Original Article

Check for updates

A General Framework for Planning the Number of Items/Subjects for Evaluating Cronbach's Alpha: Integration of Hypothesis Testing and Confidence Intervals

Wei-Ming Luh [1]

[1] *Institute of Education, National Cheng Kung University, Tainan, Taiwan.*

Corresponding Author: Wei-Ming Luh, Institute of Education, National Cheng Kung University, Tainan, 701, Taiwan. Tel.: +886-6-2757575, ext. 56221. E-mail: luhwei@mail.ncku.edu.tw

Abstract

Cronbach's alpha, widely used for measuring reliability, often operates within studies with sample information, suffering insufficient sample sizes to have sufficient statistical power or precise estimation. To address this challenge and incorporate considerations of both confidence intervals and cost-effectiveness into statistical inferences, our study introduces a novel framework. This framework aims to determine the optimal configuration of measurements and subjects for Cronbach's alpha by integrating hypothesis testing and confidence intervals. We have developed two R Shiny apps capable of considering up to nine probabilities, which encompass width, validity, and/or rejection events. These apps facilitate obtaining the required number of measurements/subjects, either by minimizing overall cost for a desired probability or by maximizing probability for a predefined cost.

Keywords

Cronbach's alpha (Cronbach, 1951) stands as one of the most widely used coefficients reflecting the interrelatedness of items (Sijtsma, 2009; Sijtsma & Pfadt, 2021). Despite considerable debate surrounding its utilization (Cho & Kim, 2015; Cortina, 1993; Green & Yang, 2009; Kelley & Pornprasertmanit, 2016; McNeish, 2018; Raykov & Marcoulides, 2019; Schmitt, 1996; Sijtsma & Pfadt, 2021), the estimation and the testing of the alpha coefficient have gained critical attention in applied settings. Ensuring the reliability of a measure remains crucial for correctly interpreting the effects of experimental variables.

However, it has been evidenced that most reliabilities fall short of ideal standards in making precise, confident decisions (Charter, 2003a). Furthermore, as investigative tasks heavily rely on sample information, determining sample size in the initial stage of research design is pivotal in reducing sampling errors. While systemic studies have discussed sample size calculation, particularly in Intraclass Correlation Coefficient (ICC) studies (Donner & Eliasziw, 1987; Shieh, 2014a; 2014b; Shoukri et al., 2004; Walter et al., 1998), the discussion regarding Cronbach's alpha remains relatively scarce.

Estimating the necessary number of participants to yield meaningful results proves challenging (Charter,1999; Cocchetti,1999; Flight & Julious, 2016; Peterson & Kim, 2013). On one hand, the number of subjects might be too small to produce sufficiently precise reliability coefficients or enough statistical power for hypothesis tests (Charter, 2003b; Heo et al., 2015; Yurdugül, 2008). On the other hand, the number of measurements (items, raters) might be too large to lack cost-effectiveness (Hsu, 1994; Overall & Dalal, 1965). Note that the magnitude of the coefficient alpha is contingent upon the number of items, with a curvilinear relationship (Komorita & Graham, 1965). This aspect necessitates further investigation into subject/item size determination. Specifically, this involves considerations of constructing confidence intervals and assessing cost-effectiveness in reliability estimation.

The conventional practice of reporting alpha coefficients as point estimates impedes interpretation and replication (Terry & Kelley, 2012) since the alpha estimate is influenced by variance sources and contains unknown-direction sampling errors. Recommending the reporting confidence intervals (CIs) aims to enhance the trustworthiness of reliability (American Psychological Association [APA], 2001; Bonett & Wright, 2015; Fan & Thompson, 2001; Iacobucci & Duhachek, 2003; Kelley et al., 2003) and to convey information related to precision and reproducibility, especially in cases of very large or small sample sizes (Mendoza & Stafford, 2001). However, the mere use of CIs does not inherently enhance statistical practice (Cumming, 2014; Morey et al., 2016) without proper sample size planning (Liu, 2009). Liu (2012) also noted that current sample size planning typically aims to achieve the power of a statistical test under specified alternative hypotheses, rather than constructing precise confidence intervals. Notably, sample size influences CI width (Charter, 1999). The process of planning sample sizes to obtain CI precision has some parallels to planning for statistical power but often results in significantly different sample size requirements (Borenstein et al., 2001; Goodman & Berlin, 1994). Moreover, researchers lack well-established criteria for determining CI widths (Smithson, 2003), and may overlook the stochastic nature of interval width, i.e., a CI width is a random variable, where approximately half the time, the computed CI width exceeds the desired width in repeated sampling (Terry & Kelley, 2012), potentially underestimating required sample sizes (Liu, 2009).

In light of the various applications of integration of hypothesis testing and confidence intervals for obtaining the needed sample sizes, the present study employed the concept

of event rejection, validity, and width (Jiroutek et al., 2003), defined as follows: an event rejection (R) is said to occur if the null value of Cronbach's alpha is rejected; an event validity (V) is said to occur if Cronbach's alpha is contained between the upper and lower CI limits, and an event width (W) is said to occur if the width of the CI is no larger than the desired width w. The probabilities of the aforementioned events are then denoted as $P(R)$, $P(V)$, and $P(W)$, respectively. There are combinations of these events for various scenarios. Consider a scenario where a researcher aims to limit the width of a CI, conditional upon Cronbach's alpha falling between the lower and upper CI bounds. Determining the necessary sample size in this case ensures $P(W|V)$ achieving a probability at a desired level of $1 - \beta$. Another instance arises when, alongside rejecting the null hypothesis, there's a desire to construct a confidence interval within a specified width. Calculating the required sample size for reporting Cronbach's alpha aims to warrant that $P(W \cap R)$ can achieve a desired probability of $1 - \beta$. This integrative approach to sample size planning, addressing multiple conditional probabilities, remains infrequently explored in literature, apart from the work by Terry and Kelley (2012), who focused on $P(W) \geq 1 - \beta$ for composite reliability coefficients. Other studies have discussed various conditional probabilities (Beal, 1989; Jiroutek et al., 2003; Liu, 2012). Our study addresses a total of nine unconditional/conditional probabilities (cases), namely:

1. $P(R)$.
2. $P(R \cap V)$.
3. $P(W)$.
4. $P(W|V)$.
5. $P(W \cap V)$.
6. $P(W|R)$.
7. $P(W \cap R)$.
8. $P(W \cap R|V)$.
9. $P(W \cap R \cap V)$.

In this regard, being able to consider these probabilities can provide researchers with a thoughtful and comprehensive scenario that was not treated in much detail in the past. Also, combining hypothesis testing and a CI estimate is essential because a power-based approach can help avoid the ethical issues raised by recruiting too many/few participants (Cesana, 2013; Maxwell & Kelley, 2011).

Yet another crucial consideration within reliability studies involves balancing the cost of data acquisition against the precision/accuracy of estimates. Surprisingly, a cost-efficient design, rooted in health economics, often remains neglected (Rezagholi & Mathiassen, 2010). In practical applications, the acquisition of raters or certain measurements entails a considerable expense. When prioritizing budget constraints, obtaining substantial information at minimal cost necessitates optimizing the configuration of both the number of measurements (i.e., items, raters) and subjects (or observations) (Shoukri

et al., 2003). To address these pivotal concerns, this study proposes the development of a framework that integrates hypothesis testing and confidence intervals. The aim is to determine the optimal number of measurements/subjects, guided by two key objectives: (a) minimizing the total cost for a desired probability, and (b) maximizing the probability of interest within a predefined total cost. Additionally, to enhance accessibility for researchers, this proposed procedure has been translated into two R Shiny apps (Diedenhofen & Musch, 2016; R Development Core Team, 2020). Leveraging rapid advancements in computing technology, there is a renewed opportunity to stimulate interest in sample size planning, specifically exploring various conditional probabilities.

The subsequent sections of this study are organized as follows. The Measurement Model section delves into an elucidation of Cronbach's alpha using a measurement model. In the section, Method for Acquiring Number of Measurements and Subjects, we detail the methodology for acquiring pairs of measurements/subjects for $P(R)$, $P(V)$, and $P(W)$, respectively. The Proposed Apps section showcases the functionality of the proposed apps concerning objectives (a) or (b) with an illustrative example. Moving to the Tables and Simulations section, we present three tables and simulation results. Finally, the Discussion and Conclusions section encapsulates the study with some best-practice suggestions.

The Measurement Model

Cronbach's α describes the reliability of a sum (or average) of m measurements (test items, raters, occasions, or alternative forms). To evaluate the Cronbach's α coefficient, a model for the parallel- measurements score y_{ij} is given as:

$$y_{ij} = t_i + e_{ij}, \tag{1}$$

where t_i is the true score of subject i and e_{ij} is the error of measurement j for subject i, $i = 1, ..., n$ (subjects); $j = 1, ..., m$ (measurements). We also assume that $\{t_i\}$ are normally and identically distributed with mean 0 and variance σ_s^2, $\{e_{ij}\}$ are normally and identically distributed with mean 0 and variance σ_e^2; and $\{t_i\}$ and $\{e_{ij}\}$ are independent. That is, the random vector $(y_{i1}, ..., y_{im})$ is distributed as a multi-normal distribution with mean 0 and a covariance of $\Sigma = \sigma_s^2 \mathbf{11}' + \sigma_e^2 \mathbf{I}$. Based on Feldt (1965, 1969) and Kraemer (1981), the estimated Cronbach's coefficient can be expressed as

$$\widehat{\alpha} = 1 - \frac{MS_{M \times S}}{MS_S}, \tag{2}$$

where $MS_{M \times S}$ is the mean square for measurement by subject, MS_S is for subjects, and $\widehat{\alpha}$ estimates the population value of Cronbach's alpha α ($0 < \alpha < 1$) as

$$\alpha = \frac{m\sigma_s^2}{m\sigma_s^2 + \sigma_e^2} = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_e^2/m}. \tag{3}$$

Based on Yurdugül (2008) and Heo et al. (2015), Equations (2) and (3) can be re-expressed as

$$\widehat{\alpha} = \frac{m}{m-1}\left(1 - \frac{tr(S)}{1'S1}\right) \text{ and } \alpha = \frac{m}{m-1}\left(1 - \frac{tr(\Sigma)}{1'\Sigma 1}\right), \tag{4}$$

where S is the unbiased sample covariance matrix; $1'$ is the transpose of column vector 1 with m unit elements; Σ is the variance-covariance matrix of the population; and $tr(\Sigma)$ is the sum of the diagonal elements of the square matrix Σ. Note that under the assumption of parallel measures (two measures have identical true scores and equal error variances), we can obtain $\Sigma = \sigma_s^2 11' + \sigma_e^2 I$, where I is an identity matrix; hence, Equation (4) is identical to Equations (2) and (3). Also note that the coefficient alpha is satisfactory if the less restrictive essentially tau-equivalent assumption (i.e., unequal variances but equal covariances) holds (Sijtsma & Pfadt, 2021) in the case of approximate unidimensionality.

From Kristof (1963) and Feldt (1965), we know that $(1-\alpha)/(1-\widehat{\alpha})$ is distributed as a central F-distribution with $(n-1)$ and $(n-1)(m-1)$ degrees of freedom. Therefore, we define the test statistic

$$F = \frac{1-\widehat{\alpha}}{1-\alpha}, \tag{5}$$

which is distributed as an F-distribution with $vn = (n-1)(m-1)$ and $vd = (n-1)$ degrees of freedom. In the present study, to find the number of measurements/subjects, we applied distribution F, based on the distributional theory derived by Feldt (1965) and Feldt et al. (1987), described in the Method for Acquiring Number of Measurements and Subjects section.

Method for Acquiring Number of Measurements and Subjects

In this section, we consider events rejection, validity, and width and their corresponding $P(R)$, $P(V)$, and $P(W)$, respectively to acquire pairs of measurements/subjects by using F distribution. First, to enhance the clinical interpretation of testing Cronbach's alpha, the null hypothesis against the right-tailed alternative hypothesis is considered in the following manner:

$$H_0: \alpha \le \alpha_0 \text{ (or } 1-\alpha \ge 1-\alpha_0) \text{ versus } H_1: \alpha > \alpha_0 \text{ (or } 1-\alpha < 1-\alpha_0). \tag{6}$$

PsychOpen GOLD

Note that the hypothesis testing described here does not invoke the nil null hypothesis that the score reliability is 0 (Fan & Thompson, 2001) but instead α_0 is a golden standard or a particular criterion (Kuijpers et al., 2013; Nunnally & Bernstein, 1994). For a significance level δ, to test the null hypothesis $H_0: 1 - \alpha \geq 1 - \alpha_0$, we have

$$P(F < F_{\delta, (vn, vd)} | H_0: 1 - \alpha \geq 1 - \alpha_0) \leq \delta, \tag{7}$$

where $F = (1 - \widehat{\alpha})/(1 - \alpha_0)$ and $F_{\delta, (vn, vd)}$ is the δth quantile of distribution F with $vn = (n - 1)(m - 1)$ and $vd = (n - 1)$ degrees of freedom. Hence, based on $F < F_{\delta, (vn, vd)}$, H_0 can be rejected when

$$\widehat{\alpha} > 1 - (1 - \alpha_0)F_{\delta, (vn, vd)}, \tag{8}$$

which can be defined as an event rejection (R). For the alternative hypothesis with a specified value $\alpha_1 > \alpha_0$, the power function will coincide with $P(R)$, the probability of the event rejection, as

$$\begin{aligned} P(R) &= P(\widehat{\alpha}_1 > 1 - (1 - \alpha_0)F_{\delta, (vn, vd)} | H_1: 1 - \alpha = 1 - \alpha_1) \\ &= P(1 - \widehat{\alpha}_1 < (1 - \alpha_0)F_{\delta, (vn, vd)} | H_1: 1 - \alpha = 1 - \alpha_1) \\ &= P(F < [(1 - \alpha_0)/(1 - \alpha_1)]F_{\delta, (vn, vd)} | H_1: 1 - \alpha = 1 - \alpha_1), \end{aligned} \tag{9}$$

where $F = (1 - \widehat{\alpha}_1)/(1 - \alpha_1)$. To achieve the desired power $1 - \beta$, we must set

$$[(1 - \alpha_0)/(1 - \alpha_1)]F_{\delta, (vn, vd)} \geq F_{1 - \beta, (vn, vd)}. \tag{10}$$

Then, we can find various pairs of measurement m with its corresponding number of subjects n to satisfy Equation (10), that is, to satisfy $P(R) \geq 1 - \beta$.

Second, for $P(V)$, various numbers of measurements/subjects for constructing two-sided and one-sided CIs of coefficient α_1 with a desired probability are described. A confidence level is set to $1 - \delta$ (i.e., the probability of the event validity, $P(V) = 1 - \delta$). To form a two-sided CI, from Equation (5) and by Feldt et al. (1987), it can be shown that

$$\begin{aligned} 1 - \delta &= P(F_{\delta/2, (vn, vd)} \leq F \leq F_{1 - \delta/2, (vn, vd)}) \\ &= P\left(\frac{1 - \widehat{\alpha}_1}{F_{1 - \delta/2, (vn, vd)}} \leq 1 - \alpha_1 \leq \frac{1 - \widehat{\alpha}_1}{F_{\delta/2, (vn, vd)}}\right) \\ &= P\left(1 - \frac{1 - \widehat{\alpha}_1}{F_{\delta/2, (vn, vd)}} \leq \alpha_1 \leq 1 - \frac{1 - \widehat{\alpha}_1}{F_{1 - \delta/2, (vn, vd)}}\right), \end{aligned}$$

where $F = (1 - \widehat{\alpha}_1)/(1 - \alpha_1)$. The lower confidence limit (LCL) and the upper confidence limit (UCL) are denoted as $LCL = 1 - (1 - \widehat{\alpha}_1)/F_{\delta/2, (vn, vd)}$ and

$UCL = 1 - (1 - \widehat{\alpha}_1)/F_{1 - \delta/2, (vn, vd)}$, respectively. Hence, a $(1 - \delta)$ 100% two-sided CI is $[LCL, UCL]$, for which the width is

$$UCL - LCL = (1 - \widehat{\alpha}_1)(1/F_{\delta/2, (vn, vd)} - 1/F_{1 - \delta/2, (vn, vd)}), \tag{11}$$

an increasing function of $(1 - \widehat{\alpha}_1)$. In some contexts, there is rationale in acquiring a one-sided CI, structured as $[LCL, 1]$. It can be shown that

$$1 - \delta = P(F_{\delta, (vn, vd)} \leq F < \infty)$$

$$= P(0 \leq 1 - \alpha_1 \leq \frac{1 - \widehat{\alpha}_1}{F_{\delta, (vn, vd)}})$$

$$= P(1 - \frac{1 - \widehat{\alpha}_1}{F_{\delta, (vn, vd)}} \leq \alpha_1 \leq 1).$$

Here, $LCL = 1 - (1 - \widehat{\alpha}_1)/F_{\delta, (vn, vd)}$ and $UCL = 1$. The width of the one-sided CI is defined as

$$1 - LCL = 1 - (1 - (1 - \widehat{\alpha}_1)/F_{\delta, (vn, vd)}) = (1 - \widehat{\alpha}_1)/F_{\delta, (vn, vd)}. \tag{12}$$

From Equations (11) and (12), it is known that the width of a CI is an increasing function of $(1 - \widehat{\alpha}_1)$ given δ, m, and n. Hence, the width is a random variable.

Third, we define the event width (W) as $UCL - LCL \leq w$ for a two-sided CI, or $1 - LCL \leq (1 - \alpha_1) + w/2$ for a one-sided CI, where w is the desired width chosen as sensibly as $0 < w \leq (1 - \alpha_0)$. The probability of the event width is

$$P(W) = P((1 - \widehat{\alpha}_1)F^a_{(vn, vd)} \leq w^a)$$

$$= P(F \leq \frac{w^a}{1 - \alpha_1}/F^a_{(vn, vd)}), \tag{13}$$

where $F = (1 - \widehat{\alpha}_1)/(1 - \alpha_1)$; $w^a = w$ and $F^a_{(vn, vd)} = (1/F_{\delta/2, (vn, vd)} - 1/F_{1 - \delta/2, (vn, vd)})$ for a two-sided CI; and $w^a = (1 - \alpha_1) + w/2$ and $F^a_{(vn, vd)} = 1/F_{\delta, (vn, vd)}$ for a one-sided CI, respectively. Then, to achieve $P(W) \geq 1 - \beta$, we must set

$$[w^a/(1 - \alpha_1)]/F^a_{(vn, vd)} \geq F_{1 - \beta, (vn, vd)} \tag{14}$$

by replacing α_1 with a planning value obtained from expert opinion or prior research (Bonett, 2002). Then, pairs of (m, n) can be obtained to satisfy Equation (14). For other unconditional/conditional probabilities, the pairs can be obtained by using the proposed apps demonstrated in The Proposed Apps section.

The Proposed Apps

In the framework of integration of hypothesis testing and CIs for nine unconditional/ conditional cases, when the budget is the primary concern, we derived an optimal pair (m, n) under cost constraint. Let c_m represent the cost per measurement, c_s represent the cost per subject, and c_{ms} represent the cost per observation. The total cost can be expressed as

$$C = c_m m + c_s n + c_{ms} mn \tag{15}$$

when a pair of (m, n) is given (Eliasziw & Donner, 1987). To be ethically and economically feasible for objectives (a) and (b) outlined in the introduction, we have developed two R Shiny apps described as follows to either minimize the total cost for a desired probability or to maximize the probability of interest within a predefined total cost.

For objective (a), App (I) (see Luh, 2024a) is designed based on the section Method for Acquiring Number of Measurements and Subjects, employing an exhaustive search method. To use App (I), researchers start by selecting the specific event of interest (case), setting a significance level, a desired probability, a planning value for an alternative α_1, and determining the number of measurements up to which all outcomes will be printed. For cases related to the event (R), researchers input a null hypothesis value which should be smaller than α_1. Regarding cases linked to the event (W), users specify the desired width of CI and whether it's one- or two-sided. Additionally, if there is a cost constraint, they can specify the unit cost of measurement, subject, and observation can be specified (refer to Figure 1). Upon entering these values, clicking "Submit" executes App (I), displaying a list of measurement-subject pairs that satisfy the desired probability. Among the pairs with the minimal total cost, an optimal pair with the maximal probability for objective (a) is highlighted at the bottom of the output.

For objective (b), the probability of interest is maximal for a given cost (C). From Equation (15), we have $C - c_m m = (c_s + c_{ms} m)n$. Thus, for given measurements m, the corresponding number of subjects is $n = (C - c_m m)/(c_s + c_{ms} m)$. The optimal pair that has a maximal probability can be derived by using an exhaustive algorithm. We offer App (II) (see Luh, 2024b) for a user-friendly application. Researchers need to specify the event of interest and the corresponding parameters. Additionally, the fixed total cost (C), as well as the unit cost of measurement, subject, and observation, are required. The output presents the optimal measurement-subject pair along with its corresponding total cost and the maximum attainable probability within this total cost.

In the following, we utilized an example from Bonett (2002) to demonstrate the functionality of the proposed apps. For objective (a), to test $H_0: \alpha \leq \alpha_0 = .7$ versus $H_1: \alpha = \alpha_1 = .8$ at $\delta = .025$, with a desired probability (power) of .9 and a given number of measurements $m = 4$, App (I) indicated the required number of subjects for $P(R)$ as $n = 170$, a result close to Bonett's 173. Additionally, aiming for a desired precision, $P(W)$,

PsychOpen GOLD

Bonett (2002) set the planning value $\alpha_1 = 0.7$, the desired absolute precision of 0.2 with 95% confidence for a two-sided CI, which yielded a total of 99 subjects. However, our simulation revealed an empirical probability of .5429, roughly equal to a probability of 1/2. Using the statpsych R package, the sample size was determined as 95 by the command size.ci.cronbach(.05, .7, 4, 0.2), still falling short of the desired probability $(1-\beta)$ of .8. Contrastingly, based on App (I), the required number of subjects was 123, resulting in an empirical probability of .8004 from our 10,000 simulations.

Figure 1

A Screenshot of the Proposed R Shiny App (I)

Planning Number of Measurements (m) and Subjects (n) for Cronbach's Alpha: Integrating Hypothesis Testing and Confidence Intervals

Case, set 1 for P(R); 2 for P(W); 3 for P(R∩V); 4 for P(W|V); 5 for P(W∩V); 6 for P(W|R); 7 for P(W∩R); 8 for P(R∩R|V); 9 for P(W∩R∩V)

```
3
```

δ, significance level
```
0.05
```

$1-\beta$, desired probability
```
0.8
```

α_0, null hypothesis value
```
0.7
```

α_1, planning value
```
0.7
```

w, desired width of CI
```
0.2
```

Side(s), set 1 for one-sided; 2 for two-sided
```
2
```

c_m, unit cost of measurement
```
1
```

c_s, unit cost of subject
```
1
```

c_{ms}, unit cost of observation
```
0
```

number of measurements up to
```
25
```

Submit

Show 25 entries   Search:

m	n	Cost	Probability
2	185	187	0.802674
3	138	141	0.801408
4	123	127	0.806175
5	115	120	0.804303
6	110	116	0.801094
7	107	114	0.802061
8	105	113	0.80419
9	103	112	0.800767
10	102	112	0.80373
11	101	112	0.804072
12	100	112	0.802465
13	100	113	0.809757
14	99	113	0.805597
15	98	113	0.800427
16	98	114	0.805117
17	98	115	0.809192
18	97	115	0.802229
19	97	116	0.805471
20	97	117	0.808357
21	96	117	0.80026
22	96	118	0.802652
23	96	119	0.804819
24	96	120	0.806791
25	96	121	0.808593

Showing 1 to 24 of 24 entries   Previous 1 Next

Optimal pair (m, n) = (11, 101)
Minimal total cost = 112
Maximal probability = 0.804072

Note. For cases relating to the event (R), leave "desired width" and "sides" there as it is. For cases relating to the event (W), leave "null hypothesis value" there as it is. If costs are not the concern, plug in 0 in "c_m", "c_s", and "c_{ms}".

Subsequently, if the primary concern is the total cost, assuming the costs of obtaining a single measurement, c_m, and a single subject, c_s, are both $1, while a single observation, c_{ms}, is $0. App (I) derived the optimal pair (11, 101) with a minimal cost of $112 for

objective (a) in the case of $P(W)$. Finally, for objective (b), with a fixed cost of \$112, employing App (II) with $\delta = .05$ and $1-\beta = .8$ for a two-sided CI, we obtained the optimal number of measurements and subjects as $m = 11$ and $n = 101$, respectively, resulting in a maximal probability of .804072.

Tables and Simulations

Tables

To aid applied researchers, three tables are presented, generated by running App (I), showcasing key characteristics regarding the interrelation of the inputted parameters with the number of measurements/subjects. Table 1 exhibits the configuration of the desired width (w) and the planning value (α_1) across nine probability cases, offering significant insights into various aspects. First of all, as anticipated, while keeping other factors constant, a wider desired width alongside the larger planning value; refer to Equations (11) and (12), generally necessitates fewer measurements/subjects, excluding Cases 1 and 2. Second, the Cases 1. $P(R)$ and 2. $P(R \cap V)$, solely involving the event rejection, are contingent upon the difference value, $\alpha_1 - \alpha_0$; the larger the difference, the fewer measurements/subjects required; refer to Equation (10). These outcomes align with the pattern demonstrated in Table 2 of Jiroutek et al. (2003). Thirdly, for cases solely involving the event width, note that due to $P(W) \geq P(W \cap V)$ and $P(W|V) \geq P(W \cap V)$, the required number of measurements/subjects in Case 5 is slightly higher than or equal to that in Case 3 and Case 4. Note that the resulting numbers are similar in Cases 1 and 2, and in Cases 3, 4, and 5 because $P(V)$ is as high as .95 (i.e., $\delta = .05$). Fourthly, for cases involving both events of width and rejection, note that $P(W|R) \geq P(W \cap R)$ and $P(W \cap R|V) \geq P(W \cap R \cap V)$. Thus, Case 7 and Case 9 needed relatively more measurements/subjects than Case 6 and Case 8, respectively. Finally, among all cases, Case 9 necessitates the largest number of measurements/subjects due to the inclusion of all three events. Put simply, the higher the corresponding probability value of the case, the fewer measurements/subjects are required to achieve the desired probability $1 - \beta$. Generally, the conditional probabilities are larger or equal to those corresponding probabilities of events with intersection, leading to a slightly reduced number of required measurements/subjects. Furthermore, due to negligible differences, the following Tables 2 and 3 do not display these conditional probabilities (Cases 4, 6, and 8).

To delve deeper into the relationship between the number of measurements and subjects, Table 2 presents the required subject sizes (n) while holding the number of measurements (m) constant. Considering the test reliability and test time, the number of measurements ranged from $m = 10$ to 25. Using the proposed App(I), it is evident that with an increase in the number of measurements, the required subject sizes decrease,

albeit inconspicuously. Likewise, as more events are considered, the larger the number of
subjects is needed.

Table 1

*Optimal Number of Measurements and Subjects (m, n) for Nine Probability Cases, Given the Configuration of the
Desired Width and the True Difference*

α_1	Case	Desired Width[a]	
		0.1	0.2
0.8	1. $P(R)$	10, 83	10, 83
	2. $P(R \cap V)$	10, 89	10, 89
	3. $P(W)$	13, 165	8, 54
	4. $P(W\vert V)$	13, 163	8, 53
	5. $P(W \cap V)$	14, 167	8, 55
	6. $P(W\vert R)$	13, 162	6, 34
	7. $P(W \cap R)$	13, 165	10, 83
	8. $P(W \cap R\vert V)$	13, 163	10, 79
	9. $P(W \cap R \cap V)$	14, 167	10, 89
0.85	1. $P(R)$	6, 32	6, 32
	2. $P(R \cap V)$	6, 34	6, 34
	3. $P(W)$	11, 101	7, 35
	4. $P(W\vert V)$	11, 100	7, 34
	5. $P(W \cap V)$	11, 104	7, 37
	6. $P(W\vert R)$	11, 101	6, 28
	7. $P(W \cap R)$	11, 101	7, 35
	8. $P(W \cap R\vert V)$	11, 100	7, 34
	9. $P(W \cap R \cap V)$	11, 104	7, 37
0.9	1. $P(R)$	4, 15	4, 15
	2. $P(R \cap V)$	5, 15	5, 15
	3. $P(W)$	8, 54	5, 21
	4. $P(W\vert V)$	8, 53	5, 21
	5. $P(W \cap V)$	8, 55	5, 22
	6. $P(W\vert R)$	8, 54	5, 18
	7. $P(W \cap R)$	8, 54	5, 21
	8. $P(W \cap R\vert V)$	8, 53	5, 21
	9. $P(W \cap R \cap V)$	8, 55	5, 22

Note. Setting $\delta = .05$, $1 - \beta = .8$, $\alpha_0 = 0.7$, $c_m = \$1$, $c_s = \$1$, and $c_{ms} = \$0$.
[a]Two-sided CIs.

Table 2

The Corresponding Subjects (n) for Six Probability Cases

	Number of Measurements (m)			
Case	10	15	20	25
$P(R)$	83	80	78	77
$P(R \cap V)$	89	86	85	84
$P(W)$	169	163	160	159
$P(W \cap V)$	173	166	163	162
$P(W \cap R)$	169	163	160	159
$P(W \cap R \cap V)$	173	166	163	162

Note. Setting $\delta = .05$, $1 - \beta = .8$, $\alpha_0 = 0.7$, $\alpha_1 = 0.8$, $w = 0.1$ (two-sided CIs), $c_m = \$1$, $c_s = \$1$, and $c_{ms} = \$0$.

Table 3

Optimal Number of Measurements and Subjects (m, n) for Six Probability Cases, Given Various Costs

Case	(1)[a]	(2)[b]	(3)[c]	(4)[d]
$P(R)$	10, 83	5, 94	20, 78	7, 87
$P(R \cap V)$	10, 89	6, 97	21, 84	7, 94
$P(W)$	10, 84	5, 94	19, 79	6, 90
$P(W \cap V)$	10, 86	5, 97	19, 81	7, 90
$P(W \cap R)$	10, 84	5, 94	19, 79	6, 90
$P(W \cap R \cap V)$	10, 89	6, 97	21, 84	7, 94

Note. Setting $\delta = .05$, $1 - \beta = .8$, $\alpha_0 = 0.7$, $\alpha_1 = 0.8$, and $w = 0.15$ (two-sided CIs).
[a]$c_m = \$1$, $c_s = \$1$, $c_{ms} = \$0$. [b]$c_m = \$4$, $c_s = \$1$, $c_{ms} = \$0$. [c]$c_m = \$1$, $c_s = \$4$, $c_{ms} = \$0$. [d]$c_m = \$1$, $c_s = \$4$, $c_{ms} = \$0.1$.

To delve deeper, Table 3 highlights the trilateral relationship among the number of measurements, subjects, and costs. It delineates four conditions based on various cost scenarios. A comparison of Columns 1 and 2 reveals that an increased unit cost for a measurement (c_m) results in a decreased number of required measurements. Similarly, a higher cost for a subject (c_s) (see Column 3) leads to a reduced number of necessary subjects to achieve objective (a) at a minimal cost. Finally, when each observation incurs a cost, i.e., $c_{ms} > 0$ (see Column 4), the multiplicative effect of measurements and subjects contributes to minimizing the total cost. In other words, the total number of observations ($m \times n$) is decreased. Taking $P(W)$ as an example, the total number of observations is reduced from 1501(= 19×79) to 540 (= 6× 90). Moreover, Figure 2 shows the comparison of the resulting total costs under the condition of $c_m = \$1$, $c_s = \$4$, and $c_{ms} = \$0.1$. It can be observed that the lowest total cost is \$420 with $m = 6$, $n = 90$. The cost can

increase to \$640.4 as the number of measurements reduces to 2, and up to \$532 as the number of measurements increases to 25. The increase rates in terms of costs are 52.5% (= (640.4-420)/420) and 26.7%, respectively. From a practical point of view, there is much to gain from cost optimization.

Figure 2

Comparison of Resulting Total Costs from a Screenshot of App (I)

Planning Number of Measurements (m) and Subjects (n) for Cronbach's Alpha: Integrating Hypothesis Testing and Confidence Intervals

Case, set 1 for P(R); 2 for P(R∩V); 3 for P(W); 4 for P(W|V); 5 for P(W∩V); 6 for P(W|R); 7 for P(W∩R); 8 for P(W∩R|V); 9 for P(W∩R∩V)

3

δ, significance level

0.05

1-β, desired probability

0.8

α_0, null hypothesis value

0.7

α_1, planning value

0.8

w, desired width of CI

0.15

Side(s), set 1 for one-sided; 2 for two-sided

2

c_m, unit cost of measurement

1

c_s, unit cost of subject

4

c_{ms}, unit cost of observation

0.1

number of measurements up to

25

Submit

Show 25 entries   Search:

m	n	Cost	Probability
2	152	640.4	0.806231
3	113	488.9	0.80246
4	100	444	0.800187
5	94	428	0.80388
6	90	420	0.801976
7	88	420.6	0.807875
8	86	420.8	0.805929
9	85	425.5	0.809994
10	84	430	0.810719
11	83	434.3	0.809047
12	82	438.4	0.805552
13	81	442.3	0.800602
14	81	451.4	0.806331
15	81	460.5	0.811202
16	80	464	0.803761
17	80	473	0.807512
18	80	482	0.810804
19	79	485.1	0.801903
20	79	494	0.804573
21	79	502.9	0.806967
22	79	511.8	0.809126
23	79	520.7	0.811082
24	78	523.2	0.80089
25	78	532	0.802565

Showing 1 to 24 of 24 entries   Previous 1 Next

Optimal pair (m, n) = (6 , 90)
Minimal total cost = 420
Maximal probability = 0.801976

Simulations

In simulations, two criteria were used to validate the proposed apps—empirical probability and coverage rates. We first executed App (I) by setting $\delta = .05$, $1 - \beta = .8$, $\alpha_0 = 0.7$, $\alpha_1 = 0.8$, $w = 0.2$, $c_m = c_s = \$1$, and $c_{ms} = 0$ to obtain the optimal pair (m, n) for Case 4.$P(W|V)$ as (10, 85) and (8, 53) for a one- and two-sided CI, respectively. Then, given m, σ_s^2 and α_1, we obtained $\sigma_e^2 = m\sigma_s^2(1 - \alpha_1)/\alpha_1$ based on Equation (3). To conduct simulation experiments, we generated t_i from a normal distribution with mean 0 and variance σ_s^2 by using the R rnorm function to form $y_{ij} = t_i + e_{ij}$, $(i = 1, ..., n, j = 1, ..., m)$, for each y_{ij}. Next, for each subject i, we generated e_{ij} for $j = 1, ..., m$, from a normal distribution with mean

PsychOpen GOLD

0 and variance σ_e^2. Finally, the observed score y_{ij} was derived by adding t_i and e_{ij} meeting the additivity condition. In the subsequent simulation, we set $\sigma_s^2 = 1$ to obtain $\sigma_e^2 = 2.5$ for one-sided CIs and $\sigma_e^2 = 2$ for two-sided CIs. Two simulations were conducted with 10,000 replications each for one- and two-sided CIs to report the empirical probabilities across nine cases. The simulation outcomes reveal nearly identical empirical probabilities to the corresponding theoretical probabilities (refer to Figure 3 for two-sided CIs) and demonstrate excellent coverage rates. Detailed results are reported below.

Figure 3

The Empirical and Theoretical Probabilities for Nine Probability Cases
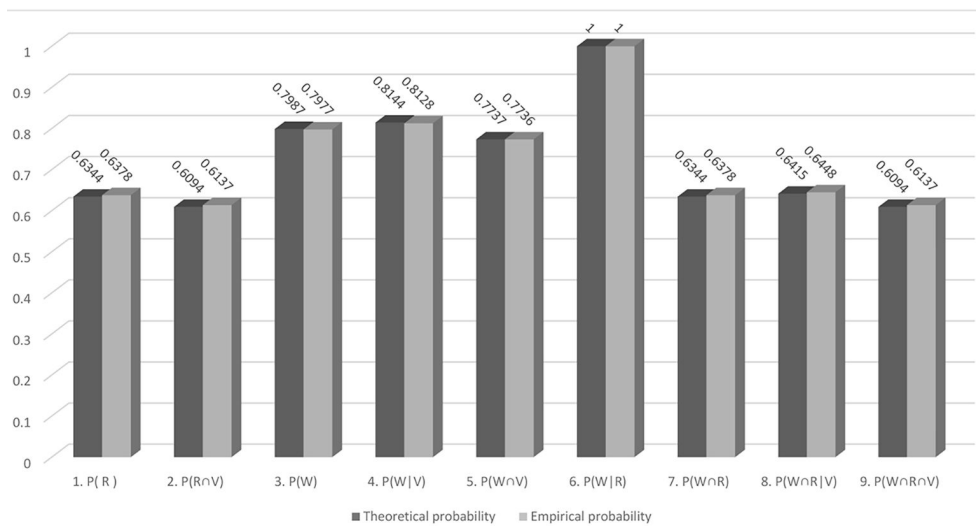
Note. Setting two-sided CIs.

For two-sided CIs, Figure 3 illustrates that Case 4 aligns with the theoretical probability as the desired level of .80. However, Cases 1, 2, 5, and 7 to 9 fall short due to insufficient measurements/subjects, whereas Case 6 exceeds the desired level due to notably larger given values of $(m, n) = (8, 53)$ compared to the intended values of $(6, 34)$, as shown in Table 1. For researchers seeking to calculate the theoretical probability, we include *R* codes in the Appendix. Moreover, we observed that the empirical distribution of estimates $\alpha_1(= .8)$ exhibited leftward skewness, featuring a mean of 0.7922 ($SD = 0.0449$), a median of 0.7978, and a coverage rate of 95.18% (i.e., empirical $P(V)$). Moreover, among the confidence intervals encompassing the planning alpha value, the average width was 0.1704, close to $w = 0.2$.

For one-sided CIs, we observed that the empirical distribution of estimates α_1 was also leftward skewness with a mean of 0.7956 ($SD = 0.0336$), a median of 0.7993, and a

coverage rate of 94.94%. Moreover, for those confidence intervals containing the planning alpha value, the average width was 0.2669, which is close to the desired width of 0.3 ($= (1 - \alpha_1) + w/2$). The empirical probabilities for the nine cases were as follows:

1. $P(R) = .8194$.
2. $P(R \cap V) = .7688$.
3. $P(W) = .8194$.
4. $P(W|V) = .8098$.
5. $P(W \cap V) = .7688$.
6. $P(W|R) = 1.0$.
7. $P(W \cap R) = .8194$.
8. $P(W \cap R|V) = .8098$.
9. $P(W \cap R \cap V) = .7688$.

As expected, Case 4 had the desired probability while Cases 2, 5, and 9 were not satisfactory because of a lack of measurements/subjects.

Discussion and Conclusion

The determination of sample size stands as a crucial and integral part of study planning, vital to achieving robust statistical power and precise estimation—a cornerstone of sound statistical practice. In addressing the need for enhanced clinical interpretation and cost-effectiveness, our study aimed to contribute to this evolving field by establishing the required number of measurements/subjects for evaluating Cronbach's alpha within a comprehensive framework encompassing hypothesis testing and confidence intervals. The introduction of our proposed apps represents a novel advancement, enabling researchers to identify optimal configurations of measurements and corresponding subjects across various events of width, validity, and rejection, crucial for achieving desired probabilities. Our empirical findings underscore the accuracy of the obtained optimal numbers, exhibiting excellent coverage rates and near-identical empirical probabilities to the desired ones. Significantly, our study illuminates the intricate interplay among the number of measurements, subjects, and costs.

It is important to note that the calculations presented here are tailored for parallel measurements and normal distributions. We acknowledge prior research (Liu et al, 2010; Olvera Astivia et al., 2020) highlighting the consequence of violating distributional assumptions and the presence of outliers. However, studies by Raykov (1997), Yuan and Bentler (2002), Osburn (2000), and our observations from simulations indicate that under specific yet verifiable conditions, coefficient alpha remains minimally affected by population deviations from scale reliability. Hence, our work stands as an initial guide for sample size planning and lays a foundation for future investigations. Subsequent studies

PsychOpen GOLD

might expand to encompass intraclass correlation coefficient (ICC) cases, facilitating adaptability across a spectrum of research designs.

Competing Interests: The author has declared that no competing interests exist.

References

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). American Psychological Association.

Beal, S. L. (1989). Sample size determination for confidence intervals on the population mean and on the difference between two population means. *Biometrics, 45*(3), 969–977. https://doi.org/10.2307/2531696

Bonett, D. G. (2002). Sample size requirements for testing and estimating coefficient alpha. *Journal of Educational and Behavioral Statistics, 27*(4), 335–340. https://doi.org/10.3102/10769986027004335

Bonett, D. G., & Wright, T. A. (2015). Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. *Journal of Organizational Behavior, 36*(1), 3–15. https://doi.org/10.1002/job.1960

Borenstein, M., Rothstein, H., & Cohen, J. (2001). *Power and precision* (p. 15). Biostat.

Cesana, B. M. (2013). Further insights in sample size calculation. *Journal of Biopharmaceutical Statistics, 23*(4), 937–939. https://doi.org/10.1080/10543406.2013.790040

Charter, R. A. (1999). Sample size requirements for precise estimates of reliability, generalizability, and validity coefficients. *Journal of Clinical and Experimental Neuropsychology, 21*(4), 559–566. https://doi.org/10.1076/jcen.21.4.559.889

Charter, R. A. (2003a). A breakdown of reliability coefficients by test type and reliability method, and the clinical implications of low reliability. *Journal of General Psychology, 130*(3), 290–304. https://doi.org/10.1080/00221300309601160

Charter, R. A. (2003b). Study samples are too small to produce sufficiently precise reliability coefficients. *Journal of General Psychology, 130*(2), 117–129. https://doi.org/10.1080/00221300309601280

Cho, E., & Kim, S. (2015). Cronbach's coefficient alpha: Well-known but poorly understood. *Organizational Research Methods, 18*(2), 207–230. https://doi.org/10.1177/1094428114555994

PsychOpen GOLD

Cocchetti, D. V. (1999). Sample size requirements for increasing the precision of reliability estimates: Problems and proposed solutions. *Journal of Clinical and Experimental Neuropsychology, 21*(4), 567–570. https://doi.org/10.1076/jcen.21.4.567.886

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334. https://doi.org/10.1007/BF02310555

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Diedenhofen, B., & Musch, J. (2016). cocron: A web interface and R package for the statistical comparison of Cronbach's alpha coefficients. *International Journal of Internet Science, 11*, 51–60.

Donner, A., & Eliasziw, M. (1987). Sample size requirements for reliability studies. *Statistics in Medicine, 6*(4), 441–448. https://doi.org/10.1002/sim.4780060404

Eliasziw, M., & Donner, A. (1987). A cost-function approach to the design of reliability studies. *Statistics in Medicine, 6*(6), 647–655. https://doi.org/10.1002/sim.4780060602

Fan, X., & Thompson, B. (2001). Confidence intervals about score reliability coefficients, please: An EPM guidelines editorial. *Educational and Psychological Measurement, 61*(4), 517–531. https://doi.org/10.1177/00131640121971365

Feldt, L. S. (1965). The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. *Psychometrika, 30*, 357–370. https://doi.org/10.1007/BF02289499

Feldt, L. S. (1969). A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests. *Psychometrika, 34*, 363–373. https://doi.org/10.1007/BF02289364

Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement, 11*(1), 93–103. https://doi.org/10.1177/014662168701100107

Flight, L., & Julious, S. (2016). Practical guide to sample size calculations: An introduction. *Pharmaceutical Statistics, 15*(1), 68–74. https://doi.org/10.1002/pst.1709

Goodman, S. N., & Berlin, J. A. (1994). The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Annals of Internal Medicine, 121*, 200–206. https://doi.org/10.7326/0003-4819-121-3-199408010-00008

Green, S. B., & Yang, Y. (2009). Commentary on coefficient alpha: A cautionary tale. *Psychometrika, 74*, 121–135. https://doi.org/10.1007/s11336-008-9098-4

Heo, M., Kim, N., & Faith, M. S. (2015). Statistical power as a function of Cronbach's alpha of instrument questionnaire items. *BMC Medical Research Methodology, 15*, Article 86. https://doi.org/10.1186/s12874-015-0070-6

Hsu, L. (1994). Unbalanced designs to maximize statistical power in psychotherapy efficacy studies. *Psychotherapy Research, 4*(2), 95–106. https://doi.org/10.1080/10503309412331333932

Iacobucci, D., & Duhachek, A. (2003). Advancing alpha: Measuring reliability with confidence. *Journal of Consumer Psychology, 13*(4), 478–487. https://doi.org/10.1207/S15327663JCP1304_14

PsychOpen GOLD

Jiroutek, M. R., Muller, K. E., Kupper, L. L., & Stewart, P. W. (2003). A new method for choosing sample size for confidence interval-based inferences. *Biometrics, 59*(3), 580–590. https://doi.org/10.1111/1541-0420.00068

Kelley, K., Maxwell, S. E., & Rausch, J. R. (2003). Obtaining power or obtaining precision: Delineating methods of sample-size planning. *Evaluation & the Health Professions, 26*(3), 258–287. https://doi.org/10.1177/0163278703255242

Kelley, K., & Pornprasertmanit, S. (2016). Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychological Methods, 21*(1), 69–92. https://doi.org/10.1037/a0040086

Komorita, S. S., & Graham, W. K. (1965). Number of scale points and the reliability of scales. *Educational and Psychological Measurement, 25*(4), 987–995. https://doi.org/10.1177/001316446502500404

Kraemer, H. C. (1981). Coping strategies in psychiatric clinical research. *Journal of Consulting and Clinical Psychology, 49*(3), 309–319. https://doi.org/10.1037/0022-006X.49.3.309

Kristof, W. (1963). The statistical theory of stepped-up reliability coefficients when a test has divided into several equivalent parts. *Psychometrika, 28*, 221–238. https://doi.org/10.1007/BF02289571

Kuijpers, R. E., van der Ark, L. A., & Croon, M. A. (2013). Testing hypotheses involving Cronbach's alpha using marginal models. *British Journal of Mathematical & Statistical Psychology, 66*(3), 503–520. https://doi.org/10.1111/bmsp.12010

Liu, X. S. (2009). Sample size and the width of the confidence interval for mean difference. *British Journal of Mathematical & Statistical Psychology, 62*(2), 201–215. https://doi.org/10.1348/000711008X276774

Liu, X. S. (2012). Implications of statistical power for confidence intervals. *British Journal of Mathematical & Statistical Psychology, 65*(3), 427–437. https://doi.org/10.1111/j.2044-8317.2011.02035.x

Liu, Y., Wu, A. D., & Zumbo, B. D. (2010). The impact of outliers on Cronbach's coefficient alpha estimate of reliability: Ordinal/rating scale item responses. *Educational and Psychological Measurement, 70*(1), 5–21. https://doi.org/10.1177/0013164409344548

Luh, W.-M. (2024a). *Planning number of measurements (m) and subjects (n) for Cronbach's alpha: integrating hypothesis testing and confidence intervals.* [Shiny App]. https://sample-size-ci.shinyapps.io/size-1-alpha-CI/

Luh, W.-M. (2024b). *Finding optimal number of measurements (m) and subjects (n) for Cronbach's alpha for fixed cost.* [Shiny App]. https://sample-size-ci.shinyapps.io/size-alpha-fixed-cost/

Maxwell, S. E., & Kelley, K. (2011). Ethics and sample size planning. In A. T. Panter (Ed.), *Handbook of ethics in quantitative methodology* (pp. 159–184). Routledge.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*(3), 412–433. https://doi.org/10.1037/met0000144

Mendoza, J. L., & Stafford, K. L. (2001). Confidence intervals, power calculation, and sample size estimation for the squared multiple correlation coefficient under the fixed and random

PsychOpen GOLD

regression models: A computer program and useful standard tables. *Educational and Psychological Measurement, 61*(4), 650–667. https://doi.org/10.1177/00131640121971419

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review, 23*, 103–123. https://doi.org/10.3758/s13423-015-0947-8

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Olvera Astivia, O. L., Kroc, E., & Zumbo, B. D. (2020). The role of item distributions on reliability estimation: The case of Cronbach's coefficient alpha. *Educational and Psychological Measurement, 80*(5), 825–846. https://doi.org/10.1177/0013164420903770

Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods, 5*(3), 343–355. https://doi.org/10.1037/1082-989X.5.3.343

Overall, J. E., & Dalal, S. N. (1965). Design of experiments to maximize power relative to cost. *Psychological Bulletin, 64*(5), 339–350. https://doi.org/10.1037/h0022527

Peterson, R. A., & Kim, Y. (2013). On the relationship between coefficient alpha and composite reliability. *Journal of Applied Psychology, 98*(1), 194–198. https://doi.org/10.1037/a0030767

R Development Core Team. (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. http://www.r-project.org.

Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence for fixed congeneric components. *Multivariate Behavioral Research, 32*(4), 329–353. https://doi.org/10.1207/s15327906mbr3204_2

Raykov, T., & Marcoulides, G. A. (2019). Thanks coefficient alpha, we still need you! *Educational and Psychological Measurement, 79*(1), 200–210. https://doi.org/10.1177/0013164417725127

Rezagholi, M., & Mathiassen, S. E. (2010). Cost-efficient design of occupational exposure assessment strategies—A review. *Annals of Occupational Hygiene, 54*(8), 858–868. https://doi.org/10.1093/annhyg/meq072

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*(4), 350–353. https://doi.org/10.1037/1040-3590.8.4.350

Shieh, G. (2014a). Optimal sample sizes for the design of reliability studies: Power consideration. *Behavior Research Methods, 46*, 772–785. https://doi.org/10.3758/s13428-013-0396-0

Shieh, G. (2014b). Sample size requirements for the design of reliability studies: Precision consideration. *Behavior Research Methods, 46*, 808–822. https://doi.org/10.3758/s13428-013-0415-1

Shoukri, M. M., Asyali, M. H., & Walter, S. D. (2003). Issues of cost and efficiency in the design of reliability studies. *Biometrics, 59*(4), 1107–1112. https://doi.org/10.1111/j.0006-341X.2003.00127.x

Shoukri, M. M., Asyali, M. H., & Donner, A. (2004). Sample size requirements for the design of reliability study: Review and new results. *Statistical Methods in Medical Research, 13*(4), 251–271. https://doi.org/10.1191/0962280204sm365ra

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

PsychOpen GOLD

Sijtsma, K., & Pfadt, J. M. (2021). Part II: on the use, the misuse, and the very limited usefulness of Cronbach's alpha: Discussing lower bounds and correlated errors. *Psychometrika, 86*, 843–860. https://doi.org/10.1007/s11336-021-09789-8

Smithson, M. (2003). *Confidence intervals.* SAGE. https://doi.org/10.4135/9781412983761

Terry, L., & Kelley, K. (2012). Sample size planning for composite reliability coefficients: Accuracy in parameter estimation via narrow confidence intervals. *British Journal of Mathematical & Statistical Psychology, 65*(3), 371–401. https://doi.org/10.1111/j.2044-8317.2011.02030.x

Walter, S. D., Eliasziw, M., & Donner, A. (1998). Sample size and optimal designs for reliability studies. *Statistics in Medicine, 17*(1), 101–110. https://doi.org/10.1002/(SICI)1097-0258(19980115)17:1<101::AID-SIM727>3.0.CO;2-E

Yuan, K. H., & Bentler, P. M. (2002). On robustness of the normal-theory based asymptotic distributions of three reliability coefficient estimates. *Psychometrika, 67*, 251–259. https://doi.org/10.1007/BF02294845

Yurdugül, H. (2008). Minimum sample size for Cronbach's coefficient alpha: A Monte-Carlo study., *H. U. Journal of Education, 35*, 397–405. http://www.efdergi.hacettepe.edu.tr/yonetim/icerik/makaleler/571-published.pdf

Appendix: R Codes for Calculating Theoretical Probabilities

```
################################################################################

    #--------------- input ----------------------------------------------------
    delta = 0.05      # significance level
    alpha0 = 0.70     # the value for the null hypothesis
    alpha1 = 0.80     # the planning value of alpha
    w = 0.2           # the desired width of CI
    m = 8             # the number of measurements
    n = 53            # the number of subjects
    side = 2          # side = 1 for one-sided CI, side = 2 for two-sided CI
    #----------------------- procedure ----------------------------------------
    vn = (n-1)*(m-1)                          # degrees of freedom for the numerator
    vd = n-1                                  # degrees of freedom for the denominator
    LF1 = ((1-alpha0)/(1-alpha1))*qf(delta, vn, vd)  # the critical value for right-tailed tests
                                             # in Equ(9)

    if (side == 1) {
        p1 = 0
        p2 = delta
        wa = (1-alpha1) + w/2
        } else {
        p1 = delta/2
        p2 = delta/2
        wa = w
          }
    RF3 = qf(1-p1, vn, vd)     # the critical value of F for left-tailed areas given 1-p1
    LF3 = qf(p2, vn, vd)       # the critical value of F for left-tailed areas given p2
    f1 = (wa/(1-alpha1))/(1/LF3 - 1/RF3)
                              # the critical value of F=(1-alpha1hat)/(1-alpha1) in Equ(13)
```

PsychOpen GOLD

```
# Nine probability cases
PV = pf(RF3, vn, vd) - pf(LF3, vn, vd)                         # for P(V)
PWF1 = pf(LF1, vn, vd)                                         # for P(R), Equ(9)
PWF2 = (pf(min(LF1, RF3), vn, vd) - pf(LF3, vn, vd))          # for P(R&V)
PWF3 = pf(f1, vn, vd)                                          # for P(W), Equ(13)
PWF4 = (pf(min(f1 ,RF3), vn, vd) - pf(LF3, vn, vd)) / (1-delta)  # for P(W|V)
PWF5 = (pf(min(f1, RF3), vn, vd) - pf(LF3, vn, vd))          # for P(W&V)
PWF6 = pf(min(f1, LF1), vn, vd) / pf(LF1, vn, vd)            # for P(W|R)
PWF7 = pf(min(f1, LF1), vn, vd)                               # for P(W&R)
PWF8 = (pf(min(f1, LF1, RF3), vn, vd) - pf(LF3, vn, vd)) / (1-delta)  # for P(W&R|V)
PWF9 = (pf(min(f1, LF1, RF3), vn, vd) - pf(LF3, vn, vd))     # for P(W&R&V)
# results
output = (round(c(PV, PWF1, PWF2, PWF3, PWF4, PWF5, PWF6, PWF7, PWF8, PWF9), 4))
names(output) = c("P(V)", "P(R)", "P(R&V)", "P(W)", "P(W|V)", "P(W&V)", "P(W|R)",
                  "P(W&R)", "P(W&R|V)", "P(W&R&V)")
output
```