fraction

- Numerator is a sum of all versions of $k\overline{\text{cov}}_k$
- Denominator is a sum of all versions of $\overline{\text{var}}_k + (k-1)\overline{\text{cov}}_k$

If we consider all subtests of length $k$
number of times a pair of items is part of a $k$-item subtest is

$$\binom{n-2}{k-2} = \frac{(n-2)!}{(k-2)!(n-k)!},$$

while number of times a single item is part of a $k$-item subtest is

$$\binom{n-1}{k-1} = \frac{(n-1)!}{(k-1)!(n-k)!}$$

# Proof

Numerator (sum of $k\overline{\text{cov}}_k$) of weighted average is

$$\binom{n-2}{k-2} \cdot \frac{2}{k(k-1)} \cdot \frac{n(n-1)}{2} \cdot k\overline{\text{cov}}_n = \binom{n}{k} k\overline{\text{cov}}_n$$

while denominator (sum of $\overline{\text{var}}_k + (k-1)\overline{\text{cov}}_k$) is

$$\binom{n}{k} \left(\overline{\text{var}}_n + (k-1)\overline{\text{cov}}_n\right)$$

Thus, weighted average is

$$\frac{k\overline{\text{cov}}_n}{\overline{\text{var}}_n + (k-1)\overline{\text{cov}}_n} = \alpha_n^*$$

# Alternative formulations

Stepped down alpha = weighted average of subtest alphas

- Interpretation is valid in general, even if parallel- or essential tau-equivalency do not hold

Reformulation:
Alpha is equal to stepped up weighted average of subtest alphas

Additional result:
Alpha is equal to weighted average of stepped up subtest alphas

Step up function and weighted average function are commuting functions on a space of alpha coefficients

# Standardized alpha

Common definition of standardized alpha is

$$\alpha_n^s = \frac{n\overline{\text{cor}}_n}{1 + (n-1)\overline{\text{cor}}_n}.$$

where

- $n$ is the number of items
- $\overline{\text{cor}}_n$ is the average correlation

Cronbach (1951, p. 321)

- if item variances are unknown
- used when big differences in item variances

# Standardized alpha

Common definition of standardized alpha is

$$\alpha_n^s = \frac{n\overline{\text{cor}}_n}{1+(n-1)\overline{\text{cor}}_n}.$$

Alternative definition of Cronbach's alpha

$$\alpha_n = \frac{n\overline{\text{cov}}_n}{\overline{\text{var}}_n+(n-1)\overline{\text{cov}}_n}$$

Since we have not used any properties of variances and covariances all results for alpha also hold for standardized alpha

# Corollary

We have $\alpha_n^* \leq \alpha_n \Leftrightarrow$

$$\frac{k}{\overline{\text{var}}_n + (k-1)\overline{\text{cov}}_n} \leq \frac{n}{\overline{\text{var}}_n + (n-1)\overline{\text{cov}}_n}$$
$$\Updownarrow$$
$$k\overline{\text{var}}_n + k(n-1)\overline{\text{cov}}_n \leq n\overline{\text{var}}_n + n(k-1)\overline{\text{cov}}_n$$
$$\Updownarrow$$
$$(n-k)\overline{\text{cov}}_n < (n-k)\overline{\text{var}}_n.$$

Since $k < n$, this inequality is equivalent to $\overline{\text{cov}}_n \leq \overline{\text{var}}_n$

# Alpha can be decreased

There exists a subtest alpha $\alpha_k$ such that $\alpha_k \leq \alpha_n$ (equality iff $\overline{\text{cov}}_n = \overline{\text{var}}_n$)

- $\alpha_n^* \leq \alpha_n$
- $\alpha_n^*$ is a weighted average of the $\alpha_k$'s

In general:
possible to decrease alpha by removing some of the items

- Alpha depends on number of items
- Makes sense to calculate 'alpha if item deleted'

# References


- Cortina JM (1993) What is coefficient alpha? An examination of theory and applications. Journal of Applied Psychology 78:98-104.
- Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. Psychometrika 16:297-334.
- Flanagan JC (1937) A proposed procedure for increasing the efficiency of objective tests. Journal of Educational Psychology 28:17-21.
- Raju NS (1977) A generalization of coefficient alpha. Psychometrika 42:549-565.
- Rulon PJ (1937) A simplified procedure for determining the reliability of a test by split-halves. Harvard Educational Review 9:99-103.
- Sijtsma K (2009) On the use, the misuse, and the very limited usefulness of Cronbach's alpha. Psychometrika 74:107-120.