Confidence intervals for Cronbach's reliability coefficient

MICHAIL TSAGRIS[1], CONSTANTINOS C. FRANGOS[2], CHRISTOS C. FRANGOS[3]
[1]School of Mathematical Sciences, University of Nottingham, Nottingham, UNITED KINGDOM
pmxmt1@nottingham.ac.uk
[2]Division of Medicine, University College London, London, UNITED KINGDOM
constantinos.frangos.09@ucl.ac.uk
[3]Department of Business Administration
Technological Educational Institute of Athens, Athens, GREECE
cfragos@teiath.gr

Abstract: - Cronbach's alpha reliability coefficient was suggested as measure of the internal consistency of a questionnaire. The focus of this work is on the interval estimation of the alpha coefficient. Iacobucci and Duhachek (2003) and Koning and Franses (2006) explored transformations for constructing confidence intervals for the true value of alpha based on the asymptotic distribution of the estimate of the coefficient van Zyl et al. (2000). Padilla et al. (2012) examined confidence intervals for the alpha coefficient using bootstrap with non normal populations. We provide an alternative transformation of the alpha coefficient and construct confidence intervals for the true value. In order to do this we also employ the asymptotic distribution of the estimate of the coefficient van Zyl et al. (2000). Simulation studies are performed to demonstrate our methodology when the data are continuous and when they are converted to a Likert scale. A comparison with some present methodologies including the bootstrap methodology is presented as well.

Key-Words: - reliability, Cronbach's alpha, transformation, asymptotic normal, delta method, confidence intervals

1 Introduction

The term repeatability refers to the replication of an experiment or test under different circumstances and the reproducibility refers to the conduction of the experiment or test by another person. In both cases, the measurements should have small deviations from replication to replication. If the latter is true, then there are grounds to believe that the measurement is very accurate and the measurement errors are rather small. This means that the procedure can be thought of as being reliable. If a test for instance is administered to some people more than once and the results do not differ significantly, then the test can be assumed to have high reliability or being reliable.

The term reliability though is due to Spearman [1], who first said that errors exist even in the non sampling cases. He noted that these errors can be estimated by the size of consecutive and repetitive measures. Each measurement consists of two elements, the true value of the measurement and the error of the measurement. If the measurement is repeated it will yield new values both for the measurement and the error associated with it. The reliability is actually the ratio of the true value divided by the observed (through a measurement) value.

Lord and Novick [2] starting from this suggestion they created the classical psychometric theory of the true value. According to them, the measurement X has two components, the true value T and the measurement error ε: $X = T + \varepsilon$.

The reliability is used as a tool of estimating the percentage of the true value in a measurement, that is

$$Reliability = \frac{True\ value}{Observed\ Value} = \frac{T}{X} \qquad (1)$$

Since the replication of a test is not always feasible, another term has been added in the literature, the internal consistency. This means that one must treat the same test as it had been administered twice and estimate the reliability of it. Due to this replication difficulty many reliability coefficients have been suggested over the years. In this work, Cronbach's reliability coefficient, also known as Cronbach's alpha or alpha coefficient is under consideration.

We will examine some of the proposed ways of constructing confidence intervals for the true value of the alpha coefficient and suggest a new way. All ways will be compared via simulation studies in

ISBN: 978-1-61804-187-6

terms of the level of coverage with the focus being on the low sample size cases. We will see the performance of the methods when the data are continuous and when they are converted to a Likert scale.

2 Cronbach's reliability coefficient

Cronbach's reliability coefficient was suggested as measure of the internal consistency of a questionnaire [3]. The test cannot always be applied twice (in order to see variance between the two measurements X . For this reason Cronbach's alpha is to be used. The formula of the Cronbach's alpha reliability coefficient is the following

$$\alpha = \frac{p}{p-1}\left\{1-\sum_{j=1}^{p} Var(Y_j) \Big/ Var\left(\sum_{j=1}^{p} Y_j\right)\right\} \qquad (2)$$

where Y_j stands for the j-th variable Y (or j-th item/question).

3 Confidence intervals for the alpha coefficient

The alpha coefficient provides a point estimate of the reliability. Being an estimate of the true reliability, its standard error should be given as well leading to an interval estimation of the true value. The alpha is only an estimate of the true value.

3.1 Confidence interval based on the asymptotic normal distribution of the sample alpha coefficient

Van Zyl at al. [4] derived the variance of the distribution of the sample alpha coefficient assuming normality of the estimator. The formula for the estimated variance is

$$Va\hat{r}(\hat{\alpha}) = \frac{p^2}{n(p-1)^2}\,\omega \qquad (3),$$

$$\omega = \frac{2}{\left(j^T \mathbf{S} j\right)^3}\left[\left(j^T \mathbf{S} j\right)\left(tr\mathbf{S}^2 + tr^2\mathbf{S}\right) - 2\left(tr\mathbf{S}\right)\left(j^T \mathbf{S}^2 j\right)\right]$$

The $\mathbf{S}$ stands for the sample unbiased estimator of the true covariance matrix ($\mathbf{\Sigma}$) of the items, tr indicates the trace of a matrix j is a p-dimensional vector of ones and n is the sample size. If the items are parallel (the covariance matrix has compound symmetry), then ω simplifies to $2(p-1)(1-\hat{\alpha})^2/p$, where $\hat{\alpha}$ is the sample estimate. The asymptotic $(1-\gamma/2)\%$ confidence interval in this case has the classical formula

$$\left(\hat{\alpha} - Z_{1-\gamma/2}\sqrt{Va\hat{r}(\hat{\alpha})}\ ,\ \hat{\alpha} + Z_{1-\gamma/2}\sqrt{Va\hat{r}(\hat{\alpha})}\right) \qquad (4)$$

where $Z_{1-\gamma/2}$ is the $1-\gamma/2$ quantile of the standard normal distribution. We will refer to (4) as the ID confidence interval.

3.2 Confidence interval suggested by Koning and Franses

van Zyl et al. [4] mentioned that under the assumption of compound symmetry and by using the delta method

$$\sqrt{n}\left[\frac{1}{2}log(1-\hat{\alpha}) - \frac{1}{2}log(1-\alpha)\right] \xrightarrow{d} N\left(0\,, \frac{p}{2(p-1)}\right) \qquad (5)$$

Koning and Franses [5] used (5) to derive a stable confidence interval

$$\left\{\begin{array}{c} 1-(1-\hat{\alpha})exp\left[Z_{1-\gamma/2}\sqrt{2p/n(p-1)}\right], \\ 1-(1-\hat{\alpha})exp\left[-Z_{1-\gamma/2}\sqrt{2p/n(p-1)}\right] \end{array}\right\} \qquad (6)$$

We will refer to (6) as the KF confidence interval.

3.3 Non parametric bootstrap confidence interval

Bias corrected and accelerated confidence intervals (BC_a) were suggested by Efron [6, 7]. They are an improved form of non-parametric confidence intervals. The percentile method relies on calculating the lower and upper $(\gamma/2)\%$ of the bootstrapped sample estimates of the parameter of interest [6, 7, 8].

$$\left(\hat{\alpha}^*_{\gamma/2}\ ,\ \hat{\alpha}^*_{1-\gamma/2}\right) \qquad (7)$$

where the upper script * indicates the bootstrap sample estimate. The BC_a is a improved version of (7)

$$\left(\hat{\alpha}^*_{\gamma_1}\ ,\ \hat{\alpha}^*_{\gamma_2}\right) \qquad (8)$$

where $\gamma_1 = \Phi\left[\hat{z}_0 + \left(\hat{z}_0 + Z_{\gamma/2}\right)/\left\{1-\hat{a}\left(\hat{z}_0 + Z_{\gamma/2}\right)\right\}\right]$, $\gamma_2 = \Phi\left[\hat{z}_0 + \left(\hat{z}_0 + Z_{1-\gamma/2}\right)/\left\{1-\hat{a}\left(\hat{z}_0 + Z_{1-\gamma/2}\right)\right\}\right]$, Φ

is the standard normal cumulative distribution function and $\hat{z}_0$ and $\hat{a}$ are the bias correction and acceleration terms respectively. The value of the bias correction term is calculated from the proportion of bootstrap estimates less than the original estimate $\hat{\alpha}$ [6], $\hat{z}_0 = \Phi^{-1}\left(\#\left\{\hat{\alpha}^*(b) < \hat{\alpha}\right\}/B\right)$, where $\hat{\alpha}^*(b)$ indicates the $b-$th bootstrap sample estimate of the alpha and B is the number of bootstrap samples. The acceleration term is calculated from the original sample using the jackknife. Let $\hat{\alpha}_{(i)}$ denote the sample estimate of alpha when the $i-$th observation is deleted and $\hat{\alpha}_{(\bullet)} = \sum_{i=1}^{n} \hat{\alpha}_{(i)}/n$. Then, a simple expression for the acceleration constant is $\hat{a} = \sum_{i=1}^{n}\left(\hat{\alpha}_{(i)} - \hat{\alpha}_{(\bullet)}\right)^3 \Big/ 6\left[\sum_{i=1}^{n}\left(\hat{\alpha}_{(i)} - \hat{\alpha}_{(\bullet)}\right)^2\right]^{3/2}$.

3.4 Logit confidence interval

We suggest a new confidence interval based on the logit transformation. The logit transformation is generally used for numbers within (0,1); proportions for example. The reliability estimates the percentage of the true value in a measurement (1) and thus is a proportion. Cronbach's reliability coefficient (2) is a number within (0,1). Hence, $\hat{\theta} = log\left[\hat{\alpha}/(1-\hat{\alpha})\right]$ and via the delta method $Va\hat{r}(\hat{\theta}) = \left(1/\hat{\alpha} + 1/(1-\hat{\alpha})\right)^2 \, Va\hat{r}(\hat{\alpha})$, where $Va\hat{r}(\hat{\alpha})$ is defined in (3). An asymptotic $(1-\gamma/2)\%$ confidence interval for θ is given by

$$\left(\hat{\theta}_{low}, \hat{\theta}_{up}\right) = \begin{pmatrix} \hat{\theta} - Z_{1-\gamma/2}\sqrt{Va\hat{r}(\hat{\theta})}, \\ \hat{\theta} + Z_{1-\gamma/2}\sqrt{Va\hat{r}(\hat{\theta})} \end{pmatrix} \quad (9)$$

Finally, an asymptotic $(1-\gamma/2)\%$ confidence interval for α (denoted by LCI) is given by back-transforming the end points of (9)

$$\left[\frac{exp\left(\hat{\theta}_{low}\right)}{1+exp\left(\hat{\theta}_{low}\right)}, \frac{exp\left(\hat{\theta}_{up}\right)}{1+exp\left(\hat{\theta}_{up}\right)}\right] \quad (10)$$

4 Simulation studies

We perform simulation studies in order to examine the coverage accuracy of each of the four methods for obtaining confidence intervals for Cronbach's alpha. The important feature of the simulations is to see the performance of the methods in the low sample case.

We start with a given covariance matrix and a known value of Cronbach's alpha and generate 10,000 samples from a multivariate normal distribution with the covariance matrix. Each time we estimate the Cronbach's alpha using (2) and then construct confidence intervals using the four aforementioned methods. We estimate the true coverage of each confidence interval from the proportion of times a confidence intervals includes the true value of alpha. The nominal coverage is set to $(1-\gamma/2) = 0.95$ and hence, when the estimate lies within $(0.946, 0.954)$ we will say the method achieves the nominal coverage. The procedure will be repeated for sample sizes ranging from 15 up 1000 for two covariance matrices and the number of dimensions for each case is equal to 5.

In reality however, the questionnaires consist of ordinal data and not continuous. Hence, for each sample we will transform the continuous data into categorical and repeat the same estimating procedure. For simulated data set, we categorized each of the variables using the 15%, 30%, 60% and 75% quantile points. The first 15% of the data were assigned to the first category, the next 15% of the data to category 2 and so on.

4.1 Example 1: Clamwest data

We used the data taken from Aitchison [9]. It is a 20 x 6 data set consisting of Colour-size compositions of 20 clam colonies from West Bay. Since the data are compositional (i.e. each row vector sums to 1) we transformed them using the alpha-transformation with alpha=-1. The correlation matrix can be supplied by the authors. The resulting Cronbach's alpha was equal to 0.641.

Table 1 summarizes the estimated coverages of the 4 methods when the data are continuous. We can see the logit method (10) performs very well it attains the coverage level even for sample sizes. BC_a had on average the lowest coverage. Casella and Berger (2001) do not suggest the use of bootstrap methodology for estimating the variance of the variance. We cannot say with certainty that this is the case for the alpha coefficient even though it is a function of variances.

Table 2 presents the estimated coverages after the data have been transformed to the ordinal scale. The results do not look so good and in fact when the sample size is 1000 the estimated coverages are very

Table 1. Estimated coverages of the 4 methods when the data are continuous. Bold numbers indicate the estimated coverage was within the acceptable limits.

	Sample sizes						
Methods	**15**	**25**	**40**	**50**	**100**	**500**	**1000**
ID	0.922	0.933	0.943	**0.947**	**0.947**	**0.947**	**0.953**
KF	0.978	0.987	0.99	0.991	0.993	0.992	0.995
BC_a	0.888	0.918	0.93	0.941	0.94	**0.945**	**0.953**
Logit	**0.951**	0.956	**0.954**	0.957	**0.95**	**0.949**	**0.954**
Estimated alpha	0.606	0.623	0.628	0.633	0.637	0.641	0.641

Table 2. Estimated coverages of the 4 methods when the data are ordinal with 5 categories. Bold numbers indicate the estimated coverage was within the acceptable limits.

	Sample sizes						
Methods	**15**	**25**	**40**	**50**	**100**	**500**	**1000**
ID	0.896	0.908	0.923	0.918	0.916	0.863	0.813
KF	0.922	0.937	**0.948**	**0.947**	**0.945**	0.895	0.844
BC_a	0.901	0.924	0.941	0.94	0.935	0.883	0.83
Logit	**0.952**	**0.947**	**0.954**	**0.946**	0.934	0.874	0.824
Estimated alpha	0.613	0.635	0.642	0.648	0.652	0.658	0.658

Table 3. Estimated coverages of the 4 methods when the data are continuous. Bold numbers indicate the estimated coverage was within the acceptable limits.

	Sample sizes						
Methods	**15**	**25**	**40**	**50**	**100**	**500**	**1000**
ID	0.928	0.935	0.941	**0.951**	**0.947**	**0.944**	**0.948**
KF	0.929	0.94	0.94	**0.95**	**0.944**	**0.943**	**0.948**
BC_a	0.905	0.921	0.929	0.938	0.938	0.944	**0.947**
Logit	0.958	**0.949**	**0.947**	**0.953**	0.946	**0.944**	**0.948**
Estimated alpha	0.809	0.822	0.827	0.83	0.833	0.835	0.836

Table 4. Estimated coverages of the 4 methods when the data are ordinal with 5 categories. Bold numbers indicate the estimated coverage was within the acceptable limits.

	Sample sizes						
Methods	**15**	**25**	**40**	**50**	**100**	**500**	**1000**
ID	0.944	**0.954**	0.961	0.964	0.958	0.881	0.792
KF	0.927	0.938	0.937	0.943	0.927	0.851	0.76
BC_a	0.919	0.933	0.938	0.942	0.928	0.845	0.75
Logit	0.962	**0.953**	**0.948**	**0.953**	0.935	0.858	0.765
Estimated alpha	0.803	0.814	0.818	0.82	0.822	0.825	0.825

low. If we take a look at the last rows of each table we will see that the estimated alpha (averaged over the 10,000 simulations for each sample size) reaches the true value of alpha as the sample size increases. This is the case for the continuous data but not for the ordinal data. For the ordinal data, when the sample sizes increases, on average the estimated alpha overestimates the true value of alpha.

4.2 Example 2: Open-closed book data

The second example uses the open-closed book data [10]. The correlation matrix of this data can be provided by the authors. The alpha coefficient is equal to 0.836 based on this dataset.

The results for the estimated coverages when the data are continuous are similar to the ones obtained before. The logit method (10) performs very well, even when the sample size is low. Its estimated coverage is almost constant around 0.95, whereas the other methods need a larger sample size to attain the nominal coverage level.

When the data are transformed into the ordinal scale, results are not as good in this case either. The estimated coverage of the 4 methods is lower than the nominal one for large sample sizes. If we see the estimated alpha coefficients for all the sample sizes we will see that when the data are continuous, the estimate equals the true value of alpha. On the other hand, when the data are ordinal, there is an underestimation of the true value of alpha.

5 Conclusions

At first we can say that the logit transformation produced confidence intervals whose estimated coverage was within the acceptable limits. However this conclusion is based on two small scale simulation studies and further simulation studies need to be performed in order to have a clearer picture. In these simulations three of the suggested (in the literature) confidence intervals did not perform very well when the sample sizes are small.

When the data are transformed to a Likert-type scale the results seem to have no validity. The estimated coverages of all methods are low. In this case the question of consistency and the unbiasedness properties of the estimator of alpha are raised. The present simulation studies showed that there is bias in the estimation of alpha when the data are in a Likert scale. This explains possibly partially why the coverage of the methods was so low. The problem of bias of the estimator of alpha is not new. Zumbo et al. [11] suggested that the alpha estimator (2) should be applied to the polychoric correlation matrix of the data.

We have not included other ways of obtaining confidence intervals, such as the Asymptotically Distribution-Free (ADF) method of Maydeu-Olivares et al. [12]. We are working towards including more methods in a larger scale simulation study. The final goal is to construct a hypothesis testing for the true lower bound of alpha.

References:

[1] Spearman, C., General intelligence, objectively determined and measured, *American Journal of Psychology*, Vol.15, No.2, 1904, pp. 201-292.

[2] Lord, F., Novick, M., *Statistical theories of mental test scores*. Addison-Wesley, 1968.

ISBN: 978-1-61804-187-6

[3] Cronbach, L., Coefficient alpha and the internal structure of tests, *Psychometrika*, Vol.16, No.3, 1951, pp. 297-334.

[4] van Zyl, J., Neudecker, H., Nel, D., On the distribution of the maximum likelihood estimator of Cronbach's alpha, *Psychometrika*, Vol.65, No.3, 2000, pp. 271-280.

[5] Koning, A., Franses, H. P., *Confidence intervals for Cronbach's coefficient alpha values*. Technical report, Erasmus Research Institute of Management – ERIM, 2006.

[6] Efron, B., Tibshirani, R.J., *An introduction to the bootstrap,* Chapman & Hall CRC, 1993.

[7] Frangos, C. C., An updated bibliography on the jackknife method, *Communications in Statistics-Theory and Methods*, Vol.16, No.6, 1987, pp. 1543-1584.

[8] Tsagris, M., Elmatzoglou, I., Frangos, C. C., The Assessment of Performance of Correlation Estimates in Discrete Bivariate Distributions Using Bootstrap Methodology, *Communications in Statistics - Theory and Methods*, Vol.41, No.1, 2012, pp. 138-152.

[9] Aitchison, J., *The Statistical Analysis of Compositional Data*, Springer Netherlands, 1986.

[10] Mardia, K. V., J. T. Kent, J. M. Bibby. *Multivariate analysis*, Academic Press, 1979.

[11] Zumbo, B. D., Gadermann, A. M., Zeisser, C., Ordinal versions of coefficients alpha and theta for Likert rating scales, *Journal of Modern Applied Statistical Methods*, Vol.6, No.1, 2007, pp. 21-29.

[12] Maydeu-Olivares, A., Coffman, D., Hartmann, W.,. Asymptotically distribution-free (ADF) interval estimation of coefficient alpha, *Psychological Methods*, Vol.12, No.2, 2007, pp. 157-176.

Appendix

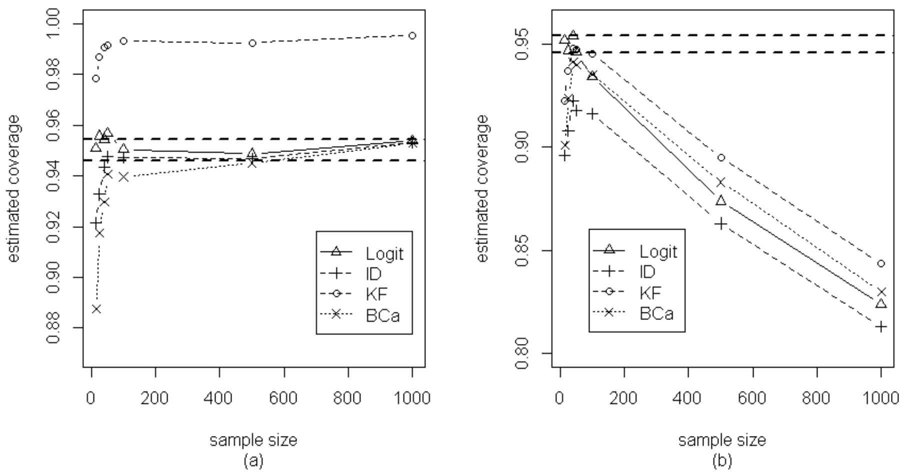

Fig 1. Estimated coverages for Example 2 when data are (a) continuous and (b) ordinal.

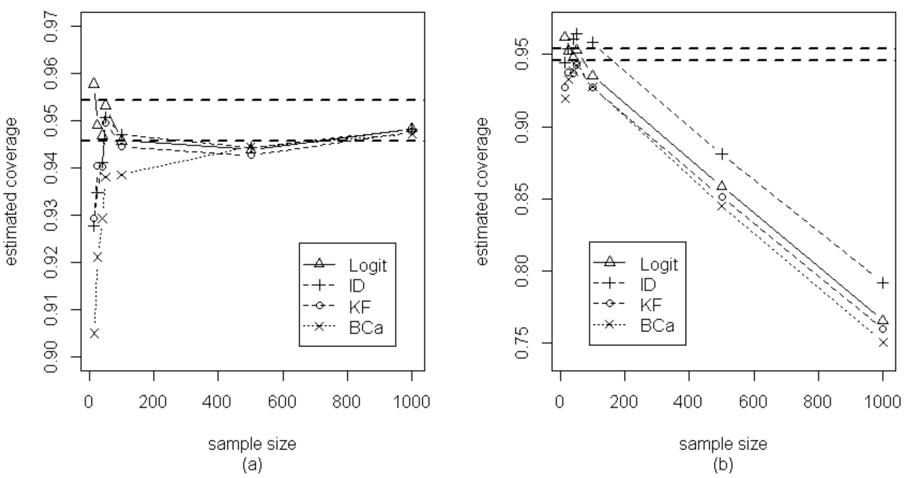

Fig 2. Estimated coverages for Example 2 when data are (a) continuous and (b) ordinal.