https://doi.org/10.70844/jmhrp.2025.2.2.42

Journal of Medical Health Research and Psychiatry

Method Article

JMHRP-25-048

An Introduction to Simulating Cronbach's Alpha Values with The Help of SimAlpha© Algorithm

Dimitris Papadopoulos*

Data Analyst, BasisPap, Kavala, Greece

Corresponding Author: Dimitris Papadopoulos, Data analyst, BasisPap, Kavala, Greece, E-mail: info@basispap.com

Received date: 14 October, 2025, **Accepted date:** 28 October, 2025, **Published date:** 04 November, 2025

Citation: Papadopoulos D (2025) An Introduction to Simulating Cronbach's Alpha Values with The Help of SimAlpha© Algorithm. J Med Health Psychiatry 2: 2.

Abstract

A questionnaire is a structured tool used in health research to systematically collect data on perceptions, behaviors and health outcomes. It serves as a fundamental tool for recording patient-reported outcomes, evaluating treatment effectiveness and monitoring public health trends. Questionnaires can be administered in a variety of formats, including paper, digital, or interactive systems and must be carefully designed to ensure reliability, validity and minimal bias. SimAlpha algorithm was designed in 2023 from Basispap to simulate Cronbach's Alpha and uses R code to do so. It has been used professionally on several occasions where a simulated study on Cronbach's Alpha was needed. Additionally, it can provide prefixed Likert scale mean values prior to the study as well as number of participants on demand. It has been a useful tool to reduce time and money needs. In this paper a brief presentation of the algorithm is given as well as potential and weaknesses. The results of the simulation show improved behavior for large number of simulated participants and replications (N > 100,000) but the number of simulated questions does remain satisfactory even in the case of small numbers such as 10 questions. The flexibility of the algorithm needs to be improved for alpha values lower than 0.7. This means that an unexperienced user might find it difficult to perform simulations for requested alpha values lower than 0.7 although a more experienced user can easily adjust requested alpha in order to obtain the exact needed value.

Keywords: Cronbach's alpha, Simulation, SimAlpha, Application

Introduction

Questionnaires are essential tools in health research, acting as structured tools for the systematic and consistent collection of information from patients or health professionals. However, one of the most common problems, especially in clinical studies, is the number of participants - registrations. This study provides a way to facilitate health professionals who wish to conduct a study but have minimal data. It provides a way to comparative studies of their existing data with a fully controlled hypothetical sample. However, even in the case where the data is sufficient, the appearance of low reliability, even in a proven reliable questionnaire, can puzzle analysts where to find the sources of this reduction such as e.g. the range of the Likert scale. The reliability of a questionnaire is the main concern of a researcher when constructing a data collection tool assuring that this tool can assure it repeatability as well as the reassurance of valid research. Validity and reliability are quite often used in the same conceptual framework but Tavakol and Dennick emphasize that an instrument cannot be valid unless it is reliable [1]. Still, the importance of reliability is undeniable since it is an indicator of good practice in research.

Cronbach's alpha

The most widely known and used indicator of reliability is Cronbach's Alpha [2]. This index incorporates the variance of each observation and each variable under the formula.

$$a = \frac{k}{k-1}\left[\frac{\sigma_\tau^2 - \sum_{i=1}^{\kappa}\sigma_i^2}{\sigma_\tau^2}\right]$$

where σ_i^2 is the variance of each variable denoted as column of n × k dataset of n observations and k variables (n; k > 1). The incorporation of each row is done with the help of σ_τ^2 which denotes row variance. Although alpha can show negative values, it produces informative results when it lies between 0 and 1 and shows a reliable result when it is greater than 0.7.

Past simulation attempts

Although not expected, not many attempts at simulating alpha values have been made in the past. Leontitsis and Page used a simulation of 20 subjects with 6 items under 1, 000 replication, did not produce alpha values grater than 0.5 [3]. The comparison of their results to other attempts presented wide ranges of alpha values with the lowest being that of Cortina and equal to 0.20 (0.40 – 0.60) [4-7].

Methodology

In order to simulate an alpha value, one needs to know the number participants and the research tool. SimAlpha allows the use of the mean value per question – variable of the questionnaire as well. In total SimAlpha can use 4 parameters. These are, number of participants [2, ∞), number of questions [2, ∞), mean value and standard deviation per item – question. In what follows simulation

https://doi.org/10.70844/jmhrp.2025.2.2.42

Journal of Medical Health Research and Psychiatry

results are presented on a 5-point Likert scale with N participants, Q items and Rep number of replications. The purpose of this simulation is to present its capabilities and its performance on requested alpha values. This simulation was done with R programming language Version 4.2.3 in Rstudio environment.

Results

Effect of replications

The first part of the results shows the distribution and behavior of SimAlpha algorithm on variable parameters. Figure 1 shows the distribution of simulated alpha values for a hypothetical questionnaire of 10 items. All items are supposed to have mean value near 3 and standard deviation near 1. We should note that SimAlpha can also use different mean values and standard deviations per item. The use of common mean and SD values is done only for comparison reasons.

The first comparison uses variable replication numbers ranging from 10 to 1,000,000 replications. According to the results of Table 1 and Figure 1 the increment in replications does not contribute to the goal alpha value of 0.95 especially after 1,000 replications. Still, it is worth noting that increment in replication shifts the distribution to the right with lower min and increased max values. This shift is depicted in Figure 1.

Replication	Mean	Min	Max	Elapsed time
10	0.94214	0.92938	0.95228	< 1 second
100	0.94519	0.92424	0.95502	1 second
1,000	0.94643	0.92287	0.96092	2 seconds
10,000	0.94643	0.91643	0.96364	23 seconds
1,00,000	0.94639	0.91643	0.96364	231 seconds
1,000,000	0.94639	0.91121	0.96615	2223 seconds

Table 1: Effect of replication number to alpha approximation (N = 200, Q = 10, Alpha = 0.95).

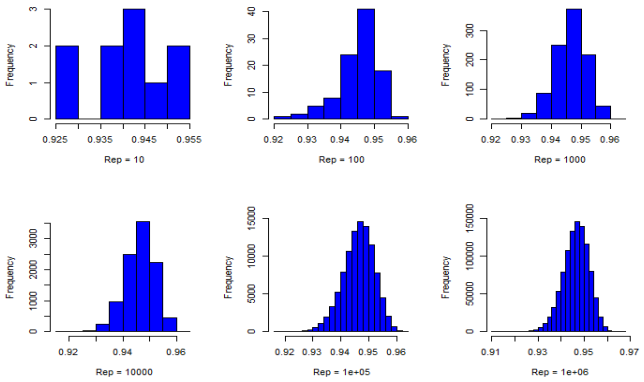

Figure 1: Distribution of alpha values for 10 to 1,000,000 replications.

Effect of participants

The next part was to investigate the behavior of SimAlpha algorithm to a variable number of participants. The range used was between 10 to 1,000,000 participants and the results for 1,000 replications are shown in Table 2 and Figure 2. According to the results, 100 participants are enough for a good approximation of the 0.95 targeted alpha value. In addition, we can see that more participants narrow down the min-max interval of simulated alpha values. Finally, figure 2 shows that 1,000 participants produce a more symmetrical distribution of alpha values compared to a larger number of participants.

Participants	Mean	Min	Max	Elapsed time
10	0.93243	0.50543	0.98792	1 second
100	0.94574	0.91291	0.96674	2 seconds
1,000	0.9469	0.93692	0.95353	4 seconds
10,000	0.94684	0.94381	0.94944	21 seconds
1,00,000	0.94686	0.94602	0.94758	260 seconds
10,00,000	0.94686	0.94665	0.94708	2286 seconds

Table 2: Effect of participant number to alpha approximation (Rep = 1,000, Q = 10, Alpha = 0.95).

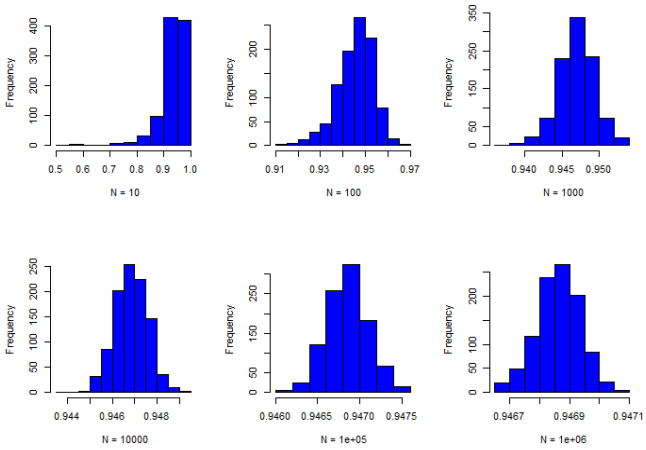

Figure 2: Distribution of alpha values for 10 to 1,000,000 participants.

Effect of number of items

The last part of the results included algorithm performance under various item number. According to the results of Table 3, there is an overestimation of alpha for questionnaires with more than 10 questions. This overestimation increases along with the increment of the number of questions but its distribution does not show improvement, that is a more symmetrical behavior as shown in Figure 3.

Questions	Mean	Min	Max	Elapsed time
10	0.94643	0.92287	0.96092	2 seconds
20	0.97253	0.96211	0.98012	4 seconds
30	0.98148	0.97348	0.98639	5 seconds
40	0.98598	0.97964	0.98979	7 seconds
50	0.98881	0.98441	0.99153	8 seconds
100	0.99439	0.99223	0.99565	2223 seconds

Table 3: Effect of participant number to alpha approximation (Rep = 1,000, N = 200, Alpha = 0.95).

https://doi.org/10.70844/jmhrp.2025.2.2.42

Journal of Medical Health Research and Psychiatry

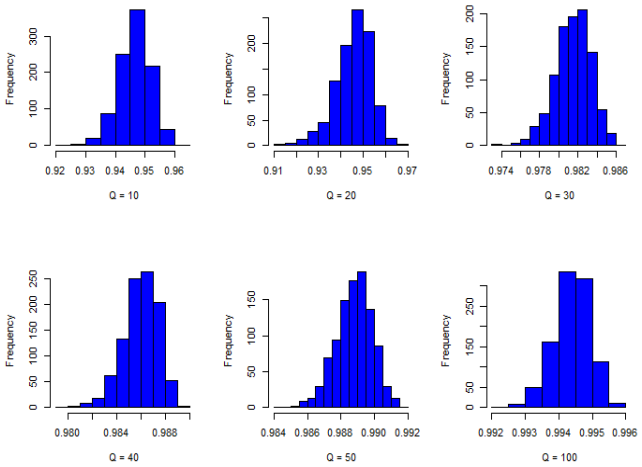

Figure 3: Distribution of alpha values for 10 to 100 items.

Benchmarking

The last step was to check deviations from the requested alpha value. Figure 4 shows mean values of calculated alpha values along with upper and lower limits expressed with the maximum and minimum values. This simulation used 0.001 to 0.999 alpha value approximation with 0.001 step, under 200 hypothetical participants, on a 10-time research tool.

According to Figure 4 there is a deviance which varies from 0.1 to 0.5 for requested alpha values from 0.1 to 0.9. The largest deviation (see also Figure 5) is shown in 0.2 to 0.4 requested alpha value. In the case of the lower critical alpha value, that is $\alpha = 0.7$ the mean deviation is near 0.25. These results clearly show the need for improvement of the target alpha value. Although a rule of thumbs can be and is, used to approximate specific alpha values even up to the third decimal place, this improvement is necessary in order to present a more user-friendly algorithm, that is an algorithm that anyone can use regardless of previous background in computing. In addition, since this is only an instance of the possible parameters can enter the simulation, more parameters should be used to compare the algorithms behavior.

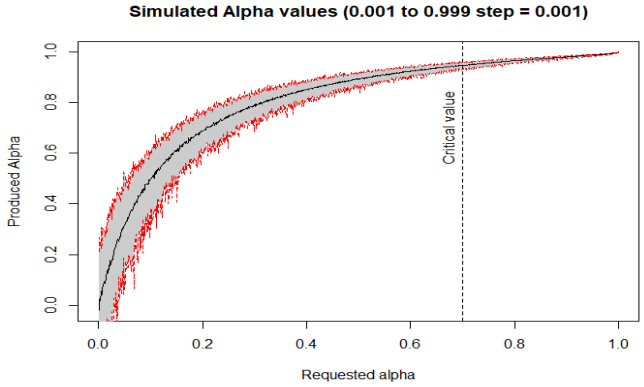

Figure 4: Simulated Alpha values under N = 200, Q = 10, Rep = 100, Step = 0.001.

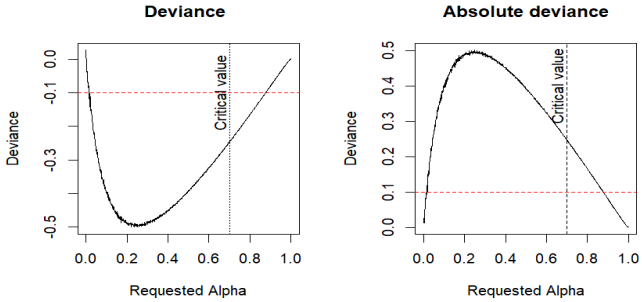

Figure 5: Deviation and absolute deviation from requested Alpha values under N = 200, Q = 10, Rep = 100, Step = 0.001.

Conclusion

SimAlpha algorithm can cope quite well with the requested task, that is to simulate a complete study with a questionnaire under a specific Alpha value. Still, it needs improvement on the requested alpha value for an inexperienced user or a user who is looking for an absolute Alpha value without deviation on the second or third decimal number. For a more lenient or a more experienced user it can be a helpful asset for pilot studies or any kind of simulation study of a new or well-established research tool. Computation time is low to tolerable for studies that want to involve a large number of hypothetical subjects. SimAlpha can support the usual parameters of a hypothetical study i.e. N = 200 participants, Q = 20 items on 1,000 replications with small deviations from the requested alpha. The time needed on these parameters is less than 4 seconds. Therefore, SimAlpha algorithm is expected to aim academics on their simulated study attempts, with the expectation of future improvement in the future on Alpha accuracy to lie below the third decimal place.

Competing Interests

The authors report no conflicts of interest in this work.

References

1. Tavakol M, Dennick R (2011) Making sense of Cronbach's alpha. Int J Med Educ 2: 53-55. [Crossref] [Googlescholar]
2. Cronbach LJ (1951) Coefficient Alpha and the Internal Structure of Tests. Psychometrika 16: 297-334. [Crossref] [Googlescholar]
3. Leontitsis A, Pagge J (2007) A simulation approach on Cronbach's alpha statistical significance. Math Comput Simul 73: 336-340. [Crossref] [Googlescholar]
4. Hakstian AR, Whalen TE (1976) A K-Sample Significance Test for Independent Alpha Coefficients. Psychometrika 41: 219-231. [Crossref] [Googlescholar]
5. Cortina JM (1993) What is coefficient alpha? An examination of theory and applications. J Appl Psychol 78: 98-104. [Crossref] [Googlescholar]
6. Charter RA (2000) Confidence Interval Formulas for Split-Half Reliability Coefficients. Psychol Rep 86: 1168-1170. [Crossref] [Googlescholar]
7. Duhachek A, Iacobucci D (2004) Alpha's Standard Error (ASE): An Accurate and Precise Confidence Interval Estimate. J Appl Psychol 89: 792-808. [Crossref] [Googlescholar]