

**Εισαγωγή στην προσομοίωση των τιμών άλφα του Cronbach με τη βοήθεια του αλγορίθμου simAlpha©**

Παπαδόπουλος Δημήτρης

Αναλυτής Δεδομένων, BasisPap, Καβάλα, Ελλάδα

DOI: 10.5281/zenodo.18164115

ΠΕΡΙΛΗΨΗ

Ένα ερωτηματολόγιο είναι ένα δομημένο εργαλείο που χρησιμοποιείται στην έρευνα υγείας για τη συστηματική συλλογή δεδομένων σχετικά με τις αντιλήψεις, τις συμπεριφορές και τα αποτελέσματα της υγείας. Χρησιμοποιείται ως θεμελιώδες εργαλείο για την καταγραφή των αποτελεσμάτων που αναφέρουν οι ασθενείς, την αξιολόγηση της αποτελεσματικότητας της θεραπείας και την παρακολούθηση των τάσεων της δημόσιας υγείας. Τα ερωτηματολόγια μπορούν να χορηγηθούν σε διάφορες μορφές, συμπεριλαμβανομένων των έντυπων, ψηφιακών ή διαδραστικών συστημάτων, και πρέπει να σχεδιάζονται προσεκτικά ώστε να διασφαλίζεται η αξιοπιστία, η εγκυρότητα και η ελάχιστη μεροληψία. Ο αλγόριθμος simAlpha σχεδιάστηκε το 2023 από την Basispap για να προσομοιώσει το Alpha του Cronbach και χρησιμοποιεί κώδικα R για την υλοποίησή του. Έχει χρησιμοποιηθεί επαγγελματικά σε αρκετές περιπτώσεις όπου χρειαζόταν μια προσομοιωμένη μελέτη για το άλφα του Cronbach. Επιπλέον, μπορεί να παρέχει συγκεκριμένες μέσες τιμές της κλίμακας Likert πριν από τη μελέτη, καθώς και τον αριθμό των συμμετεχόντων κατά παραγγελία. Αποτελεί ένα χρήσιμο εργαλείο για τη μείωση του χρόνου συλλογής συγκριτικών δεδομένων π.χ. σε μια πιλοτική έρευνα. Ακολουθεί μια σύντομη παρουσίαση του αλγορίθμου καθώς και των δυνατοτήτων και των αδυναμιών του. Τα αποτελέσματα της προσομοίωσης δείχνουν βελτιωμένη συμπεριφορά για μεγάλο αριθμό προσομοιωμένων συμμετεχόντων και επαναλήψεων ($N > 100.000$), αλλά ο αριθμός των προσομοιωμένων ερωτήσεων παραμένει ικανοποιητικός ακόμη και στην περίπτωση μικρών αριθμών, όπως π.χ. 10 ερωτήσεις. Η ευελιξία του αλγορίθμου πρέπει να βελτιωθεί για τιμές άλφα χαμηλότερες από 0,7. Αυτό σημαίνει ότι ένας άπειρος χρήστης μπορεί να δυσκολευτεί να εκτελέσει προσομοιώσεις για ζητούμενες τιμές άλφα χαμηλότερες από 0,7, αν και ένας πιο έμπειρος χρήστης μπορεί εύκολα να προσαρμόσει το ζητούμενο άλφα για να λάβει την ακριβή απαιτούμενη τιμή.

Λέξεις Κλειδιά: Cronbach's Alpha, προσομοίωση, simAlpha.**Υπεύθυνος αλληλογραφίας:** Παπαδόπουλος Δημήτρης, E-mail: info@basispap.com**Introduction to simulating Cronbach's alpha values using the simAlpha© algorithm**

Papadopoulos Dimitrios

Data Analyst, Kavala, Greece

DOI:10.5281/zenodo.18164115

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. The simAlpha algorithm was designed in 2023 by Basispap to simulate Cronbach's Alpha and uses R code for its implementation. It has been used professionally in several cases where a simulated study of Cronbach's alpha was needed. In addition, it can provide specific mean values of the Likert scale before the study, as well as the number of participants on demand. It is a useful tool for reducing the time of collecting comparative data e.g. in a pilot study. A brief presentation of the algorithm as well as its strengths and weaknesses is presented. The simulation results show improved behavior for a large number of simulated participants and repetitions ($N > 100,000$), but the number of simulated questions remains satisfactory even in the case of small numbers, such as e.g. 10 questions. The flexibility of the algorithm needs to be improved for alpha values lower than 0.7. This means that an inexperienced user may have difficulty performing simulations for desired alpha values lower than 0.7, although a more experienced user can easily adjust the desired alpha to obtain the exact required value.

Keywords: Cronbach's Alpha, simulation, simAlpha.

Corresponding Author: Papadopoulos Dimitrios, E-mail: info@basispap.com

ΕΙΣΑΓΩΓΗ

Τα ερωτηματολόγια είναι απαραίτητα εργαλεία στην έρευνα για την υγεία, λειτουργώντας ως δομημένα εργαλεία για τη συστηματική και συνεπή συλλογή πληροφοριών από ασθενείς ή επαγγελματίες υγείας. Ένα όμως από τα πιο συχνά προβλήματα, ιδιαίτερα σε κλινικές μελέτες, αποτελεί το πλήθος των συμμετεχόντων – εγγραφών. Σε αυτή την μελέτη παρέχεται ένας τρόπος διευκόλυνσης προς τους επαγγελματίες υγείας που επιθυμούν να κάνουν μια έρευνα έχοντας όμως ελάχιστα δεδομένα. Παρέχει τον τρόπο συγκριτικών μελετών των υπάρχοντων δεδομένων τους με ένα πλήρως ελεγχόμενο υποθετικό δείγμα. Ακόμη όμως και στην περίπτωση που τα δεδομένα είναι επαρκή η εμφάνιση χαμηλής αξιοπιστίας, ακόμη και σε ένα αποδεδειγμένα αξιόπιστο ερωτηματολόγιο, μπορεί να προβληματίσει τους αναλυτές όπου θα πρέπει να βρει τις πηγές αυτής της μείωσης όπως πχ. το εύρος της κλίμακας Likert.

Η αξιοπιστία ενός ερωτηματολογίου είναι το κύριο μέλημα ενός ερευνητή κατά την κατασκευή ενός εργαλείου συλλογής δεδομένων, διασφαλίζοντας ότι αυτό το εργαλείο μπορεί να εξασφαλίσει την

επαναληψιμότητα καθώς και τη διασφάλιση της έγκυρης έρευνας. Η εγκυρότητα και η αξιοπιστία χρησιμοποιούνται αρκετά συχνά στο ίδιο εννοιολογικό πλαίσιο, αλλά οι Tavakol and Dennick¹ τονίζουν ότι ένα μέσο δεν μπορεί να είναι έγκυρο αν δεν είναι αξιόπιστο. Έτσι, η σημασία της αξιοπιστίας είναι αναμφισβήτητη, καθώς αποτελεί δείκτη ορθής πρακτικής στην έρευνα.

Το άλφα του Cronbach

Ο πιο ευρέως γνωστός και χρησιμοποιούμενος δείκτης αξιοπιστίας είναι ο Alpha του Cronbach.² Αυτός ο δείκτης ενσωματώνει τη διακύμανση κάθε παρατήρησης και κάθε μεταβλητής κάτω από τον τύπο

$$\alpha = \frac{k}{k-1} \left[\frac{\sigma_t^2 - \sum_{i=1}^k \sigma_i^2}{\sigma_t^2} \right] \quad (1)$$

όπου σ_i^2 είναι η διακύμανση κάθε μεταβλητής που συμβολίζεται ως στήλη συνόλου δεδομένων n παρατηρήσεων και k μεταβλητών ($n ; k > 1$). Η ενσωμάτωση κάθε σειράς γίνεται με τη βοήθεια της οποίας υποδηλώνει τη διακύμανση της σειράς. Αν και το άλφα μπορεί να δείξει αρνητικές τιμές, παράγει ενημερωτικά αποτελέσματα όταν βρίσκεται μεταξύ 0 και 1 και δείχνει ένα

αξιόπιστο αποτέλεσμα όταν είναι μεγαλύτερο από $0,7 \cdot n \times k \sigma_{\tau}^2$

Προηγούμενες προσπάθειες προσομοίωσης

Αν και μη αναμενόμενο, δεν έχουν γίνει πολλές προσπάθειες προσομοίωσης τιμών άλφα στο παρελθόν. Οι Lentitsis & Page³ χρησιμοποίησαν μια προσομοίωση 20 ατόμων με 6 ερωτήματα σε 1.000 επαναλήψεις, η οποία όμως δεν παρήγαγε τιμές άλφα μεγαλύτερες του 0,5. Η σύγκριση των αποτελεσμάτων τους με άλλες προσπάθειες⁴⁻⁷ παρουσίασε μεγάλα εύρη τιμών άλφα με τη χαμηλότερη να είναι αυτή της Cortina⁸ και ίση με 0,20 [0,40 – 0,60].

ΜΕΘΟΔΟΛΟΓΙΑ

Για να προσομοιωθεί μια τιμή άλφα, πρέπει κάποιος να γνωρίζει τον αριθμό των συμμετεχόντων και το εργαλείο έρευνας. Το simAlpha επιτρέπει τη χρήση της μέσης τιμής ανά ερώτηση – μεταβλητή του ερωτηματολογίου. Συνολικά το simAlpha μπορεί να χρησιμοποιήσει 4 παραμέτρους. Αυτές είναι, ο αριθμός συμμετεχόντων [2, ∞), ο αριθμός ερωτήσεων [2, ∞), η μέση τιμή και τυπική απόκλιση ανά θέμα – ερώτηση. Στη συνέχεια, τα αποτελέσματα της προσομοίωσης παρουσιάζονται σε κλίμακα Likert 5 (ή και περισσότερων) σημείων με N συμμετέχοντες, Q ερωτήσεις και Rep αριθμό επαναλήψεων. Σκοπός αυτής της

προσομοίωσης είναι να παρουσιάσει τις δυνατότητές του και την απόδοσή του σε ζητούμενες τιμές άλφα. Η προσομοίωση αυτή έγινε με τη γλώσσα προγραμματισμού R Version 4.2.3 σε περιβάλλον Rstudio.

ΑΠΟΤΕΛΕΣΜΑΤΑ

Επίδραση των επαναλήψεων

Το πρώτο μέρος των αποτελεσμάτων δείχνει την κατανομή και τη συμπεριφορά του αλγορίθμου simAlpha σε μεταβλητές παραμέτρους. Το Σχήμα 1 δείχνει την κατανομή των προσομοιωμένων τιμών άλφα για ένα υποθετικό ερωτηματολόγιο 10 ερωτήσεων. Όλα τα στοιχεία υποτίθεται ότι έχουν μέση τιμή κοντά στο 3 και τυπική απόκλιση κοντά στο 1. Θα πρέπει να σημειώσουμε ότι το simAlpha μπορεί επίσης να χρησιμοποιήσει διαφορετικές μέσες τιμές και τυπικές αποκλίσεις ανά περίπτωση. Η χρήση κοινών μέσων τιμών και τιμών SD γίνεται μόνο για λόγους σύγκρισης.

Η πρώτη σύγκριση χρησιμοποιεί μεταβλητούς αριθμούς αναπαραγωγής που κυμαίνονται από 10 έως 1.000.000 επαναλήψεις. Σύμφωνα με τα αποτελέσματα του πίνακα 1 και του σχήματος 1, η αύξηση των επαναλήψεων δεν συμβάλλει στην τιμή του στόχου άλφα του 0,95, ειδικά μετά από 1.000 επαναλήψεις. Ωστόσο, αξίζει να σημειωθεί ότι η αύξηση του αριθμού των επαναλήψεων μετατοπίζει την κατανομή προς τα δεξιά με χαμηλότερες ελάχιστες και

αυξημένες μέγιστες τιμές. Αυτή η μετατόπιση απεικονίζεται στο σχήμα 1.

Επίδραση των συμμετεχόντων

Το επόμενο μέρος ήταν η διερεύνηση της συμπεριφοράς του αλγορίθμου simAlpha σε μεταβλητό αριθμό συμμετεχόντων. Το εύρος που χρησιμοποιήθηκε ήταν μεταξύ 10 και 1.000.000 συμμετεχόντων και τα αποτελέσματα για 1.000 επαναλήψεις φαίνονται στον πίνακα 2 και στο σχήμα 2. Σύμφωνα με τα αποτελέσματα, 100 συμμετέχοντες είναι αρκετοί για μια καλή προσέγγιση της στοχευμένης τιμής άλφα 0,95. Επιπλέον, μπορούμε να δούμε ότι περισσότεροι συμμετέχοντες περιορίζουν το διάστημα ελάχιστου-μέγιστου των προσομοιωμένων τιμών άλφα. Τέλος, το σχήμα 2 δείχνει ότι 1.000 συμμετέχοντες παράγουν μια πιο συμμετρική κατανομή των τιμών άλφα σε σύγκριση με μεγαλύτερο αριθμό συμμετεχόντων.

Επίδραση του αριθμού των ερωτήσεων

Το τελευταίο μέρος των αποτελεσμάτων περιελάμβανε την απόδοση του αλγορίθμου κάτω από διάφορους αριθμούς στοιχείων. Σύμφωνα με τα αποτελέσματα του πίνακα 3, υπάρχει υπερεκτίμηση του άλφα για ερωτηματολόγια με περισσότερες από 10 ερωτήσεις. Αυτή η υπερεκτίμηση αυξάνεται μαζί με την αύξηση του αριθμού των ερωτήσεων, αλλά η κατανομή της δεν παρουσιάζει βελτίωση, δηλαδή μια πιο

συμμετρική συμπεριφορά όπως φαίνεται στο σχήμα 3.

Συγκριτική αξιολόγηση

Το τελευταίο βήμα ήταν ο έλεγχος των αποκλίσεων από την ζητούμενη τιμή άλφα. Στο σχήμα 4 παρουσιάζονται οι μέσες τιμές των υπολογισμένων τιμών άλφα μαζί με τα ανώτερα και κατώτερα όρια που εκφράζονται με τις μέγιστες και ελάχιστες τιμές. Αυτή η προσομοίωση χρησιμοποίησε προσέγγιση τιμής άλφα 0,001 έως 0,999 με βήμα 0,001, κάτω από 200 υποθετικούς συμμετέχοντες, σε ένα ερευνητικό εργαλείο 10 φορές.

Σύμφωνα με το σχήμα 4 υπάρχει μια απόκλιση που κυμαίνεται από 0,1 έως 0,5 για ζητούμενες τιμές άλφα από 0,1 έως 0,9. Η μεγαλύτερη απόκλιση (βλ. επίσης σχήμα 5) φαίνεται σε 0,2 έως 0,4 ζητούμενη τιμή άλφα. Στην περίπτωση της χαμηλότερης κρίσιμης τιμής άλφα, δηλαδή $\alpha = 0,7$, η μέση απόκλιση είναι κοντά στο 0,25. Αυτά τα αποτελέσματα δείχνουν ξεκάθαρα την ανάγκη βελτίωσης της τιμής-στόχου άλφα. Αν και ένας εμπειρικός κανόνας μπορεί να χρησιμοποιηθεί και χρησιμοποιείται για την προσέγγιση συγκεκριμένων τιμών άλφα ακόμη και μέχρι το τρίτο δεκαδικό ψηφίο, αυτή η βελτίωση είναι απαραίτητη προκειμένου να παρουσιαστεί ένας πιο φιλικός προς τον χρήστη αλγόριθμος, δηλαδή ένας αλγόριθμος που μπορεί να χρησιμοποιήσει ο καθένας ανεξάρτητα από το προηγούμενο υπόβαθρο

στην πληροφορική. Επιπλέον, δεδομένου ότι αυτό είναι μόνο ένα στιγμιότυπο των πιθανών παραμέτρων που μπορούν να εισέλθουν στην προσομοίωση, θα πρέπει να χρησιμοποιηθούν περισσότερες παράμετροι για τη σύγκριση της συμπεριφοράς των αλγορίθμων.

Περιορισμοί

Αν και ο αλγόριθμός καλύπτει τις προσδοκίες ενός ερευνητή που θέλει να ετοιμάσει ένα προσομοιωμένο δείγμα, χρειάζεται βελτίωση της ζητούμενης τιμής άλφα όταν η χρήση του γίνεται από έναν άπειρο χρήστη ή έναν χρήστη που αναζητά μια απόλυτη τιμή άλφα χωρίς απόκλιση στον δεύτερο ή τρίτο δεκαδικό ψηφίο. Για έναν πιο επιεική ή πιο έμπειρο χρήστη, μπορεί να είναι ένα χρήσιμο πλεονέκτημα για πιλοτικές μελέτες ή κάθε είδους μελέτη προσομοίωσης ενός νέου ή καθιερωμένου ερευνητικού εργαλείου.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Ο αλγόριθμος simAlpha μπορεί να ανταπεξέλθει αρκετά καλά στη συγκεκριμένη περίπτωση, δηλαδή να προσομοιώσει μια πλήρη μελέτη με ένα ερωτηματολόγιο κάτω από μια συγκεκριμένη τιμή Alpha. Ωστόσο, χρειάζεται βελτίωση της ζητούμενης τιμής άλφα για έναν άπειρο χρήστη ή έναν χρήστη που αναζητά μια απόλυτη τιμή άλφα χωρίς απόκλιση στον δεύτερο ή τρίτο δεκαδικό ψηφίο. Για έναν πιο επιεική ή πιο έμπειρο

χρήστη, μπορεί να είναι ένα χρήσιμο πλεονέκτημα για πιλοτικές μελέτες ή κάθε είδους μελέτη προσομοίωσης ενός νέου ή καθιερωμένου ερευνητικού εργαλείου. Ο χρόνος υπολογισμού είναι χαμηλός έως ανεκτός για μελέτες που θέλουν να περιλαμβάνουν μεγάλο αριθμό υποθετικών θεμάτων. Το simAlpha μπορεί να υποστηρίξει τις συνήθεις παραμέτρους μιας υποθετικής μελέτης, δηλαδή $N = 200$ συμμετέχοντες, $Q = 20$ στοιχεία σε 1.000 επαναλήψεις με μικρές αποκλίσεις από το ζητούμενο άλφα. Ο χρόνος που απαιτείται σε αυτές τις παραμέτρους είναι μικρότερος από 4 δευτερόλεπτα. Ως εκ τούτου, ο αλγόριθμος simAlpha αναμένεται να στοχεύσει τους ακαδημαϊκούς στις προσομοιωμένες προσπάθειες μελέτης τους, με την προσδοκία μελλοντικής βελτίωσης στο μέλλον της ακρίβειας Alpha να βρίσκεται κάτω από το τρίτο δεκαδικό ψηφίο.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Tavakol M, Dennick R. Making sense of Cronbach's alpha. International Journal of Medical Education. 2011 Jun 27;2:53–5.
2. Cronbach LJ. Coefficient Alpha and the Internal Structure of Tests. Psychometrika. 1951 Sep 1;16(3):297–334.
3. Leontitsis A, Pagge J. A simulation approach on Cronbach's alpha statistical significance. Mathematics and Computers in Simulation. 2007 Jan;73(5):336–40.

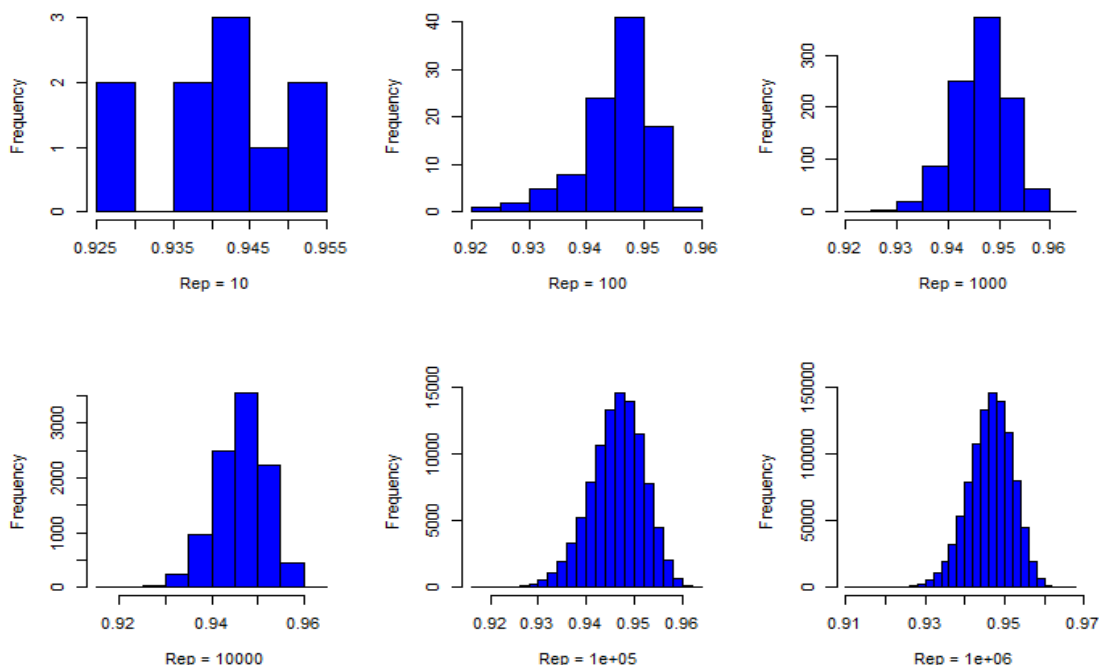
4. Hakstian AR, Whalen TE. A K-Sample Significance Test for Independent Alpha Coefficients. Psychometrika. 1976 Jun 1;41(2):219-31.
5. Duhachek A, Iacobucci D. Alpha's Standard Error (ASE): An Accurate and Precise Confidence Interval Estimate. Journal of Applied Psychology. 2004;89(5):792-808.
6. Cortina JM. What is coefficient alpha? An examination of theory and applications. Journal of Applied Psychology. 1993 Feb;78(1):98-104.
7. Charter RA. Confidence Interval Formulas for Split-Half Reliability Coefficients. Psychological Reports. 2000 Jun 1;86(3_suppl):1168-70.
8. Cortina JM. What is coefficient alpha? An examination of theory and applications. Journal of Applied Psychology. 1993 Feb;78(1):98-104.

ΠΑΡΑΡΤΗΜΑ

ΠΙΝΑΚΑΣ 1. Επίδραση του αριθμού επαναλήψεων στην προσέγγιση το άλφα ($N = 200$, $Q = 10$, $\text{Alpha} = 0,95$).

Επαναλήψεις	MT	Ελάχιστη τιμή	Μέγιστη τιμή	Χρόνος υλοποίησης
10	0.9421422	0.9293807	0.9522801	< 1 δευτερόλεπτο
100	0.9451850	0.9242415	0.9550180	1 δευτερόλεπτο
1,000	0.9464342	0.9228699	0.9609187	2 δευτερόλεπτα
10,000	0.9464284	0.9164307	0.9626803	23 δευτερόλεπτα
100,000	0.9463887	0.9164307	0.9636376	231 δευτερόλεπτα
1,000,000	0.9463948	0.9112057	0.9661475	2223 δευτερόλεπτα

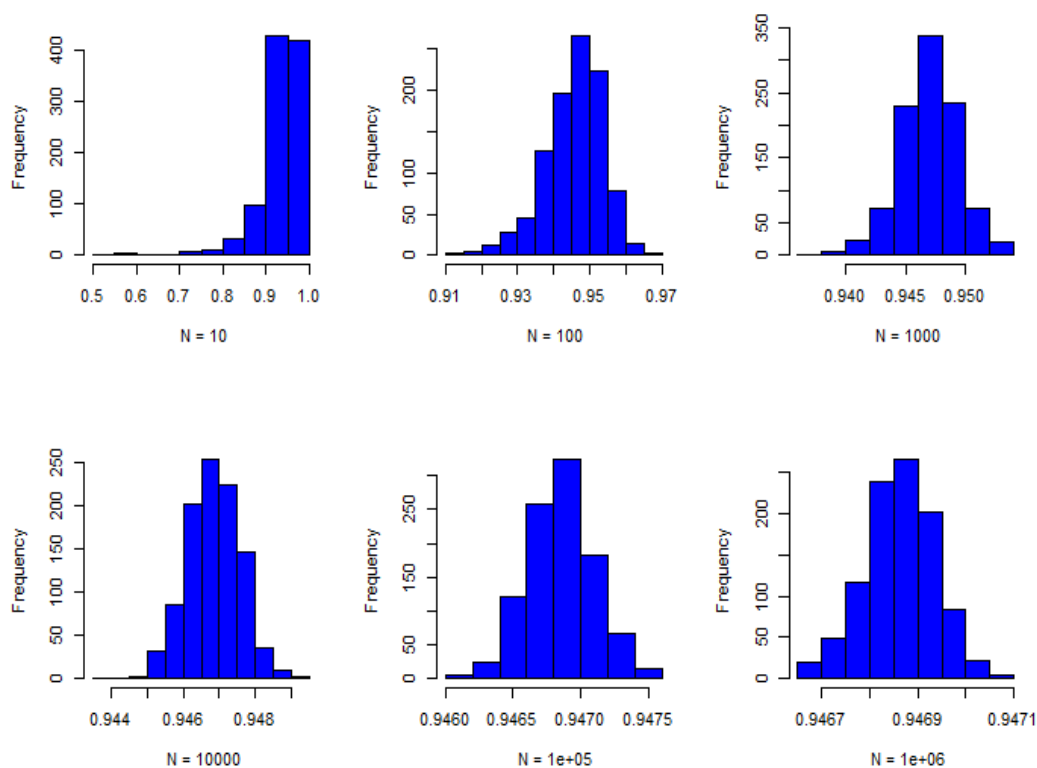
ΣΧΗΜΑ 1. Κατανομή τιμών άλφα για 10 έως 1.000.000 επαναλήψεις.



ΠΙΝΑΚΑΣ 2. Επίδραση του αριθμού των συμμετεχόντων στην προσέγγιση του άλφα (Rep = 1.000, Q = 10, Alpha = 0,95).

Πλήθος συμμετεχόντων	MT	Ελάχιστη τιμή	Μέγιστη τιμή	Χρόνος υλοποίησης
10	0.9324295	0.5054251	0.9879183	1 δευτερόλεπτο
100	0.9457435	0.9129090	0.9667423	2 δευτερόλεπτα
1,000	0.9468957	0.9369180	0.9535305	4 δευτερόλεπτα
10,000	0.9468373	0.9438134	0.9494351	21 δευτερόλεπτα
100,000	0.9468561	0.9460229	0.9475808	260 δευτερόλεπτα
1,000,000	0.9468629	0.9466537	0.9470779	2286 δευτερόλεπτα

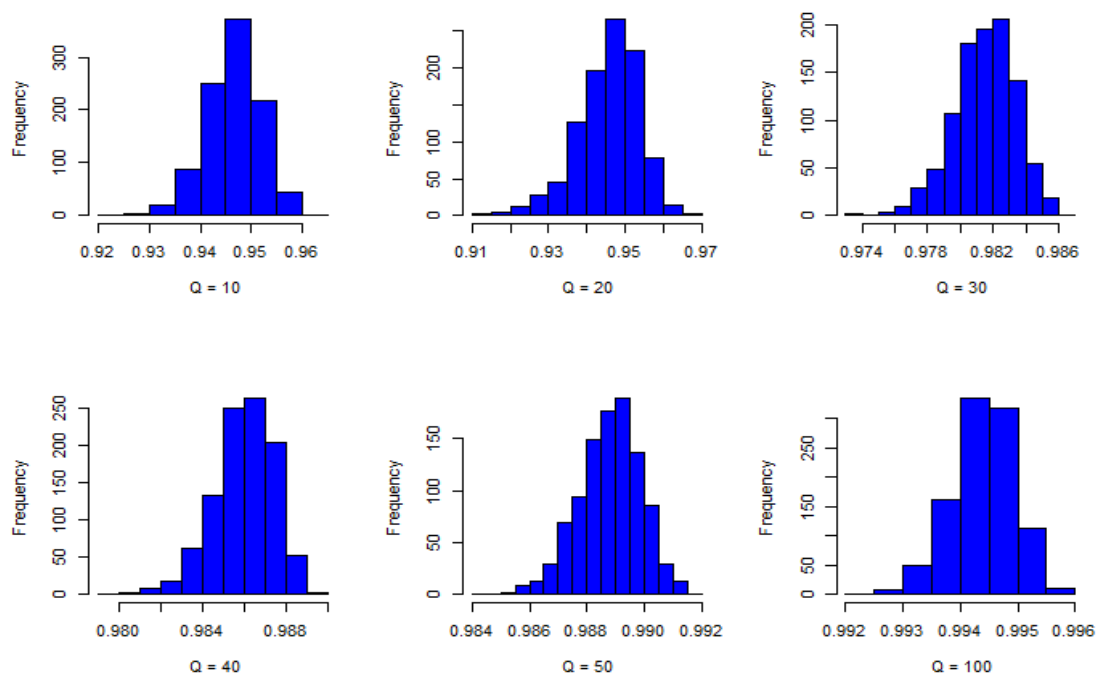
ΣΧΗΜΑ 2. Κατανομή τιμών άλφα για 10 έως 1.000.000 συμμετέχοντες.



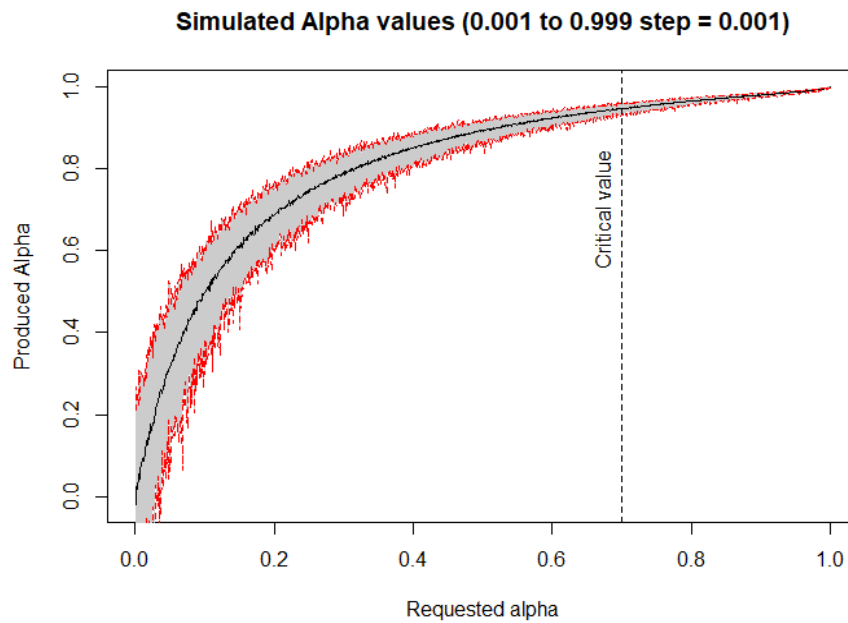
ΠΙΝΑΚΑΣ 3. Επίδραση του αριθμού των ερωτήσεων στην προσέγγιση του άλφα ($Rep = 1.000$, $N = 200$, $Alpha = 0,95$).

Αριθμός ερωτήσεων	MT	Ελάχιστη τιμή	Μέγιστη τιμή	Χρόνος υλοποίησης
10	0.9464342	0.9228699	0.9609187	2 δευτερόλεπτα
20	0.9725333	0.9621068	0.9801231	4 δευτερόλεπτα
30	0.9814750	0.9734824	0.9863947	5 δευτερόλεπτα
40	0.9859774	0.9796366	0.9897918	7 δευτερόλεπτα
50	0.9888061	0.9844060	0.9915331	8 δευτερόλεπτα
100	0.9943914	0.9922301	0.9956507	2223 δευτερόλεπτα

ΣΧΗΜΑ 3. Κατανομή τιμών άλφα για 10 έως 100 ερωτήσεις.



ΣΧΗΜΑ 4. Προσομοιωμένες τιμές άλφα κάτω από $N = 200$, $Q = 10$, $Rep = 100$, $Step = 0,001$.



ΣΧΗΜΑ 5. Απόκλιση και απόλυτη απόκλιση από τις ζητούμενες τιμές Alpha κάτω από $N = 200$, $Q = 10$, $Rep = 100$, $Step = 0,001$.

