WIKIPEDIA

Cronbach's alpha

Cronbach's alpha (Cronbach's α), also known as **tau-equivalent reliability** (ρ_T) or **coefficient alpha** (coefficient α), is a reliability coefficient that provides a method of measuring internal consistency of tests.[1][2][3]

Numerous studies warn against using it unconditionally.[4][5][6][7][8][9] Reliability coefficients based on structural equation modeling (SEM) are often recommended as an alternative.

Contents

Formula and calculation

Systematic and conventional formula

Let X_i denote the observed score of item i and $X = X_1 + X_2 + \cdots + X_k$ denote the sum of all items in a test consisting of k items. Let σ_{ij} denote the covariance between X_i and X_j, $\sigma_i^2 (= \sigma_{ii})$ denote the variance of X_i, and σ_X^2 denote the variance of X. σ_X^2 consists of item variances and inter-item covariance:

$$\sigma_X^2 = \sum_{i=1}^k \sum_{j=1}^k \sigma_{ij} = \sum_{i=1}^k \sigma_i^2 + \sum_{i=1}^k \sum_{j \neq i}^k \sigma_{ij}.$$

Let $\bar{\sigma}_{ij}$ denote the average of the inter-item covariance:

$$\bar{\sigma}_{ij} = \frac{\sum_{i=1}^k \sum_{j \neq i}^k \sigma_{ij}}{k(k-1)}.$$

ρ_T's systematic[3] formula is

$$\rho_T = \frac{k^2 \bar{\sigma}_{ij}}{\sigma_X^2}.$$

The more frequently used version of the formula is

$$\rho_T = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_X^2}\right).$$

Calculation example

When applied to appropriate data

ρ_T is applied to the following data that satisfies the condition of being tau-equivalent.

Observed covariance matrix

	X_1	X_2	X_3	X_4
X_1	10	6	6	6
X_2	6	11	6	6
X_3	6	6	12	6
X_4	6	6	6	13

$k = 4, \bar{\sigma}_{ij} = 6,$

$$\sigma_X^2 = \sum_{i=1}^{k} \sigma_i^2 + \sum_{i=1}^{k} \sum_{j \neq i}^{k} \sigma_{ij} = (10 + 11 + 12 + 13) + 4 \cdot (4 - 1) \cdot 6 = 118,$$

and $\rho_T = \dfrac{4^2 \cdot 6}{118} = 0.8135.$

When applied to inappropriate data

ρ_T is applied to the following data that does not satisfy the condition of being tau-equivalent.

Observed covariance matrix

	X_1	X_2	X_3	X_4
X_1	10	4	5	7
X_2	4	11	6	8
X_3	5	6	12	9
X_4	7	8	9	13

$k = 4, \bar{\sigma}_{ij} = (4 + 5 + 6 + 7 + 8 + 9)/6 = 6.5,$

$\sigma_X^2 = (10 + 11 + 12 + 13) + 2(4 + 5 + 6 + 7 + 8 + 9) = 124,$

and $\rho_T = \dfrac{4^2 \cdot 6.5}{124} = 0.8387.$

Compare this value with the value of applying congeneric reliability to the same data.

Prerequisites for using Cronbach's alpha

In order to use ρ_T as a reliability coefficient, the data must satisfy the following conditions.

1. Uni-dimensionality

2. Tau-equivalence (essential)
3. Independence between errors

Parallel condition

At the population level, parallel data have equal inter-item covariance (i.e., non-diagonal elements of the covariance matrix) and equal variances (i.e., diagonal elements of the covariance matrix). For example, the following data satisfy the parallel condition.

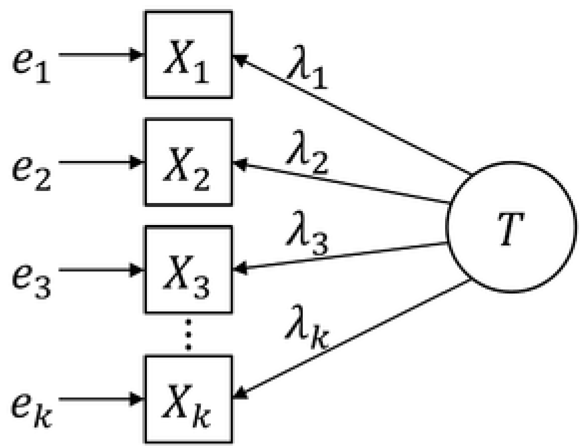
Observed covariance matrix

	X_1	X_2	X_3	X_4
X_1	10	6	6	6
X_2	6	10	6	6
X_3	6	6	10	6
X_4	6	6	6	10

In parallel data, even if a correlation matrix is used instead of a covariance matrix, there is no loss of information. All parallel data are tau-equivalent, but the reverse is not true. That is, among the three conditions, the parallel condition is most difficult to meet.

Tau-equivalent condition

At the population level, tau-equivalent data have equal covariance, but their variances may have different values. For example, the following data satisfies the condition of being tau-equivalent. All items in tau-equivalent data have equal discrimination or importance. All tau-equivalent data are also congeneric, but the reverse is not true.



A tau-equivalent measurement model is a special case of a congeneric measurement model, hereby assuming all factor loadings to be the same, i.e. $\lambda = \lambda_1 = \lambda_2 = \lambda_3 = \cdots = \lambda_k$

Observed covariance matrix

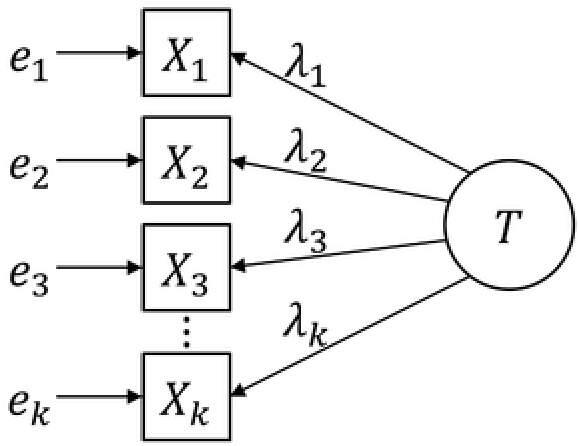
	X_1	X_2	X_3	X_4
X_1	10	6	6	6
X_2	6	12	6	6
X_3	6	6	9	6
X_4	6	6	6	10

Congeneric condition

At the population level, congeneric data need not have equal variances or covariance, provided they are unidimensional. For example, the following data meet the condition of being congeneric. All items in congeneric data can have different discrimination or importance.

Observed covariance matrix

	X_1	X_2	X_3	X_4
X_1	5	4	3	2
X_2	4	20	12	8
X_3	3	12	13	6
X_4	2	8	6	8



Congeneric measurement model

Relationship with other reliability coefficients

Classification of single-administration reliability coefficients

Conventional names

There are numerous reliability coefficients. Among them, the conventional names of reliability coefficients that are related and frequently used are summarized as follows:[3]

Conventional names of reliability coefficients

	Split-half	Unidimensional	Multidimensional
Parallel	Spearman-Brown formula	Standardized α	(No conventional name)
Tau-equivalent	Flanagan formula Rulon formula Flanagan-Rulon formula Guttman's λ_4	Cronbach's α coefficient α Guttman's λ_3 KR-20 Hoyt reliability	Stratified α
Congeneric	Angoff-Feldt coefficient Raju(1970) coefficient	composite reliability construct reliability congeneric reliability coefficient ω unidimensional ω Raju(1977) coefficient	coefficient ω ω total McDonald's ω multidimensional ω

Combining row and column names gives the prerequisites for the corresponding reliability coefficient. For example, Cronbach's α and Guttman's λ_3 are reliability coefficients derived under the condition of being unidimensional and tau-equivalent.

Systematic names

Conventional names are disordered and unsystematic. Conventional names give no information about the nature of each coefficient, or give misleading information (e.g., composite reliability). Conventional names are inconsistent. Some are formulas, and others are coefficients. Some are named after the original developer, some are named after someone who is not the original developer, and others do not include the name of any person. While one formula is referred to by multiple names, multiple formulas are referred to by one notation (e.g., alphas and omegas). The proposed systematic names and their notation for these reliability coefficients are as follows:[3]

Systematic names of reliability coefficients

	Split-half	Unidimensional	Multidimensional
Parallel	split-half parallel reliability (ρ_{SP})	parallel reliability (ρ_P)	multidimensional parallel reliability (ρ_{MP})
Tau-equivalent	split-half tau-equivalent reliability (ρ_{ST})	tau-equivalent reliability (ρ_T)	multidimensional tau-equivalent reliability (ρ_{MT})
Congeneric	split-half congeneric reliability (ρ_{SC})	congeneric reliability (ρ_C)	**Bifactor model** Bifactor reliability (ρ_{BF}) **Second-order factor model** Second-order factor reliability (ρ_{SOF}) **Correlated factor model** Correlated factor reliability (ρ_{CF})

Parallel reliability

ρ_T is often referred to as coefficient alpha and ρ_P is often referred to as standardized alpha. Because of the standardized modifier, ρ_P is often mistaken for a more standard version than ρ_T. There is no historical basis to refer to ρ_P as standardized alpha. Cronbach (1951)[10] did not refer to this coefficient as alpha, nor did he recommend using it. ρ_P was rarely used before the 1970s. As SPSS began to provide ρ_P under the name of standardized alpha, this coefficient began to be used occasionally.[11] The use of ρ_P is not recommended because the parallel condition is difficult to meet in real-world data.

Split-half tau-equivalent reliability

ρ_T equals the average of the ρ_{ST} values obtained for all possible split-halves. This relationship, proved by Cronbach (1951),[10] is often used to explain the intuitive meaning of ρ_T. However, this interpretation overlooks the fact that ρ_{ST} underestimates reliability when applied to data that are not tau-equivalent. At the population level, the maximum of all possible ρ_{ST} values is closer to reliability than the average of all possible ρ_{ST} values.[7] This mathematical fact was already known even before the publication of Cronbach (1951).[12] A comparative study[13] reports that the maximum of ρ_{ST} is the most accurate reliability coefficient.

Revelle (1979) refers to the minimum of all possible ρ_{ST} values as coefficient β,[14] and recommends that β provides complementary information that ρ_T does not.[6]

Congeneric reliability

If the assumptions of uni-dimensionality and tau-equivalence are satisfied, ρ_T equals ρ_C.

If uni-dimensionality is satisfied but tau-equivalence is not satisfied, ρ_T is smaller than ρ_C.[7]

ρ_C is the most commonly used reliability coefficient after ρ_T. Users tend to present both, rather than replacing ρ_T with ρ_C.[3]

A study investigating studies that presented both coefficients reports that ρ_T is 0.02 smaller than ρ_C on average.[15]

Multidimensional reliability coefficients

If ρ_T is applied to multidimensional data, its value is smaller than multidimensional reliability coefficients and larger than ω_T.[3]

Intraclass correlation

ρ_T is said to be equal to the stepped-up consistency version of the intraclass correlation coefficient, which is commonly used in observational studies. But this is only conditionally true. In terms of variance components, this condition is, for item sampling: if and only if the value of the item (rater, in the case of rating) variance component equals zero. If this variance component is negative, ρ_T will underestimate the stepped-up intra-class correlation coefficient; if this variance component is positive, ρ_T will overestimate this stepped-up intra-class correlation coefficient.

History

Before 1937

ρ_{SP}[16][17] was the only known reliability coefficient. The problem was that the reliability estimates depended on how the items were split in half (e.g., odd/even or front/back). Criticism was raised against this unreliability, but for more than 20 years no fundamental solution was found.[18]

Kuder and Richardson (1937)

Kuder and Richardson (1937)[19] developed several reliability coefficients that could overcome the problem of ρ_{SP}. They did not give the reliability coefficients particular names. Equation 20 in their article is ρ_T. This formula is often referred to as Kuder–Richardson Formula 20, or KR-20. They dealt with cases where the observed scores were dichotomous (e.g., correct or incorrect), so the expression of KR-20 is slightly different from the conventional formula of ρ_T. A review of this paper reveals that they did not present a general formula because they did not need to, not because they were not able to. Let p_i denote the correct answer ratio of item i, and q_i denote the incorrect answer ratio of item i ($p_i + q_i = 1$). The formula of KR-20 is as follows.

$$\rho_{\text{KR-20}} = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} p_i q_i}{\sigma_X^2}\right)$$

Since $p_i q_i = \sigma_i^2$, KR-20 and ρ_T have the same meaning.

Between 1937 and 1951

Other versions of the general formula of KR-20

Kuder and Richardson (1937) made unnecessary assumptions to derive ρ_T. Several studies have derived ρ_T in a different way from Kuder and Richardson (1937).

Hoyt (1941)[20] derived ρ_T using ANOVA (Analysis of variance). Cyril Hoyt may be considered the first developer of the general formula of the KR-20, but he did not explicitly present the formula of ρ_T.

The first expression of the modern formula of ρ_T appears in Jackson and Ferguson (1941).[21] The version they presented is as follows. Edgerton and Thompson (1942)[22] used the same version.

$$\rho_T = \frac{k}{k-1}\left(\frac{\sigma_X^2 - \sum_{i=1}^{k}\sigma_i^2}{\sigma_X^2}\right)$$

Guttman (1945)[12] derived six reliability formulas, each denoted by $\lambda_1, \cdots, \lambda_6$. Louis Guttman proved that all of these formulas were always less than or equal to reliability, and based on these characteristics, he referred to these formulas as 'lower bounds of reliability'. Guttman's λ_4 is ρ_{ST}, and λ_3 is ρ_T. He proved that λ_2 is always greater than or equal to λ_3 (i.e., more accurate). At that time, all calculations were done with paper and pencil, and since the formula of λ_3 was simpler to calculate, he mentioned that λ_3 was useful under certain conditions.

$$\lambda_3 = \rho_T = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k}\sigma_i^2}{\sigma_X^2}\right)$$

Gulliksen (1950)[23] derived ρ_T with fewer assumptions than previous studies. The assumption he used is essential tau-equivalence in modern terms.

Recognition of KR-20 at the time

The original and general KR-10 formulas were recognized to be exactly identical, and the expression of general formula of KR-20 was not used. Hoyt[20]:156 explained that his method "gives precisely the same result" as KR-20. Jackson and Ferguson[21]:74 stated that the two formulas are "identical". Guttman[12]:275 said λ_3 is "algebraically identical" to KR-20. Gulliksen[23]:224 also admitted that the two formulas are "identical".

Even studies critical of KR-20 did not point out that the original formula of KR-20 could only be applied to dichotomous data.[24]

Criticism of underestimation of KR-20

Developers[19] of this formula reported that ρ_T consistently underestimates reliability. Hoyt[25] argued that this characteristic alone made ρ_T more recommendable than the traditional split-half technique, which was unknown whether to underestimate or overestimate reliability.

Cronbach (1943)[24] was critical of the underestimation of ρ_T. He was concerned that it was not known how much ρ_T underestimated reliability. He criticized that the underestimation was likely to be excessively severe, such that ρ_T could sometimes lead to negative values. Because of these problems, he argued that ρ_T could not be recommended as an alternative to the split-half technique.

Cronbach (1951)

As with previous studies,[20][12][21][23] Cronbach (1951)[10] invented another method to derive ρ_T. His interpretation was more intuitively attractive than those of previous studies. That is, he proved that ρ_T equals the average of ρ_{ST} values obtained for all possible split-halves. He criticized that the name KR-20 was weird and suggested a new name, coefficient alpha. His approach has been a huge success. However, he not only omitted some key facts, but also gave an incorrect explanation.

First, he positioned coefficient alpha as a general formula of KR-20, but omitted the explanation that existing studies had published the precisely identical formula. Those who read only Cronbach (1951) without background knowledge could misunderstand that he was the first to develop the general formula of KR-20.

Second, he did not explain under what condition ρ_T equals reliability. Non-experts could misunderstand that ρ_T was a general reliability coefficient that could be used for all data regardless of prerequisites.

Third, he did not explain why he changed his attitude toward ρ_T. In particular, he did not provide a clear answer to the underestimation problem of ρ_T, which he himself[24] had criticized.

Fourth, he argued that a high value of ρ_T indicated homogeneity of the data.

After 1951

Novick and Lewis (1967)[26] proved the necessary and sufficient condition for ρ_T to be equal to reliability, and named it the condition of being essentially tau-equivalent.

Cronbach (1978)[2]:263 mentioned that the reason Cronbach (1951) received a lot of citations was "mostly because [he] put a brand name on a common-place coefficient".[3] He explained that he had originally planned to name other types of reliability coefficients (e.g., inter-rater reliability or test-retest reliability) in consecutive Greek letter (e.g., β, γ, $\ldots$), but later changed his mind.

Cronbach and Schavelson (2004)[27] encouraged readers to use generalizability theory rather than ρ_T. He opposed the use of the name Cronbach's alpha. He explicitly denied the existence of studies that had published the general formula of KR-20 prior to Cronbach (1951).

Common misconceptions

[7]

The value of Cronbach's alpha ranges between zero and one

By definition, reliability cannot be less than zero and cannot be greater than one. Many textbooks mistakenly equate ρ_T with reliability and give an inaccurate explanation of its range. ρ_T can be less than reliability when applied to data that are not tau-equivalent. Suppose that X_2 copied the value of X_1 as it is, and X_3 copied by multiplying the value of X_1 by -1. The covariance matrix between items is as follows, $\rho_T = -\mathbf{3}$.

Observed covariance matrix

	X_1	X_2	X_3
X_1	1	1	−1
X_2	1	1	−1
X_3	−1	−1	1

Negative ρ_T can occur for reasons such as negative discrimination or mistakes in processing reversely scored items.

Unlike ρ_T, SEM-based reliability coefficients (e.g., ρ_C) are always greater than or equal to zero.

This anomaly was first pointed out by Cronbach (1943)[24] to criticize ρ_T, but Cronbach (1951)[10] did not comment on this problem in his article, which discussed all conceivable issues related ρ_T and he himself[27]:396 described as being "encyclopedic".

If there is no measurement error, the value of Cronbach's alpha is one

This anomaly also originates from the fact that ρ_T underestimates reliability. Suppose that X_2 copied the value of X_1 as it is, and X_3 copied by multiplying the value of X_1 by two. The covariance matrix between items is as follows, $\rho_T = 0.9375$.

Observed covariance matrix

	X_1	X_2	X_3
X_1	1	1	2
X_2	1	1	2
X_3	2	2	4

For the above data, both ρ_P and ρ_C have a value of one.

The above example is presented by Cho and Kim (2015).[7]

A high value of Cronbach's alpha indicates homogeneity between the items

Many textbooks refer to ρ_T as an indicator of homogeneity between items. This misconception stems from the inaccurate explanation of Cronbach (1951)[10] that high ρ_T values show homogeneity between the items. Homogeneity is a term that is rarely used in the modern literature, and related studies interpret the term as referring to uni-dimensionality. Several studies have provided proofs or counterexamples that high ρ_T values do not indicate unidimensionality.[28][7][29][30][31][32] See counterexamples below.

Unidimensional data

	X_1	X_2	X_3	X_4	X_5	X_6
X_1	10	3	3	3	3	3
X_2	3	10	3	3	3	3
X_3	3	3	10	3	3	3
X_4	3	3	3	10	3	3
X_5	3	3	3	3	10	3
X_6	3	3	3	3	3	10

$\rho_T = 0.72$ in the unidimensional data above.

Multidimensional data

	X_1	X_2	X_3	X_4	X_5	X_6
X_1	10	6	6	1	1	1
X_2	6	10	6	1	1	1
X_3	6	6	10	1	1	1
X_4	1	1	1	10	6	6
X_5	1	1	1	6	10	6
X_6	1	1	1	6	6	10

$\rho_T = 0.72$ in the multidimensional data above.

Multidimensional data with extremely high reliability

	X_1	X_2	X_3	X_4	X_5	X_6
X_1	10	9	9	8	8	8
X_2	9	10	9	8	8	8
X_3	9	9	10	8	8	8
X_4	8	8	8	10	9	9
X_5	8	8	8	9	10	9
X_6	8	8	8	9	9	10

The above data have $\rho_T = 0.9692$, but are multidimensional.

Unidimensional data with unacceptably low reliability

	X_1	X_2	X_3	X_4	X_5	X_6
X_1	10	1	1	1	1	1
X_2	1	10	1	1	1	1
X_3	1	1	10	1	1	1
X_4	1	1	1	10	1	1
X_5	1	1	1	1	10	1
X_6	1	1	1	1	1	10

The above data have $\rho_T = 0.4$, but are unidimensional.

Uni-dimensionality is a prerequisite for ρ_T. You should check uni-dimensionality before calculating ρ_T, rather than calculating ρ_T to check uni-dimensionality.[3]

A high value of Cronbach's alpha indicates internal consistency

The term internal consistency is commonly used in the reliability literature, but its meaning is not clearly defined. The term is sometimes used to refer to a certain kind of reliability (e.g., internal consistency reliability), but it is unclear exactly which reliability coefficients are included here, in addition to ρ_T. Cronbach (1951)[10] used the term in several senses without an explicit definition. Cho and Kim (2015)[7] showed that ρ_T is not an indicator of any of these.

Removing items using "alpha if item deleted" always increases reliability

Removing an item using "alpha if item deleted" may result in 'alpha inflation,' where sample-level reliability is reported to be higher than population-level reliability.[33] It may also reduce population-level reliability.[34] The elimination of less-reliable items should be based not only on a statistical basis, but also on a theoretical and logical basis. It is also recommended that the whole sample be divided into two and cross-validated.[33]

Ideal reliability level and how to increase reliability

Nunnally's recommendations for the level of reliability

The most frequently cited source of how high reliability coefficients should be is Nunnally's book.[35][36][37] However, his recommendations are cited contrary to his intentions. What he meant was to apply different criteria depending on the purpose or stage of the study. However, regardless of the nature of the research, such as exploratory research, applied research, and scale development research, a criterion of 0.7 is universally used.[38] 0.7 is the criterion he recommended for the early stages of a study, which most studies published in the journal are not. Rather than 0.7, the criterion of 0.8 referred to applied research by Nunnally is more appropriate for most empirical studies.[38]

Nunnally's recommendations on the level of reliability

	1st edition[35]	2nd[36] & 3rd[37] edition
Early stage of research	0.5 or 0.6	0.7
Applied research	0.8	0.8
When making important decisions	0.95 (minimum 0.9)	0.95 (minimum 0.9)

His recommendation level did not imply a cutoff point. If a criterion means a cutoff point, it is important whether or not it is met, but it is unimportant how much it is over or under. He did not mean that it should be strictly 0.8 when referring to the criteria of 0.8. If the reliability has a value near 0.8 (e.g., 0.78), it can be considered that his recommendation has been met.[39]

Cost to obtain a high level of reliability

Nunnally's idea was that there is a cost to increasing reliability, so there is no need to try to obtain maximum reliability in every situation.

Trade-off with validity

Measurements with perfect reliability lack validity.[7] For example, a person who take the test with the reliability of one will get a perfect score or a zero score, because the examinee who gives the correct answer or incorrect answer on one item will give the correct answer or incorrect answer on all other items. The phenomenon in which validity is sacrificed to increase reliability is called attenuation paradox.[40][41]

A high value of reliability can be in conflict with content validity. For high content validity, each item should be constructed to be able to comprehensively represent the content to be measured. However, a strategy of repeatedly measuring essentially the same question in different ways is often used only for the purpose of increasing reliability.[42][43]

Trade-off with efficiency

When the other conditions are equal, reliability increases as the number of items increases. However, the increase in the number of items hinders the efficiency of measurements.

Methods to increase reliability

Despite the costs associated with increasing reliability discussed above, a high level of reliability may be required. The following methods can be considered to increase reliability.

Before data collection:

- Eliminate the ambiguity of the measurement item.
- Do not measure what the respondents do not know.
- Increase the number of items. However, care should be taken not to excessively inhibit the efficiency of the measurement.
- Use a scale that is known to be highly reliable.[44]
- Conduct a pretest - discover in advance the problem of reliability.
- Exclude or modify items that are different in content or form from other items (e.g., reverse-scored items).

After data collection:

- Remove the problematic items using "alpha if item deleted". However, this deletion should be accompanied by a theoretical rationale.
- Use a more accurate reliability coefficient than ρ_T. For example, ρ_C is 0.02 larger than ρ_T on average.[15]

Which reliability coefficient to use

ρ_T is used in an overwhelming proportion. A study estimates that approximately 97% of studies use ρ_T as a reliability coefficient.[3]

However, simulation studies comparing the accuracy of several reliability coefficients have led to the common result that ρ_T is an inaccurate reliability coefficient.[45][13][6][46][47]

Methodological studies are critical of the use of ρ_T. Simplifying and classifying the conclusions of existing studies are as follows.

1. Conditional use: Use ρ_T only when certain conditions are met.[3][7][9]
2. Opposition to use: ρ_T is inferior and should not be used.[48][5][49][6][4][50]

Alternatives to Cronbach's alpha

Existing studies are practically unanimous in that they oppose the widespread practice of using ρ_T unconditionally for all data. However, different opinions are given on which reliability coefficient should be used instead of ρ_T.

Different reliability coefficients ranked first in each simulation study[45][13][6][46][47] comparing the accuracy of several reliability coefficients.[7]

The majority opinion is to use SEM-based reliability coefficients as an alternative to ρ_T.[3][7][48][5][49][9][6][50]

However, there is no consensus on which of the several SEM-based reliability coefficients (e.g., unidimensional or multidimensional models) is the best to use.

Some people suggest ω_H[6] as an alternative, but ω_H shows information that is completely different from reliability. ω_H is a type of coefficient comparable to Revelle's β.[14][6] They do not substitute, but complement reliability.[3]

Among SEM-based reliability coefficients, multidimensional reliability coefficients are rarely used, and the most commonly used is ρ_C.[3]

Software for SEM-based reliability coefficients

General-purpose statistical software such as SPSS and SAS include a function to calculate ρ_T. Users who don't know the formula of ρ_T have no problem in obtaining the estimates with just a few mouse clicks.

SEM software such as AMOS, LISREL, and MPLUS does not have a function to calculate SEM-based reliability coefficients. Users need to calculate the result by inputting it to the formula. To avoid this inconvenience and possible error, even studies reporting the use of SEM rely on ρ_T instead of SEM-based reliability coefficients.[3] There are a few alternatives to automatically calculate SEM-based reliability coefficients.

1. R (free): The psych package[51] calculates various reliability coefficients.
2. EQS (paid):[52] This SEM software has a function to calculate reliability coefficients.
3. RelCalc (free):[3] Available with Microsoft Excel. ρ_C can be obtained without the need for SEM software. Various multidimensional SEM reliability coefficients and various types of ω_H can be calculated based on the results of SEM software.

Derivation of formula

The following is the derivation of formula.[3]

Assumptions:

1. The observed score of an item consists of the true score of the item and the error of the item, which is independent of the true score, $X_i = T_i + e_i$ Lemma. $\operatorname{Cov}(T_i, e_i) = \operatorname{Cov}(T_i, e_j) = 0, \forall i \neq j$
2. Errors are independent of each other, $\operatorname{Cov}(e_i, e_j) = 0, \forall i \neq j$
3. The assumption of being essentially tau-equivalent: the true score of an item consists of the true score common to all items and the constant of the item, $T_i = \mu_i + t$

Let T denote the sum of the item true scores. $T = \sum_{i=1}^{k} T_i$

The variance of T is called the true score variance.

Definition. Reliability is the ratio of true score variance to observed score variance. $\rho = \frac{\sigma_T^2}{\sigma_X^2}$

The following relationship is established from the above assumptions.

$$\sigma_i^2 = \operatorname{Var}(\mu_i + t + e_i) = \sigma_t^2 + \sigma_{e_i}^2 (\because \operatorname{Var}(\mu_i) = \operatorname{Cov}(t, e_i) = \operatorname{Cov}(\mu_i, e_i) = \operatorname{Cov}(\mu_i, t) =$$

$$\sigma_{ij} = \operatorname{Cov}(T_i + e_i, T_j + e_j) = \sigma_t^2 (\because \operatorname{Cov}(T_i, e_j) = \operatorname{Cov}(T_j, e_i) = \operatorname{Cov}(e_i, e_j) = 0)$$

Therefore, the covariance matrix between items is as follows.

Observed covariance matrix

	X_1	X_2	$\dots$	X_k
X_1	$\sigma_t^2+\sigma_{e_1}^2$	σ_t^2	$\dots$	σ_t^2
X_2	σ_t^2	$\sigma_t^2+\sigma_{e_2}^2$	$\dots$	σ_t^2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
X_k	σ_t^2	σ_t^2	$\dots$	$\sigma_t^2+\sigma_{e_k}^2$

You can see that σ_t^2 equals the mean of the co-variances between items. That is, $\sigma_t^2 = \overline{\sigma}_{ij}$

$$\sigma_T^2 = \text{Var}\left(\sum_{i=1}^{k} t\right) = k^2\sigma_t^2 = k^2\overline{\sigma}_{ij}$$

Let ρ_T denote the reliability when satisfying the above assumptions. ρ_T is:

$$\rho_T = \frac{k^2\overline{\sigma}_{ij}}{\sigma_X^2}$$

References

1. Cronbach, Lee J. (1951). "Coefficient alpha and the internal structure of tests". *Psychometrika*. Springer Science and Business Media LLC. **16** (3): 297–334. doi:10.1007/bf02310555 (https://doi.org/10.1007%2Fbf02310555). hdl:10983/2196 (https://hdl.handle.net/10983%2F2196). ISSN 0033-3123 (https://www.worldcat.org/issn/0033-3123). S2CID 13820448 (https://api.semanticscholar.org/CorpusID:13820448).
2. Cronbach, L. J. (1978). "Citation Classics" (http://garfield.library.upenn.edu/classics1978/A1978EQ39200002.pdf) (PDF). *Current Contents*. **13**: 263.
3. Cho, Eunseong (2016-07-08). "Making Reliability Reliable". *Organizational Research Methods*. SAGE Publications. **19** (4): 651–682. doi:10.1177/1094428116656239 (https://doi.org/10.1177%2F1094428116656239). ISSN 1094-4281 (https://www.worldcat.org/issn/1094-4281). S2CID 124129255 (https://api.semanticscholar.org/CorpusID:124129255).
4. Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. Psychometrika, 74(1), 107–120. doi:10.1007/s11336-008-9101-0 (https://doi.org/10.1007%2Fs11336-008-9101-0)
5. Green, S. B., & Yang, Y. (2009). Commentary on coefficient alpha: A cautionary tale. Psychometrika, 74(1), 121–135. doi:10.1007/s11336-008-9098-4 (https://doi.org/10.1007%2Fs11336-008-9098-4)
6. Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. Psychometrika, 74(1), 145–154. doi:10.1007/s11336-008-9102-z (https://doi.org/10.1007%2Fs11336-008-9102-z)
7. Cho, E., & Kim, S. (2015). Cronbach's coefficient alpha: Well known but poorly understood. Organizational Research Methods, 18(2), 207–230. doi:10.1177/1094428114555994 (https://doi.org/10.1177%2F1094428114555994)

8. McNeish, D. (2017). Thanks coefficient alpha, we'll take it from here. Psychological Methods, 23(3), 412–433. doi:10.1037/met0000144 (https://doi.org/10.1037%2Fmet0000144)
9. Raykov, T., & Marcoulides, G. A. (2017). Thanks coefficient alpha, we still need you! Educational and Psychological Measurement, 79(1), 200–210. doi:10.1177/0013164417725127 (https://doi.org/10.1177%2F0013164417725127)
10. Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16 (3), 297–334. doi:10.1007/BF02310555 (https://doi.org/10.1007%2FBF02310555)
11. Cho, E. and Chun, S. (2018), Fixing a broken clock: A historical review of the originators of reliability coefficients including Cronbach's alpha. Survey Research, 19(2), 23–54.
12. Guttman, L. (1945). A basis for analyzing test-retest reliability. Psychometrika, 10(4), 255–282. doi:10.1007/BF02288892 (https://doi.org/10.1007%2FBF02288892)
13. Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. Psychological Methods, 5(3), 343–355. doi:10.1037/1082-989X.5.3.343 (https://doi.org/10.1037%2F1082-989X.5.3.343)
14. Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. Multivariate Behavioral Research, 14(1), 57–74. doi:10.1207/s15327906mbr1401_4 (https://doi.org/10.1207%2Fs15327906mbr1401_4)
15. Peterson, R. A., & Kim, Y. (2013). On the relationship between coefficient alpha and composite reliability. Journal of Applied Psychology, 98(1), 194–198. doi:10.1037/a0030767 (https://doi.org/10.1037%2Fa0030767)
16. Brown, W. (1910). Some experimental results in the correlation of mental abilities. British Journal of Psychology, 3(3), 296–322. doi:10.1111/j.2044-8295.1910.tb00207.x (https://doi.org/10.1111%2Fj.2044-8295.1910.tb00207.x)
17. Spearman, C. (1910). Correlation calculated from faulty data. British Journal of Psychology, 3(3), 271–295. doi:10.1111/j.2044-8295.1910.tb00206.x (https://doi.org/10.1111%2Fj.2044-8295.1910.tb00206.x)
18. Kelley, T. L. (1924). Note on the reliability of a test: A reply to Dr. Crum's criticism. Journal of Educational Psychology, 15(4), 193–204. doi:10.1037/h0072471 (https://doi.org/10.1037%2Fh0072471)
19. Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. Psychometrika, 2(3), 151–160. doi:10.1007/BF02288391 (https://doi.org/10.1007%2FBF02288391)
20. Hoyt, C. (1941). Test reliability estimated by analysis of variance. Psychometrika, 6(3), 153–160. doi:10.1007/BF02289270 (https://doi.org/10.1007%2FBF02289270)
21. Jackson, R. W. B., & Ferguson, G. A. (1941). Studies on the reliability of tests. University of Toronto Department of Educational Research Bulletin, 12, 132.
22. Edgerton, H. A., & Thomson, K. F. (1942). Test scores examined with the lexis ratio. Psychometrika, 7(4), 281–288. doi:10.1007/BF02288629 (https://doi.org/10.1007%2FBF02288629)
23. Gulliksen, H. (1950). Theory of mental tests. John Wiley & Sons. doi:10.1037/13240-000 (https://doi.org/10.1037%2F13240-000)
24. Cronbach, L. J. (1943). On estimates of test reliability. Journal of Educational Psychology, 34(8), 485–494. doi:10.1037/h0058608 (https://doi.org/10.1037%2Fh0058608)
25. Hoyt, C. J. (1941). Note on a simplified method of computing test reliability: Educational and Psychological Measurement, 1(1). doi:10.1177/001316444100100109 (https://doi.org/10.1177%2F001316444100100109)
26. Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. Psychometrika, 32(1), 1–13. doi:10.1007/BF02289400 (https://doi.org/10.1007%2FBF02289400)

27. Cronbach, L. J., & Shavelson, R. J. (2004). My Current Thoughts on Coefficient Alpha and Successor Procedures. Educational and Psychological Measurement, 64(3), 391–418. doi:10.1177/0013164404266386 (https://doi.org/10.1177%2F0013164404266386)
28. Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. Journal of Applied Psychology, 78(1), 98–104. doi:10.1037/0021-9010.78.1.98 (https://doi.org/10.1037%2F0021-9010.78.1.98)
29. Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an Index of test unidimensionality. Educational and Psychological Measurement, 37(4), 827–838. doi:10.1177/001316447703700403 (https://doi.org/10.1177%2F001316447703700403)
30. McDonald, R. P. (1981). The dimensionality of tests and items. The British Journal of Mathematical and Statistical Psychology, 34(1), 100–117. doi:10.1111/j.2044-8317.1981.tb00621.x (https://doi.org/10.1111%2Fj.2044-8317.1981.tb00621.x)
31. Schmitt, N. (1996). Uses and abuses of coefficient alpha. Psychological Assessment, 8(4), 350–353. doi:10.1037/1040-3590.8.4.350 (https://doi.org/10.1037%2F1040-3590.8.4.350)
32. Ten Berge, J. M. F., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. Psychometrika, 69(4), 613–625. doi:10.1007/BF02289858 (https://doi.org/10.1007%2FBF02289858)
33. Kopalle, P. K., & Lehmann, D. R. (1997). Alpha inflation? The impact of eliminating scale items on Cronbach's alpha. Organizational Behavior and Human Decision Processes, 70(3), 189–197. doi:10.1006/obhd.1997.2702 (https://doi.org/10.1006%2Fobhd.1997.2702)
34. Raykov, T. (2007). Reliability if deleted, not 'alpha if deleted': Evaluation of scale reliability following component deletion. The British Journal of Mathematical and Statistical Psychology, 60(2), 201–216. doi:10.1348/000711006X115954 (https://doi.org/10.1348%2F000711006X115954)
35. Nunnally, J. C. (1967). Psychometric theory. New York, NY: McGraw-Hill.
36. Nunnally, J. C. (1978). Psychometric theory (2nd ed.). New York, NY: McGraw-Hill.
37. Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.). New York, NY: McGraw-Hill.
38. Lance, C. E., Butts, M. M., & Michels, L. C. (2006). What did they really say? Organizational Research Methods, 9(2), 202–220. doi:10.1177/1094428105284919 (https://doi.org/10.1177%2F1094428105284919)
39. Cho, E. (2020). A comprehensive review of so-called Cronbach's alpha. Journal of Product Research, 38(1), 9–20.
40. Loevinger, J. (1954). The attenuation paradox in test theory. Psychological Bulletin, 51(5), 493–504. doi:10.1002/j.2333-8504.1954.tb00485.x (https://doi.org/10.1002%2Fj.2333-8504.1954.tb00485.x)
41. Humphreys, L. (1956). The normal curve and the attenuation paradox in test theory. Psychological Bulletin, 53(6), 472–476. doi:10.1037/h0041091 (https://doi.org/10.1037%2Fh0041091)
42. Boyle, G. J. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? Personality and Individual Differences, 12(3), 291–294. doi:10.1016/0191-8869(91)90115-R (https://doi.org/10.1016%2F0191-8869%2891%2990115-R)
43. Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. Journal of Personality Assessment, 80(1), 99–103. doi:10.1207/S15327752JPA8001_18 (https://doi.org/10.1207%2FS15327752JPA8001_18)
44. Lee, H. (2017). Research Methodology (2nd ed.), Hakhyunsa.
45. Kamata, A., Turhan, A., & Darandari, E. (2003). Estimating reliability for multidimensional composite scale scores. Annual Meeting of American Educational Research Association, Chicago, April 2003, April, 1–27.

46. Tang, W., & Cui, Y. (2012). A simulation study for comparing three lower bounds to reliability. Paper Presented on April 17, 2012 at the AERA Division D: Measurement and Research Methodology, Section 1: Educational Measurement, Psychometrics, and Assessment, 1–25.
47. van der Ark, L. A., van der Palm, D. W., & Sijtsma, K. (2011). A latent class approach to estimating test-score reliability. Applied Psychological Measurement, 35(5), 380–392. doi:10.1177/0146621610392911 (https://doi.org/10.1177%2F0146621610392911)
48. Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. British Journal of Psychology, 105(3), 399–412. doi:10.1111/bjop.12046 (https://doi.org/10.1111%2Fbjop.12046)
49. Peters, G. Y. (2014). The alpha and the omega of scale reliability and validity comprehensive assessment of scale quality. The European Health Psychologist, 1(2), 56–69.
50. 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 (https://doi.org/10.1177%2F0734282911406668)
51. Revelle, William (7 January 2017). "An overview of the psych package" (http://personality-project.org/r/overview.pdf) (PDF).
52. "Multivariate Software, Inc" (https://web.archive.org/web/20010521070751/http://www.mvsoft.com/eqs60.htm). *www.mvsoft.com*. Archived from the original (http://www.mvsoft.com/eqs60.htm) on 2001-05-21.

External links

- Cronbach's alpha SPSS tutorial (http://www.open.ac.uk/socialsciences/spsstutorial/files/tutorials/cronbachs-alpha.pdf)
- The free web interface and R package cocron (http://comparingcronbachalphas.org) allows to statistically compare two or more dependent or independent Cronbach alpha coefficients.