Ελληνικό Στατιστικό Ινστιτούτο
Πρακτικά 18ου Πανελληνίου Συνεδρίου Στατιστικής (2005) σελ.435-440

THE DETERMINATION OF THE OPTIMAL SAMPLE SIZE FOR RELIABILITY SCALES IN SOCIAL SCIENCES

Athanassios Katsis
Department of Social and Educational Policy
University of Peloponnese
katsis@uop.gr

Aristea Limakopoulou
Educational Research Center
aristea@kee.gr

ABSTRACT

In many areas of research, the precise measurement of hypothesized processes or variables (theoretical *constructs, latent variables*) poses a challenge by itself. In social sciences, unreliable measurements of people's beliefs or intentions obviously hamper efforts to predict their behavior. Each measurement (response to an item) reflects to some extent the true score for the intended concept and to some extent esoteric, random error. It can be expressed in an equation as: $\mathrm{X} = \tau + \varepsilon$, where X refers to the respective actual measurement, that is subject's response to a particular item, τ is commonly used to refer to the true score and ε refers to the random error component in the measurement. An index of reliability can be defined in terms of the proportion of true score variability that is captured across subjects or respondents, relative to the total observed variability. This can be written as follows: $\mathrm{Re}\,liability = \sigma^2_{(true\ \ score)} / \sigma^2_{(total\ \ observed)}$. A measurement is reliable if it reflects mostly true score, relative to the error. The most widely used estimator of the above mentioned index is provided by Cronbach's alpha [Cronbach (1951)]. The aim of this study is the determination of the optimal sample size to estimate the probability that Cronbach's alpha exceeds a pre-specified desirable level of accuracy. The Bayesian point of view is employed.

1. INTRODUCTION

In many areas of research, the precise measurement of latent variables is usually challenging and intriguing. Typically, the reliability of a scale is assessed through the coefficient alpha [Cronbach (1951), Psarrou and Zafiropoulos (2001, p. 171)], which measures the correlation among the items of the scale.

Unfortunately, the coefficient alpha is almost never known beforehand. A random sample of the population is collected and its population value is estimated through the classical statistical techniques. On the other hand, the size of the sample is an issue of particular importance in applied research and generally in statistical theory. Furthermore, the Bayesian approach seems like a natural choice for determining the sample size since any prior beliefs about the unknown parameter are accounted for by the researcher in the planning stages.

The literature on this topic includes the work of Feldt, Woodruff and Salih (1987) for small-sample tests, Feldt and Ankenmann (1998, 1999) for comparing two reliability coefficients and Bonnet (2002) who provides a hypothesis test and interval estimation for a single coefficient alpha. In all of the above they examine if a reliability coefficient is equal to a constant or to another coefficient through a classical hypothesis testing set-up.

In this article we determine the optimal sample size regarding the reliability coefficient from the Bayesian point of view. We assume that a cluster sampling is employed, which is very common in practice. The optimality criterion focuses on the probability that the coefficient alpha exceeds a pre-specified desirable level of accuracy. Thus, the researcher obtains a very precise idea about the value of the coefficient alpha while incorporating her/his own prior beliefs about the study.

2. BAYESIAN CALCULATIONS

Let α denote the coefficient alpha based on a scale having a certain number of items. The researcher wishes to have α being greater than or equal to a specific value κ, usually around 0.70 for most practical applications. Let X define the following Bernoulli random variable:

$$X = \begin{cases} 1, & \text{if } \alpha \geq \kappa \\ 0, & \text{otherwise} \end{cases}$$

Since we are interested in the likelihood of α exceeding a pre-determined value, we focus on the parameter $\theta = P(\alpha \geq \kappa)$, which represents the mean of *X*. This parameter denotes the likelihood of the alpha coefficient exceeding the threshold κ. The prior distribution on θ is supposed to be Beta with parameters λ and ν, i.e., θ ~ Beta(λ, ν).

A random sample of n clusters is drawn and let α_i denote the sample alpha coefficient of the i-th cluster. Evidently, $X_1, ..., X_n$ are i.i.d observations from X where each X_i is defined as

$$X_i = \begin{cases} 1, & \text{if } \alpha_i \geq \kappa \\ 0, & \text{otherwise} \end{cases} \quad (i = 1, ..., n)$$

where the sample alpha coefficient of the i-th cluster is given by $\alpha_i = \dfrac{N_i \bar{r}_i}{1+(N_i - 1)\bar{r}_i}$ with N_i denoting the number of items and $\bar{r}_i$ the average iter-item correlation among the items in the i-th cluster.

Goldstein (1981) derived an upper bound for the sample size needed to estimate the mean of any distribution based exclusively on the prior measure of the mean and certain positive constants imposed by the researcher. It is imperative to note that since this is a pre-experimental procedure, the data X are unknown therefore rendering their use in any condition prohibitive.

The bound is based on the following sequential argument:

Initially, after collecting a sample of n observations, we shall terminate sampling if a bigger sample will not substantially alter the estimates based on n observations. Hence, using the constants c_1 and c_2, this is described in the condition

$$P\{\sup_{m>n} | E_m(\theta) - E_n(\theta) | > c_1 | X_1,...,X_n\} \langle c_2 \quad (1)$$

where the expectations are with respect to the posterior distribution of θ. Furthermore, we add a further condition (expand the above condition) to ensure that any future sampling will satisfy (1).

In the next step, we denote by A_m the sample of size m for which (1) is satisfied. We choose as the optimal sample size the smallest value n such as, for a positive constant c_3, the following condition is satisfied:

$$P\{X_1,...,X_m \in A_m \; for\ all\ m \geq n\} \geq 1 - c_3 \qquad (2)$$

Equation (2) aims at ensuring that any future observations will also satisfy (1). The constant quantities c_i (i=1, 2, 3) determine how strict the researcher is in estimating θ. Using results from martingale theory, an upper bound for n is given by the following:

$$n \leq V(X)[\frac{1}{c_1{}^2 c_2 c_3} - \frac{1}{V(\theta)}] \qquad (3)$$

where $V(\theta)$ is the prior variance of θ and $V(X)$ is the expected variance of X with respect to the prior distribution, i.e., $V(X) = E_\theta V(X | \theta)$. A complete proof of (3) is provided in Goldstein (1981). Straightforward calculations yield that

$$V(\theta) = \frac{\lambda \nu}{(\lambda + \nu)^2 (\lambda + \nu + 1)} \qquad (4)$$

$$\text{and} \quad V(X) = \frac{\lambda \nu}{(\lambda+\nu)(\lambda+\nu+1)}. \qquad (5)$$

Hence, in our case (3) becomes:

$$n \leq \frac{1}{c_1{}^2 c_2 c_3} \cdot \frac{\lambda \nu}{(\lambda+\nu)(\lambda+\nu+1)} - (\lambda+\nu) \quad (6)$$

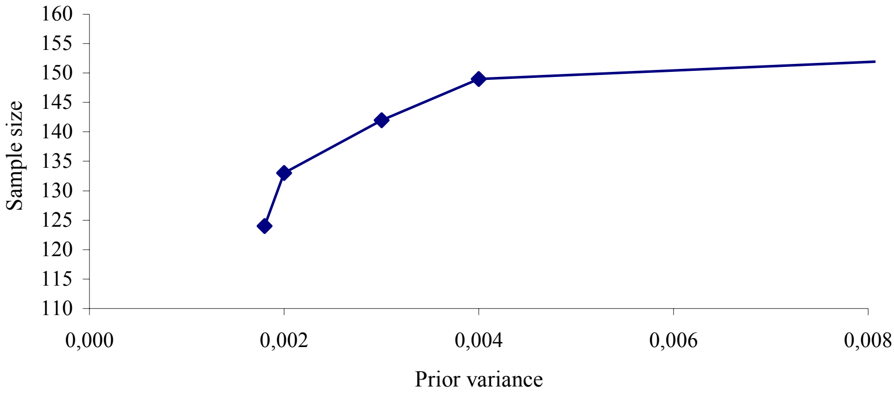
Specific results are obtained using the equality in (6) and are presented in TABLE 1 in the form of the optimal number of clusters. We have used moderate values of the constant quantities setting c_i (i=1, 2, 3) equal to 0.15. Most cases have unequal values for λ and ν to describe the sample size pattern for various values of $E(\theta)$.

TABLE 1. *Sample sizes for various prior values*

λ	ν	$E(\theta)$	$V(\theta)$	$V(\mathrm{X})$	n
9	1	0.90	0.008	0.082	152
18	2	0.90	0.004	0.086	149
27	3	0.90	0.003	0.087	142
36	4	0.90	0.002	0.088	133
45	5	0.90	0.002	0.088	124
7	3	0.70	0.019	0.191	367
14	6	0.70	0.010	0.200	375
28	12	0.70	0.005	0.205	365
56	24	0.70	0.003	0.207	330
112	48	0.70	0.001	0.209	252
3	2	0.60	0.040	0.200	390
6	4	0.60	0.022	0.218	421
12	8	0.60	0.011	0.229	431
1	1	0.50	0.083	0.167	327
0.50	0.50	0.50	0.125	0.125	246

We observe that for large values of $E(\theta)$, the optimal sample size increases as the prior variance increases as well (i.e., we become less certain about the parameter θ). We note that when the researcher strongly believes that the alpha coefficient will a priori exceed the imposed threshold (i.e., the cases in the table that $E(\theta) = 0.90$) and the prior variance of θ increases (ranging from 0.002 to 0.008), the respective sample sizes also tend to become bigger (from 124 to 152). In all these cases $V(\mathrm{X})$ remains approximately the same (see FIGURE 1). This intuitive result is not followed when $E(\theta)$ is around the non-informative value of 0.50 which indicates that the researcher is rather "neutral" about the fact that the alpha coefficient will a priori exceed the imposed threshold. In this range the prior knowledge of θ is not an important factor in determining the optimal number of clusters.

FIGURE 1: *Optimal sample sizes when E(θ)=0.90*



3. CONCLUSIONS

The above analysis is an analytical, fully Bayesian approach to the problem of deriving the optimal sample size for the reliability coefficient in the common situation of cluster sampling. It achieves the twofold purpose of attaining a minimum level for the probability of the alpha coefficient exceeding a pre-specified threshold while providing an analytical and flexible criterion for sample size calculations. Unlike the frequentist approach (its contributions are mentioned initially), the proposed methodology does not focus on the classical hypotheses testing argument or in the number of items that constitute a scale since the researcher is more interested in the likelihood that the reliability coefficient attains a specific value. The optimality criterion examines the difference that a bigger sample will entail in the parameter of interest, terminating when it is unlikely that such a "meaningful" difference will occur.

ΠΕΡΙΛΗΨΗ

Σε πολλούς τομείς της έρευνας, η ακριβής μέτρηση των υποτιθέμενων διαδικασιών ή των μεταβλητών (*constructs, latent variables*) παρουσιάζει ιδιαίτερο ενδιαφέρον. Στις κοινωνικές επιστήμες, οι αναξιόπιστες μετρήσεις των πεποιθήσεων ή των προθέσεων των ανθρώπων παρακωλύουν προφανώς τις προσπάθειες να προβλεφθεί η συμπεριφορά τους. Κάθε μέτρηση (απάντηση σε ένα ερώτημα) απεικονίζει ως ένα ορισμένο βαθμό το αληθινό αποτέλεσμα για την προοριζόμενη έννοια και ως ένα ορισμένο βαθμό το εσωτερικό, τυχαίο σφάλμα. Μπορεί να εκφραστεί με μια εξίσωση της μορφής: $X = \tau + \varepsilon$, όπου το X αναφέρεται στην αντίστοιχη πραγματική μέτρηση, δηλαδή στην απάντηση σε ένα συγκεκριμένο ερώτημα, το

τ χρησιμοποιείται συνήθως για να αναφερθεί στο σωστό αποτέλεσμα ενώ το ε αναφέρεται στο τυχαίο σφάλμα της μέτρησης. Ένας δείκτης αξιοπιστίας μπορεί να καθοριστεί ως το ποσοστό της μεταβλητότητας του αποτελέσματος σχετικά με τη συνολική παρατηρούμενη μεταβλητότητα. Αυτό μπορεί να γραφτεί ως εξής: $\mathrm{Re}\,liability = \sigma^2_{(true \quad score)} / \sigma^2_{(total \quad observed)}$. Μια μέτρηση είναι αξιόπιστη εάν απεικονίζει το σωστό αποτέλεσμα και όχι το τυχαίο σφάλμα. Ο εκτιμητής του συγκεκριμένου δείκτη που ευρύτατα χρησιμοποιείται είναι ο συντελεστής Cronbach's alpha [Cronbach (1951)]. Ο στόχος αυτής της μελέτης είναι η εκτίμηση του βέλτιστου μεγέθους δείγματος ώστε να υπολογιστεί η πιθανότητα ο συντελεστής Cronbach alpha να υπερβαίνει ένα προκαθορισμενο επιθυμητό επίπεδο ακρίβειας. Στην εργασία αυτή υιοθετείται η Μπεϋζιανή προσέγγιση.

REFERENCES

Bonnet, D.G. (2002). Sample Size Requirements for Testing and Estimating Coefficient Alpha. *Journal of Educational and Behavioral Statistics*, 27(4), 335-340.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.

Feldt, L. S., & Ankenmann, R. D. (1998). Appropriate Sample Size for Comparing Alpha Reliabilities. *Applied Psychological Measurement*, 22(2), 170-178.

Feldt, L. S., & Ankenmann, R. D. (1999). Determining sample size for a test of equality of alpha coefficients when the number of part-tests is small. *Psychological Methods*, 4, 366-377.

Feldt, L. S., Woodruff, D. J., & Salih, F.A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement*, 11, 93-103.

Goldstein, M. (1981). A Bayesian criterion for sample size. *The Annals of Statistics*, 9(3), 670-672.

Psarrou, M., & Zafiropoulos, K. (2001). Scientific research, *Theory and Applications in Social Sciences*, Dardanos, Athens (in Greek).