

取代还是共存

——信度估计系数阿尔法与欧米伽

高清辉

厦门大学，厦门

摘要 | 在众多信度指标中，克隆巴赫 α (Cronbach's α) 系数最为常用，但 α 系数在使用中常常忽视重要的“基本 τ 等价”假设，导致 α 系数并非真正的信度；并且存在着误用与滥用、阈值不确定等问题，受到很多批评，被长期广泛应用是因其易于计算、易于理解等特性。麦克唐纳 ω 克服了 α 系数的上述缺点，信度估计上明显优于 α 系数，但麦克唐纳 ω 也存在计算要求较高等缺点，尚无法取代克隆巴赫 α 。两个信度指标 ω 与 α 将长期共存，在研究报告中应同时提供 ω 与 α 的数值，互相补充。

关键词 | 信度估计；克隆巴赫 α ；麦克唐纳 ω

Copyright © 2022 by author (s) and SciScan Publishing Limited

This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/). <https://creativecommons.org/licenses/by-nc/4.0/>



1 引言

在心理学和一些实证行为科学研究中，经常要关注一些无法直接测量的量或结构，例如某种能力、心理压力等。多项目测量量表是研究这些量或结构的常用方法：设置若干个能表达隐性目标量的项目，考察对象的反应，运用某种方式进行数字的聚合（常见如加总或平均），以生成每个被测量对象的测量值。在此类测试中，测量结果的稳定程度即信度是一个非常重要的指标。在众多信度指标中，克隆巴赫 α (Cronbach's α) 系数最为常用。自从克隆巴赫在 1951 年提出 α 系数^[1]之后， α 系数得到了广泛的应用，且被引用的次数极高，按谷歌学术搜索的统计，到 2021 年其被引用次数已超过 54000 次。虽然 α 系数得到如此多的引用，但对其的批评也很多，有的甚至很尖锐。孟庆茂和刘红云指出“用 α 系数对测验进行评价的确存在一些问题”^[2]。西茨玛 (Sijtsma) 认为“想必没有其他统计数据受到如此多的误解和混淆”^[3]。彼得斯 (Peters) 建议放弃使用 α 系数，因为“Cronbach's α 值的

基金项目：国家社科基金资助项目（20BYY117）。

作者简介：高清辉，厦门大学副教授，博士，研究方向：教育测评、国际中文教育。

文章引用：高清辉. 取代还是共存——信度估计系数阿尔法与欧米伽 [J]. 中国心理学前沿, 2022, 4 (8): 954-961.

<https://doi.org/10.35534/pc.0408114>

用处十分有限”^[4]。许多学者提议用麦克唐纳的 ω ^[5]取代 α ，认为研究人员应该从 α 转向 ω ，但因种种原因目前仍是 α 占主导地位。在2021年，前述克隆巴赫这篇开创性的论文^[1]发表70周年之际，心理测量学界的权威刊物“Psychometrika”以“克隆巴赫 α 的再审视”为专题发表了5篇专栏论文，表明了心理测量学界对该问题的关切。正如温忠麟所言， α 是“晃而不倒的信度标杆”^[6]，“晃”是因为其自身的局限性而受到多方批评，“不倒”是它在新条件下仍有存在的价值。那么，作为信度指标， α 系数具体有哪些缺点？ α 系数既然有那么多的局限性，为何还会被广泛应用呢？能否用 ω 取代 α ？我们将对这些问题作探讨。

2 作为信度指标的 α 系数

2.1 克隆巴赫 α 系数

在许多文献中， α 通常被直接看作测试的信度，但这是有条件的。信度的定义来自经典测试理论（CTT），一个测试的信度是一个被测团体真分数的变异数与实际分数变异数的比值。根据信度的理论定义实际求出一个测试的信度是困难的，测试中我们只能得到实际分数，而真分数与测量值的误差都是未知的，解决方法一是添加一些假定，缩小不确定性；二是估计信度的下界。克隆巴赫Cronbach对信度的估计用 α 表示^[1]，即

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \text{Var}(X_i)}{\text{Var}(X)} \right) \quad (1)$$

其中， X 是由 k 个项目组成的测试， $X_1, X_2, X_3, \dots, X_k$ 是观测到的项目分数。他证明了 α 是所有可能的分半信度的平均值。因此 α 在概念上等同于同一场合实施的两平行形式的测试之间的相关性。在没有计算机辅助计算的年代，通过简单的运算就可得到所有可能分半信度的平均值，这是一种“快速且安全”的估计信度的方法。此后，它就被称为“Cronbach's α ”而广泛应用于表示测试的信度，但也不断受到质疑，主要有以下四个方面。

2.1.1 重要的假定被忽视

腾伯格与索康（TenBerge and Socan）证明了在CTT假设下 α 和信度之间有如下关系^[7]。

$$\alpha + \frac{k}{(k-1) \text{Var}(X)} \sum_{i < j} \text{Var}(T_i - T_j) = \frac{\text{Var}(T)}{\text{Var}(X)} \quad (2)$$

X 是由 k 个项目组成的测试， $X_1, X_2, X_3, \dots, X_k$ 是观测到的项目分数； T_i 是第 i 项目的真分数。由式（2）可得两个结论：（1） α 是信度的下界，式（2）的右边是信度 $\rho(X)$ ，左边的第二项是非负数（因为 $\text{Var}(\cdot) \geq 0$ ），因此， $\alpha \leq \rho(X)$ ；

$$(2) \text{ 当且仅当 } \text{Var}(T_i - T_j) = 0 \quad (\forall i \neq j) \quad (3)$$

时， $\alpha = \rho(X)$ 。

条件（3）实际意义为“任意两个项目的真分数只相差一个常数”即 $T_i = T_j + C_{ij} (\forall i \neq j)$ ，其中 C_{ij} 是与第 i 第 j 项目有关但与被试无关的常数。这就是所谓的“基本 τ 等价”（essentially τ equivalent）。只有在“基本 τ 等价”的假设满足时 α 才是真正的信度。要求任意两个项目的真分数都

相等的“ τ 等价”实在是要求过高，即使加上了“基本”，允许相差一个常数，也是过于严苛的假设。若违反了“基本 τ 等价”， α 可能高估或低估了信度，对于项目数较少的测试，偏差程度可能很大。格林（Green）与杨（Yang）的模拟发现，违反基本 τ 等价，甚至可能使 α 低估真实信度达 0.2 之多^[8]，将动摇测试者对此测试可靠性的看法。

2.1.2 α 的误用

α 也有相当多的误用，误用最多的是关于 α 与同质性和内部一致性的关系，内部一致性（Internal Consistency）是指项目之间的关系，各项目得分之间有较高的相关性，则说其内部一致性高。高内部一致性是测量测试项目样本同质性或一维性的必要条件，但不是充分条件。将同质性和内部一致性看作同一概念，并将 α 看作是同质性或内部一致性的度量，这是典型的误用。达文波特（Davenport）证明了式（4）。

$$\alpha = k\bar{r} / [1 + (k-1)\bar{r}] \quad (4)$$

其中， \bar{r} 是所有项目成对相关系数的平均值，可作为内部一致性的度量^[9]。从式（4）可以看出，在固定 $\bar{r} > 0$ 的条件下，只要增加项目数 k 就可使 α 无限接近 1。可见 α 是不能作为内部一致性度量的。

2.1.3 阈值问题

将 α 看作测试的信度时，就面临如何确定阈值问题： α 多大才可以认为测试是可以接受的？科蒂纳（Cortina）认为 α 大于等于 70% 是可接受的^[10]。该阈值源于努纳利（Nunnally）的建议^[11]，并在很多论文中得到认可和应用，然而该建议只是从一般的信度考虑，而非归因于 α 本身的性质。努纳利的书在 1967 年版时指出信度在 0.5 或 0.6 就足以进行探索性研究，但在 1978 年版时增加到 0.7，人们选择引用哪一个版本，往往取决于他们的 α 高于还是低于 0.7。文图拉·莱昂（Ventura-León）指出，判断描述一个测试的质量仅靠一个 α 的值是不够，认为“世界不应当围着‘ $\alpha \geq 0.70$ ’转”^[12]。实际上在某些情况下， α 水平较低的测试可能非常有用。 α 的阈值在不同情况下应有不同，需要进行的区分越精细，信度就必须越高。

2.1.4 删除项目问题

如果测试的 α 值不够高，达不到测试人所希望的值（例如 0.7），研究人员常常会删除原测试中某些项目来提高 α 值，这样做并非无懈可击，删除项目间相关性较低的项目可能会导致 α 值的夸大，会产生样本的 α 水平比总体的 α 水平更高的“ α 膨胀”现象。雷科夫（Raykov）认为即使项目数量减少 α 似乎也在增加，但实际上量表的真实信度可能反而降低^[13]。

2.2 α 系数被广泛引用的原因

α 系数有如此多被质疑之处，但 α 系数仍被大量引用，有以下几个原因。

一是论文发表的需要。多项目测量量表是心理学和许多社会科学学科经常使用的重要方法，可以说有多项目测量量表的地方就有测量信度的要求。许多期刊和组织规定，发表关于多项目测量量表的论文必须报告信度数据。科蒂纳回顾了从 1966 年到 1990 年社会科学文献引用 α 系数的情况，“在 278 种不同期刊上被引用”，列举出引用的领域有“教育、工业、社会、临床、儿童、社区和异常心理学、实验心理学、社会学、统计学、医学、咨询、护理、经济学、政治学、犯罪学、老年学、广播学、人

类学和会计学等”^[10]。

二是教学上有关统计学的训练较少而导致 α 的误用。在心理学教育系统中, 数学和统计学训练不是重点, 这就导致只应用 α 而不顾 α 的应用条件, 以至于误用。针对 α 系数的严厉批评文章大部分只能在心理测量学的期刊(如“Psychometrika”等)上或在以方法论家和统计学家为目标读者的期刊上找到。如前述关于“基本 τ 等价”的问题, 虽然在理论上是一个非常重要的假设, 但只在测量理论圈受到关注, 在实际测量中社会科学家们很少注意这个问题, 形成“你讲你的新思想, 我用我的老方法”的尴尬局面。

三是 α 的计算较为便捷, 配套软件较多。如 SAS、SPSS 等软件可以帮助研究者们不必深究统计概念、编写计算机程序, 只要轻点鼠标就可以轻松得到所需要的参数, 便于应用。

3 α 的挑战者 ω

3.1 挑战 α 地位的新指标

70 年来虽然有许多学者指出 α 的诸多缺点, 但要用新的指标来取代 α 系数却很不容易。例如: 陈希镇的 β 系数^[14], 谢小庆的 γ 系数^[15], 丁树良和周新莲的 ξ 系数^[16]等所提出的新系数都只是改进了 α 的部分功能, 无法得到普及, 当然也无法挑战 α 了。

本特勒(Bentler)和伍德沃德(Woodward)提出了信度最大下限 glb(Greatest Lower Bound), 可以证明, $\alpha \leq \lambda_2 \leq \text{glb} \leq \rho(X)$, 因此作为信度的下界, glb 比 α 要好^[17]。雷弗尔(Revelle)和津巴格(Zinbarg)则建议用麦克唐纳的 ω ^[18]。

3.2 麦克唐纳的 ω

1970 年, 麦克唐纳(McDonald)在论及因子分析的论文的附录中给出了系数 θ ^[19], 后来在其 1999 年的论文中, 他将 θ 改称为 ω ^[5], 中文文献称为组合信度或合成信度, 大多数文献将其称为麦克唐纳的欧米伽 McDonalds ω , ω 的含义如下:

考虑单因子模型, 设 $T_i = \mu_i + \lambda_i T \quad i=1, \dots, k;$ (5)

其中 μ_i 是常数, 满足 $\sum \mu_i = 0$ (本节中, \sum 表示 i 从 1 到 k 取和, 下同), λ_i 是因子负荷, 满足 $\sum \lambda_i = 1$, τ 是唯一隐变量, 于是可将 X_i 分解为式(6)。

$$X_i = \mu_i + \lambda_i T + e_i \quad (6)$$

利用因子分析求出 λ_i , 便可依下式求出 ω :

$$\omega = \frac{\left(\sum \lambda_i\right)^2}{\left(\sum \lambda_i\right)^2 + \sum \text{Var}(e_i)} \quad (7)$$

上式中, 分子是所有项目的非标准因子载荷和的平方, 分母是它再加上项目剩余方差之和, 是信度的表达式。麦克唐纳(1999)指出, 在假定一维的情况下, 若所有 λ_i 都相等, 则(7)式中的 ω 就是 Cronbach 的 α ^[5]。这就是说, α 是 ω 的特例。注意到 ω 没有假定“基本 τ 等价”, 而当“基本 τ 等价”被满足时, 所有 λ_i 都相等, ω 就成为 α 。这就说明 ω 在信度估计中是优于 α 的选项。满足“基

本 τ 等价”的模型, ω 的性能至少能和 α 一样好, 而在违反“基本 τ 等价”时 ω 优于 α 。

因为 ω 明显优于 α , 近年来不仅得到理论工作者的大力推荐, 随着方便的计算方法的出现, 也得到了实证工作者的青睐。例如古伯 (Goodboy) 和马丁 (Martin) 希望“通信学者应该提供信度本身的计算, 即系数 ω ”以取代“学科对 α 作为信度估计的惯常依赖”^[20]。泰勒 (Taylor) 提出了过度依赖于克隆巴赫 α 的担忧, 鼓励研究人员更多地使用 ω 系数^[21]。目前邓恩 (Dunn) 等人所希望的“研究人员应该从 α 转向 ω ^[22]”的现象开始出现, 越来越多的人已接受了 ω , 只报告 α 的数据可能是“过时研究”, 低于标准, 因此提供 ω 系数势在必行。

3.3 ω 计算障碍的扫除

ω 的计算必须使用验证性因子分析 (CFA), 计算复杂, 没有专用的软件包可计算, 在推广普及上不及 α 。为改变这种状态, 许多学者在 ω 计算软件方面做了不少工作。例如, 温忠麟和叶宝娟给出了单位测验计算 ω 的 LISREL 程序^[23]; 古伯和马丁提供了如何利用 MPLUS 软件在可用数据集上计算 ω 的方法^[20]; 麦克尼什给出了用 R 软件包求得 ω 的方法^[24]等, 这些工具大大方便了 ω 的计算。

4 取代还是共存—— α 和 ω 的未来

4.1 争论还在继续

α 理论上受到严厉的批评, 又被证明只是 ω 的特例, 计算容易的优势正在消失, 在应用领域“用 ω 取代 α ”的呼声渐高, α 被 ω 取代看来只是时间的问题, 然而实际上并非如此。在前述以“克隆巴赫 α 的再审视”为专题发表在 2021 年的“Psychometrika”刊物上的 5 篇专栏论文中, 西茨玛的态度转变引起了人们的关注。西茨玛是 Tilburg 大学社会科学学院统计学教授, 2009 年西茨玛对“克隆巴赫 α 非常有限的有用性”^[3]的评论在跨学科信度研究人员中产生了巨大影响, 截至 2021 年 3 月, 该文是 Psychometrika 上发表的被引用次数最多的文章, 文中他严厉批评 α , “ α 不是内部一致性的衡量标准, 它也不是单位程度的度量”。建议用其他指标来替换 α 。12 年后, 他却为 α 的缺点做了诸多辩护, 认为“使用 CTT 还是 FA 因子分析取决于偏好; 两者在科学上是一致的……系数 α 提供了一个下限, 当测试通过近似测量一维或系数时, 该下限很有用”^[24]。西茨玛态度的变化, 说明在学术界关于是否用 ω 取代 α 还存在不同的意见。

“挺 α 派”的一个重要理由是: ω 与 α 实际相差不大。这个问题其实麦克唐纳已经觉察到了, 他指出, α 值远低于 ω 的实际例子“很难找到”^[5]。在海斯 (Hayes) 和考茨 (Coutts) 的研究中, 计算了 17 个量表的 α 和 ω 值 (用 CFA 估计), 当四舍五入到两位小数时有 11 个量表的 α 和 ω 没有差别, 而其余 6 个的差异不过 0.01 而已^[26]。经实验研究表明, α 和 ω 似乎不太可能有意义的差异。彼得森 (Peterson) 和金 (Kim) 用元分析 (Meta-analysis) 方法研究了这个问题, 从 24 种期刊 327 篇文章中获得了共 2524 对的 α 和 ω 值, 得出结论: 在相同研究条件下, 在估计信度上, ω 比 α 明确得更好, 但一般说来, 差异很小^[27], α 严重低估了真实信度的说法难以让人信服。

4.2 α 与 ω 将长期共存

虽然 α 有许多缺点,但 ω 的计算要求更高,当样本数和总体信度较低时, ω 可能显示出更多的估计失败,估计信度预期信度之间可能有更大距离,因此 ω 并不能完全取代 α 。此外, α 还提供了所有条件下的一致低估,确保 α 提供内部一致性的下限估计,而 ω 在同一方向上没有始终如一的误差(即时高时低)。Cho 指出,几乎没有经验证据表明 ω 信度比 α 更准确^[28]。过去 70 年发展起来与 α 、 ω 相关的研究显示, α 、 ω 与信度之间的关系相当复杂,无论如何,不应简单化地抛弃 α 或不愿前进固守 α ,而是应该对测量质量的评价通过多种渠道采用多种方法互相参照,体现为 α 与 ω 的共存。未来实际应用的论文应同时将 α 和 ω 的数据给出,使它们提供的信息互相补充,以示信度达到要求。 ω 与 α 将长期共存,互相补充。

参考文献

- [1] Cronbach L J. Coefficient alpha and the internal structure of tests [J]. Psychometrika, 1951, 16 (3) .
- [2] 孟庆茂,刘红云. α 系数在使用中存在的问题 [J]. 心理学探新, 2002, 22 (3) .
- [3] Sijtsma K. On the use, the misuse, and the very limited usefulness of Cronbach's alpha [J]. Psychometrika, 2009, 74 (1) .
- [4] Peters G. The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality [J]. European Health Psychologist, 2014, 16 (2) .
- [5] McDonald R P. Test theory: A unified treatment [M]. Mahwah, NJ: Lawrence Erlbaum, 1999.
- [6] 温忠麟. α 系数: 晃而不倒的信度标杆 [N]. 中国社会科学报, 2011-10-13 (12) .
- [7] TenBerge J M F, Socan G. The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality [J]. Psychometrika, 2004, 69 (4) .
- [8] Green S B, Yang Y. Commentary on coefficient alpha: A cautionary tale [J]. Psychometrika, 2009, 74 (1) .
- [9] Davenport E C, Davison M L, Liou P Y, et al. Reliability, Dimensionality, and Internal Consistency as Defined by Cronbach: Distinct Albeit Related Concepts [J]. Educational Measurement: Issues and Practice, 2015, 34 (4) .
- [10] Cortina J M. What is coefficient alpha? An examination of theory and applications [J]. Journal of Applied Psychology, 1993, 78 (1) .
- [11] Nunnally J C. Psychometric theory [M]. New York, NY: McGraw-Hill, 1978.
- [12] Ventura-León J, Pea-Calero B N. The world should not revolve around Cronbach's alpha ≥ 70 [J]. Adicciones, 2020, 33 (4) .
- [13] Raykov T. Reliability if deleted, not "alpha if deleted": Evaluation of scale reliability following component deletion [J]. British Journal of Mathematical and Statistical Psychology, 2007, 60 (2) .
- [14] 陈希镇. 如何正确使用信度估计公式 [J]. 心理学报, 1991 (1) .
- [15] 谢小庆. 信度估计得 γ 系数 [J]. 心理学报, 1998, 30 (2) .

- [16] 丁树良, 周新莲. 一种新的信度估计 [J]. 江西师范大学学报 (自然科学版), 2002, 26 (3): 222-224.
- [17] Bentler P M, Woodward J A. Inequalities among lower bounds to reliability: With applications to test construction and factor analysis [J]. Psychometrika, 1980, 45 (2).
- [18] Revelle W, Zinbarg R E. Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma [J]. Psychometrika, 2009, 74 (1).
- [19] McDonald R P. The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis [J]. British Journal of Mathematical and Statistical Psychology, 1970, 23 (1).
- [20] Goodboy A K, Martin M M. Omega over alpha for reliability estimation of unidimensional communication measures [J]. Annals of the International Communication Association, 2020, 44 (4).
- [21] Taylor J M. Coefficient Omega [J]. Journal of Nursing Education, 2021, 60 (8).
- [22] Dunn T J, Baguley T, Brunsden V. From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation [J]. British Journal of Psychology, 2014, 105 (3).
- [23] 温忠麟, 叶宝娟. 测验信度估计: 从 α 系数到内部一致性信度 [J]. 心理学报, 2011, 43 (7).
- [24] McNeish D. Thanks coefficient alpha, we'll take it from here [J]. Psychological Methods, 2018, 23 (3).
- [25] Sijtsma K, Pfadt J M. Rejoinder: The Future of Reliability [J]. Psychometrika, 2021, 86 (4).
- [26] Hayes A F, Coutts J J. Use Omega Rather than Cronbach's Alpha for Estimating Reliability, But…… [J]. Communication Methods and Measures, 2020, 14 (1).
- [27] Peterson R A, Kim Y. On the relationship between coefficient alpha and composite reliability [J]. The Journal of applied psychology, 2013, 98 (1).
- [28] Cho E. Neither Cronbach's alpha nor McDonald's omega: A commentary on Sijtsma and Pfadt [J]. Psychometrika, 2021, 86 (4).

Replace or Coexist —Cronbach's Alpha and McDonald's Omega

Gao Qinghui

Xiamen University, Xiamen

Abstract: Coefficient Cronbach's α is the most widely used for estimating reliability. Researchers have used the index extensively in the papers that need to report its reliability, but many scholars have also questioned it. The α coefficient is used as reliability and must satisfy the "essential tau-equivalence" assumption. This assumption is too strict and difficult to meet, and its violation may lead to α overestimating or underestimating the reliability. Using Cronbach's α to estimate internal consistency is inappropriate. The acceptable lower bound to the reliability of a test is often set empirically, and there is no precise standard. Researchers increase the α value by deleting items, which may also lead to a decrease in the actual reliability of the scale. Although these problems exist, α has been widely used in related research for a long time. This is due to the following reasons: many research fields involve reporting reliability coefficients, researchers have not been taught how to use α correctly for a long time; in addition, standard statistical software has the function of calculating α , which is convenient for calculation; the editors also have requirements for reporting α in the paper. McDonald pointed out that α is a particular case of MacDonal's ω , and ω becomes α when the "essential tau-equivalence" is satisfied. ω is better than α when the "essential tau-equivalence" cannot be satisfied in reliability estimation. However, the calculation of ω must use confirmatory factor analysis (CFA), which is challenging to implement in the pre-computer era. Researchers have gradually started to use ω instead of α in their research, and more and more people have accepted ω . However, whether MacDonal's ω should be used instead of Cronbach's α , there is still a heated debate in the academic community. Opinions on ω mainly focus on the fact that the actual values of ω and α are not significantly different in the calculation. That ω may show more estimation failures when the sample size and overall reliability are small. It is unreasonable to abandon α or be unwilling to move forward and stick to α . In future research, use McDonald's omega and Cronbach's alpha for reliability estimation to coexist for a long time and complement each other.

Key words: Reliability estimation; Cronbach's α ; MacDonal's ω