TESTING RELIABILITY HYPOTHESES BASED ON COEFFICIENT ALPHA

IE Business School Working Paper WP08-08 10-01-2008

Alberto Maydeu Olivares

IE Business School
Marketing Dept.
C/ Maria de Molina, 11-15
28006, Madrid – Spain
Alberto.Maydeu@ie.edu

Abstract

We show how to test hypotheses for coefficient alpha in three different situations: 1. Hypothesis tests of whether coefficient alpha equals a prespecified value. 2. Hypothesis tests involving two statistically independent sample alphas as may arise when testing the equality of coefficient alpha across groups. 3. Hypothesis tests involving two statistically dependent sample alphas as may arise when testing the equality of alpha across time, or when testing the equality of alpha for two test scores within the same sample. We illustrate how these hypotheses may be tested in a structural equation modeling framework under the assumption of normally distributed responses and also under asymptotically distribution free (ADF) assumptions. The formulas for the hypothesis tests and computer code are given for four different applied examples.

Keywords

coefficient alpha, hypothesis testing, structural equation modeling

1. Introduction

Assessing the reliability of a questionnaire or test score is one of the most frequent tasks in psychological research. Often, researchers wish to go beyond providing a point estimate of the reliability of their test score and are interested in testing hypotheses concerning the reliability of their test score. A typical situation is when a researcher wishes to determine whether the reliability of her test score is larger than some predetermined cut-off value (say .80). Another commonly encountered situation is when a researcher wishes to determine whether the reliability of her test score is equal across two or more populations. For instance, a researcher may wish to determine whether the reliability of a test score is equal across genders. Or, she may wish to determine whether the reliability of a test score is the same across several countries. Finally, sometimes researchers are interested in determining whether, within a population, the reliabilities of two test scores are equal. For instance, the researcher may wish to test whether the reliability of a test score based on p items equals the reliability of a test score based on a subset of those p items (as when a full form and a short form of a questionnaire are available). A special case of this instance is when a researcher wishes to test whether the reliability changes when a single item is removed from the scale. As another example, a researcher may wish to test whether the scores based on two subsets of items drawn from the same item domain are equally reliable.

Most often, reliability assessment is performed by means of coefficient alpha (Hogan, Benjamin & Brezinski, 2000). Consequently, in this paper we focus on performing hypothesis tests on the reliability of a test score based on coefficient alpha. Coefficient alpha (α) was first proposed by Guttman (1945) with important contributions by Cronbach (1951). For some recent discussions on α, see Cortina (1993), Miller (1995), Schmitt (1996), Shevlin, Miles, Davies & Walker (2000), and ten Berge (2000). Coefficient alpha is a population parameter and thus an unknown quantity. In applications, it is typically estimated using the sample coefficient alpha, a point estimator of the population coefficient alpha. As with any point estimator, sample coefficient alpha is subject to variability around the true parameter, particularly in small samples. Methods for performing hypothesis testing based on coefficient alpha rely on the estimation of the variability of sample coefficient alpha (i.e., its standard error). The initial proposals for estimating the standard error of coefficient alpha were based on model as well as distributional assumptions (see Duhachek and Iacobucci, 2004 for an overview). Thus, if a particular model held for the covariance matrix among the test items, *and* the test items followed a particular distribution, then a confidence interval for coefficient alpha could be obtained. The sampling distribution for coefficient alpha was first derived (independently) by Kristof (1963) and Feldt (1965) who assumed that the test items were strictly parallel (see Lord & Novick, 1968) and normally distributed. This model implies that all the item variances are equal and all item covariances are equal. However, Barchard and Haskstian (1997) found that standard errors for coefficient alpha obtained using these results were not sufficiently accurate when model assumptions were violated (i.e. the items were not strictly parallel). The lack of robustness of the standard errors for coefficient alpha to violations of model assumptions have hindered the widespread use of hypotheses tests for alpha in applications.

A major breakthrough occurred when van Zyl, Neudecker, and Nel (2000) derived the asymptotic (i.e. large sample) distribution of sample coefficient alpha without model assumptions [1]. In particular, these authors assumed only that the items composing the test were normally distributed. Duhachek and Iacobucci (2004) recently compared the performance of these model-free standard errors for coefficient alpha vs. those of model-based procedures proposed by Feldt (1965) and Hakstian and Whalen (1976) under violations of the model assumptions underlying coefficient alpha. They found that the model-free, normal theory (NT) interval estimator proposed by van Zyl et al. (2000) uniformly outperformed competing procedures across all conditions.

However, the results of van Zyl et al. (2000) assume that the items composing the test can be well approximated by a normal distribution. In practice, tests are most often composed of binary or Likert-type items for which the normal distribution can be a poor approximation. Yuan, Guarnaccia, and Hayslip (2003) have proposed a model-free and asymptotically distribution free (ADF) standard error for sample coefficient alpha that overcomes this limitation. Maydeu-Olivares, Coffman and Hartmann (2007) have shown that for sample sizes over 100 observations, ADF standard errors are preferable to NT standard errors because the latter are not sufficiently accurate when the skewness or excess kurtosis of the items is larger than one.

The purpose of the current study is to show how reliability hypotheses based on coefficient alpha can be tested using the NT results of van Zyl et al. (2000), and also using the ADF results of Yuan et al. (2003). In particular, we show how to perform hypothesis testing in three cases. Case 1 involves a single sample alpha. Case 2 involves two statistically independent sample alphas. Case 3 involves two statistically dependent sample alphas. Case 2 arises when comparing the population alpha across two independent samples, such as males and females, or across countries. Case 3 arises when comparing the population alpha for two sets of items in a single sample. Typical examples are testing the equality of population alpha when an item is dropped, testing the equality of alpha for a full scale score and a reduced scale score, or testing the equality of alpha for the same score measured at two time points. We do so by adopting a structural equations modeling (SEM) framework. A SEM framework is not needed for testing Case 1 and 2 hypotheses. Indeed, the formulae involved are straightforward. All that is needed are the standard errors of sample alpha which can be computed using the code provided by Duhachek and Iacobucci (2004) and Maydeu-Olivares et al. (2007). However, adopting a SEM framework for Cases 1 and 2 is convenient because it provides a link to Case 3 hypothesis testing, which can not be easily performed without using a SEM framework. Also, by adopting a SEM framework we can integrate the results of van Zyl et al. (2000) and Yuan et al. (2003) with the large literature on reliability assessment using SEM.

The three cases considered are illustrated using four examples. The test statistics discussed in this paper are based on large sample theory and may not be accurate in small samples. Since it is questionable to present results using arbitrary parameter values and draw generalizable conclusions from them, we show how a Monte-Carlo investigation can be performed using the simulation capabilities of SEM packages to determine the accuracy of the obtained p-values and we do so for each of the examples presented.

2. Coefficient alpha

Consider a test or questionnaire composed of p items $Y_1, \ldots, Y_p$ intended to measure a single attribute. The reliability of the test score, $X = Y_1 + \cdots + Y_p$, is defined as the percentage of variance of X that is due to the attribute of which the items are indicators. The most widely used procedure to assess the reliability of X is coefficient alpha (Guttman, 1945; Cronbach, 1951). In the population of respondents, coefficient alpha is

$$\alpha = \frac{p}{p-1}\left(1 - \frac{\sum_i \sigma_{ii}}{\sum_i \sigma_{ii} + 2\sum_{i<j}\sigma_{ij}}\right), \tag{1}$$

where $\sum_i \sigma_{ii}$ denotes the sum of the p item variances in the population, and $\sum_{i<j}\sigma_{ij}$ denotes the sum of the $\frac{p(p-1)}{2}$ distinct item covariances. In applications, a sample of N respondents from the population is available, and a point estimator of the population α given in Equation (1) can be obtained using the sample coefficient alpha

$$\hat{\alpha} = \frac{p}{p-1}\left(1 - \frac{\sum_i s_{ii}}{\sum_i s_{ii} + 2\sum_{i<j} s_{ij}}\right), \tag{2}$$

where s_{ij} denotes the sample covariance between items i and j, and s_{ii} denotes the sample variance of item i.

2.1 Coefficient alpha and the reliability of a test score

A necessary and sufficient condition for coefficient alpha to equal the reliability of the test score is that the items are *true-score equivalent* (i.e. *essentially tau-equivalent* items) in the population (Lord & Novick, 1968: p. 50; McDonald, 1999: Chapter 6). A true-score equivalent model is a one factor model in which the factor loadings are equal for all items. The model implies that the population covariances are all equal, but the population variances need not be equal for all items. Coefficient alpha also equals the reliability of the test score when the items are parallel and strictly parallel, as these are special cases of the true-score equivalent model [2].

When the items do not conform to a true score equivalent model, coefficient alpha does not equal the reliability of the test score. For instance, if the items conform to a one factor model with distinct factor loadings (i.e. *congeneric* items) then the reliability of the test score is given by coefficient omega (see McDonald, 1999). Under a congeneric measurement model, coefficient alpha underestimates the true reliability. However, the difference between coefficient alpha and coefficient omega are generally in the third decimal except in the rare cases where one of the factor loadings is very large (e.g. .9) and all the other factor loadings are very small (e.g. .2) (Raykov, 1997).

2.2 The large sample distribution of sample alpha

Equation (2) shows that $\hat{\alpha}$ is a function of the sample variances and covariances. These variances and covariances are normally distributed in large samples, not only when the item responses are normally distributed, but also under the ADF assumptions set forth by Browne (1982, 1984) [3]. As a result, and without any model assumptions, in large samples, $\hat{\alpha}$ is normally distributed with mean α and variance φ^2. The standard error of sample alpha, $\hat{\varphi}$, can be estimated using the delta method (e.g., Agresti, 2002) from the large sample covariance matrix of the sample variances and covariances. This matrix is different under NT and ADF assumptions. As a result, when the normality assumption for the items is replaced by the milder ADF assumption, the standard error for $\hat{\alpha}$ will differ and we will use $\hat{\varphi}_{NT}$ and $\hat{\varphi}_{ADF}$ to distinguish them. However, the point estimate of sample coefficient alpha, $\hat{\alpha}$, remains unchanged when NT or ADF assumptions are invoked.

The accuracy of statistical inferences for α rests on the accuracy of the standard errors for $\hat{\alpha}$. Both the NT and ADF model-free standard errors for $\hat{\alpha}$ proposed by van Zyl et al. (2000) and Yuan et al. (2003), respectively, are based on large sample theory. Fortunately, Duhacheck and Iacobucci (2004) showed that the NT standard errors can be well estimated with sample sizes as small as 30 provided the item responses are approximately normally distributed. Also, sample sizes as small as 100 observations (and in some cases 50 observations) suffice to adequately estimate ADF standard errors (Maydeu-Olivares et al., 2007).

3. Hypothesis testing for coefficient alpha

In this section, we describe the statistical theory underlying hypothesis testing for α based on the results of van Zyl et al. (2000) and Yuan et al. (2003).

Case 1: Hypothesis testing involving a single $\hat{\alpha}$

Consider testing whether α equals some a priori value, α_0 (e.g. .8 or .7). The null and alternative hypotheses are $H_0 : \alpha_{\text{dif}} \leq 0$ and $H_1 : \alpha_{\text{dif}} > 0$, where $\alpha_{\text{dif}} = \alpha - \alpha_0$. Since in large samples $\hat{\alpha}$ is normally distributed, a suitable test statistic is

$$z = \frac{\hat{\alpha}_{\text{dif}}}{\hat{\varphi}_{\text{dif}}} = \frac{\hat{\alpha} - \alpha_0}{\hat{\varphi}} \tag{3}$$

where $\hat{\varphi}_{\text{dif}}$ is the standard error for $\hat{\alpha}_{\text{dif}}$, and $\hat{\varphi}$ is either the NT standard error, $\hat{\varphi}_{NT}$, or the ADF standard error, $\hat{\varphi}_{ADF}$, depending on the distributional assumptions made. Then, the observed significance level (p-value) for the test is the area under the standard normal curve to the left of the observed z value.

Case 2: Hypothesis testing involving two statistically independent $\hat{\alpha}$'s

This case arises when a researcher is interested in comparing α in two populations (e.g., males vs. females), or in two disjoint samples from the same population. For testing the equality of alpha across two populations, the null and alternative hypotheses are $H_0 : \alpha_{\text{dif}} = 0$ and $H_1 : \alpha_{\text{dif}} \neq 0$, where $\alpha_{\text{dif}} = \alpha_1 - \alpha_2$, and α_1 and α_2 are the alpha coefficients for a test score in populations 1 and 2, respectively. An appropriate test statistic is

$$z = \frac{\hat{\alpha}_{\text{dif}}}{\hat{\varphi}_{\text{dif}}} = \frac{\hat{\alpha}_1 - \hat{\alpha}_2}{\sqrt{\hat{\varphi}_1^2 + \hat{\varphi}_2^2}} \tag{4}$$

where $\hat{\varphi}_1$ and $\hat{\varphi}_2$ are the (NT or ADF) standard errors for the estimates $\hat{\alpha}_1$ and $\hat{\alpha}_2$. For this two-tailed alternative, the *p*-value of the test is obtained as twice the area under the standard normal curve to the left of $|z|$.

Case 3: Hypothesis testing involving two statistically dependent $\hat{\alpha}$'s

Armed with the code for NT standard errors provided by Duhacheck and Iacobucci (2004), and with the code for ADF standard errors provided by Maydeu-Olivares et al. (2007), Case 1 and 2 hypothesis testing can be readily performed. In particular, for testing the equality of alpha across two populations we used the fact that the variance of the difference between $\hat{\alpha}_1$ and $\hat{\alpha}_2$ equals the sum of the variances of each sample alpha. However, a researcher may be interested in testing whether the alpha coefficients for two test scores obtained from the same sample are equal. In this case, the null and alternative hypotheses are, as in Case 2, $H_0 : \alpha_{\text{dif}} = 0$ and $H_1 : \alpha_{\text{dif}} \neq 0$, where $\alpha_{\text{dif}} = \alpha_1 - \alpha_2$, and α_1 and α_2 are the alpha coefficients for two different test scores in the same population. An appropriate test statistic is

$$z = \frac{\hat{\alpha}_{\text{dif}}}{\hat{\varphi}_{\text{dif}}} = \frac{\hat{\alpha}_1 - \hat{\alpha}_2}{\sqrt{\hat{\varphi}_1^2 + \hat{\varphi}_2^2 + 2\operatorname{cov}(\hat{\alpha}_1, \hat{\alpha}_2)}}, \tag{5}$$

where $\hat{\varphi}_1$ and $\hat{\varphi}_2$ are the (NT or ADF) standard errors of $\hat{\alpha}_1$ and $\hat{\alpha}_2$. The *p*-values are obtained as in Case 2. Notice however that in this case the variance for the difference between $\hat{\alpha}_1$ and $\hat{\alpha}_2$ also depends on the covariance between the two sample alphas because they are obtained from the same sample. There are a variety of situations where Case 3 hypotheses testing can arise. These situations are easily handled by adopting a SEM framework. The simpler Cases 1 and 2 can also be tested using a SEM framework that directly yields the *z* test statistics.

4. Hypothesis testing for coefficient alpha using a SEM framework

In this section, we describe how to test hypotheses concerning α using SEM and the model free approach of van Zyl et al. (2000) and Yuan et al. (2003). All that is needed is a SEM package that has capabilities for defining additional parameters that are functions of the parameters of the model. In this paper we used Mplus version 4 (Muthén & Muthén, 2006). We provide the annotated Mplus input files as supplementary material that can be downloaded from <psychological methods website>. The files are easy for applied researchers to use. We also provide the data used in the examples as supplementary material so that the examples below can be reproduced.

Case 1: Hypothesis testing involving a single $\hat{\alpha}$

Within a SEM framework, a model-free standard error for coefficient α can be obtained as follows:

1) Specify the model to be a $p \times p$ symmetric matrix.
2) Following Equation (1), define three additional parameters: $\gamma_1 = \sum_i \sigma_{ii}$, $\gamma_2 = \sum_{i<j} \sigma_{ij}$, and $\alpha = \frac{p}{p-1}\left(1 - \frac{\gamma_1}{\gamma_1 + 2\gamma_2}\right)$.
3) Define $\alpha_{\text{dif}} = \alpha - \alpha_0$.

The z statistic given by Equation (3) appears in the computer output as the ratio of the estimated α_{dif} divided by its standard error. The p-value for the statistic can be readily obtained from a table of standard normal probabilities.

Notice that since a fully saturated model is used in step 1) there are zero degrees of freedom and the model fits perfectly. Also, the additional parameters in step 2) do not introduce additional constraints on the model.

Different estimation methods can be used to estimate the parameters. Some popular choices are generalized least squares (GLS) estimation, maximum likelihood (ML) estimation, or weighted least squares (WLS) estimation. GLS and ML estimation can be performed under normality assumptions or with standard errors that are robust to normality (e.g. ADF) assumptions. WLS estimation assumes ADF assumptions. Because a saturated model is being fitted, all estimators (GLS, ML or WLS) lead to the same point estimate for coefficient alpha, as given in Equation (2). Also, when estimating the model under normality assumptions, GLS and ML lead to the same standard error for $\hat{\alpha}$, as given by van Zyl et al. (2000). Similarly, when estimating the model without normality assumptions, robust GLS, robust ML and WLS lead to the same standard error for $\hat{\alpha}$, as given by Yuan et al. (2003). Mplus 4 implements NT GLS estimation, NT ML estimation, robust ML estimation, and WLS estimation [4]. Also, Mplus yields as optional output confidence intervals for any parameter in the model (including additional parameters, such as α or α_{dif}).

Case 2: Hypothesis testing involving two statistically independent $\hat{\alpha}$'s

For two populations, we need to extend the previous SEM setup to two populations as follows:

1) For each population, specify the model to be a $p \times p$ symmetric matrix.
2) Define three additional parameters as above for each population. Thus, for the first population define γ_{11}, γ_{21}, and α_1. For the second population define γ_{12}, γ_{22}, and α_2.
3) Define $\alpha_{\text{dif}} = \alpha_1 - \alpha_2$.

Again, the model fits perfectly, and the z statistic given by Equation (4) appears in the Mplus output as the ratio of the estimated α_{dif} divided by its standard error. Then, the p-value for the statistic can be readily obtained from a table of standard normal probabilities. Also, a confidence interval for α_{dif} may be requested.

Case 3: Hypothesis testing involving two statistically dependent $\hat{\alpha}$'s

Consider two test scores computed on the same sample of respondents. This may occur when the two test scores being compared are alternate forms of the same test (possibly with no items in common), or when the two test scores correspond to a pre-test and post-test administration of the same test. The first test score is based on p_1 items, and the second is based on p_2 items. Some items may appear on both test scores, so that the overall number of items is $p \leq p_1 + p_2$. When no item appears on both test scores, the overall number of items is $p = p_1 + p_2$. On the other hand, $p < p_1 + p_2$ when one or more items appears on both test scores. This may occur when one test score corresponds to the full form of a test and the other test score corresponds to a reduced form of the test.

The procedure involved for testing a hypothesis involving the difference between the α's is very similar to the previous ones:

1) Specify the model to be a $p \times p$ symmetric matrix.
2) Define three additional parameters for each test score: γ_{11}, γ_{21}, and α_1 for the first test score, and γ_{12}, γ_{22}, and α_2 for the second test score.
3) Define $\alpha_{\text{dif}} = \alpha_1 - \alpha_2$.

Again, we do not impose any constraints among the p items. The model fits perfectly. The z statistic given by Equation (5) appears in the Mplus output as the ratio of the estimated α_{dif} divided by its standard error. The p value for the test statistic can be obtained from a table of standard normal probabilities.

In the next section we provide numerical examples to illustrate hypothesis testing for α under both normality assumptions and the less stringent ADF assumptions. As an example of Case 1, we test whether the population coefficient α of a scale score equals .9. As an example of Case 2, we test whether the population coefficient α's across genders are equal. We provide two examples of Case 3. In the first example we test whether the population α of the scale score equals the scale score when only half the items are used. These scale scores correspond to the full and short forms of a questionnaire. In the second example we test whether the population α of a scale score changes when the questionnaire is administered to the same respondents at two time points.

5. A numerical example: Testing reliability hypotheses based on coefficient alpha for the NPO scale scores

The Negative Problem Orientation (NPO) scale is one of the five scales of the Social Problem Solving Inventory (SPSI-R: D'Zurilla, Nezu & Maydeu-Olivares, 2002). Two forms of this inventory are available, the full form and the short form. In its full form, the NPO scale consists of 10 items. Each item is to be answered using a 5-point response scale. The short scale consists of a subset of five items. Here, we shall use two random samples, 100 male and 100 female respondents, from the normative US sample. The correlations and standard deviations among the 10 NPO items in these samples are provided in Table 1 [5]. The first five items shown in Table 1 correspond to the items composing the short form.

Insert Table 1 about here

Example 1: Testing the hypothesis that α = .9 for the full NPO scale in the female population (case 1)

Using ML and assuming the items are approximately normally distributed, $\hat{\alpha} = .88$ and $\hat{\varphi}_{NT} = .02$. The z-statistic of Equation (3) is $z = -1.08$, yielding a p-value of 0.14. Thus, we can not reject the hypothesis that α equals .9 in the female population. Because in this sample the NPO items do not markedly depart from a normal distribution we obtain almost identical results when using the milder ADF assumptions. In that case, the z-statistic is -1.04 and $p = 0.14$. Mplus also yields (upon request) confidence intervals for α. A 95% confidence interval under normality assumptions is (.85; .92) and the ADF interval is the same (to two significant digits).

Example 2: Testing the hypothesis that the population α for the full NPO scale is equal across genders (case 2).

For the male sample, under normality assumptions, $\hat{\alpha} = .84$ and $\hat{\varphi}_{NT} = .02$. The Mplus output yields an estimated α difference (males – females) of -.045 with a standard error of .03 under normality assumptions. The z-statistic from Equation (4) is -1.52, yielding a p-value of .13. We can not reject the hypothesis that population α is equal across genders. In the male sample, the NPO items do not markedly depart from a normal distribution either. As a result, when we replace the NT assumptions by ADF assumptions, a similar result is obtained; z = -1.48 and p = .14.

Example 3: Testing the equality of α between the full and short forms of the NPO scale in the male population (case 3)

For the short form in the male sample, $\hat{\alpha} = .72$ and $\hat{\varphi}_{NT} = .04$. Also, the estimated α difference (full – short) is .12 with a standard error of .03. The z statistic from Equation (5) is 4.21 and $p < .01$. We reject the hypothesis of equality of α for the full and short NPO scale scores in the male population. Again, similar results are obtained under ADF assumptions; $z = 3.60$ and $p < .01$.

As a final example, we provide another example of Case 3. This example involves testing the hypothesis of equality of α for two repeated administrations of the short forms of the NPO scale. The two administrations are 3 weeks apart. The sample includes both male and female respondents ($N = 138$). Table 2 provides the correlations and standard deviations among the five items at each administration. The first five items shown in Table 2 correspond to the first administration and the last five items correspond to the second administration.

Insert Table 2 about here

Example 4: Testing the equality of α in two repeated administrations of the short form of the NPO scale (case 3)

Under ADF assumptions a 95% confidence interval for α for the first administration is (.69; .82), whereas a 95% confidence interval for the second administration is (.79; .89). Given these intervals, it is difficult to determine whether coefficient alpha is equal across administrations. In contrast, the z statistic from Equation (5) is -3.06, $p < .01$. We clearly reject the hypothesis of invariance of α across administrations. A higher α was obtained for the second administration. For these data, a similar result is obtained under normality assumptions, $z = -2.98$, $p < .01$.

5.1 Accuracy of the p-values

As we have pointed out, the accuracy of the tests statistics rely on the accuracy of the standard errors. Duhachek and Iacobucci (2004) and Maydeu-Olivares et al. (2007) investigated the accuracy of the NT and ADF standard errors in a variety of situations and reported that they are accurate with sample sizes of 100 (and in some cases even fewer observations). However, when applying these test statistics, the applied researcher may be in doubt as to whether the conditions confronted in her study match those investigated in previous studies. In other words, the *p*-values obtained may be in doubt.

To verify the accuracy of the *p*-values for a particular study, a simulation study can be performed using the estimated parameters of the model as the true parameter values. Using the capabilities of Mplus for Monte-Carlo simulation, we investigated the accuracy of the *p*-values in each of our four examples using the actual sample size from each of the studies as the sample size in our simulations [6]. Because in our examples the items were approximately normally distributed, multivariate normal data was generated in each case using the estimated mean and covariance matrix from each example as the true mean and covariance matrix. In each case, 1000 random samples were drawn. Table 3 provides the empirical rejection rates for each of the examples. As can be seen in this table, given the small sample sizes considered, the *p*-values obtained are reliable. Also notice that the *p*-values for example 2 are somewhat more accurate than those for the remaining three examples. This may be related to the sample sizes involved. In example 2 the sample size is larger (100 males and 100 females) than in the other examples (100 females in example 1, 138 individuals in example 3, and 100 males in example 4).

Insert Table 3 about here

----

6. Discussion

Two strategies exist for drawing statistical inferences about the reliability of a test score. One strategy involves using coefficient alpha. Another strategy is to use a model-based reliability coefficient.

The model-based approach begins by fitting a measurement model to the items composing the test. When a model can not be rejected at the usual significance level, then a reliability coefficient based on the fitted model can be employed. For instance, suppose interest lies in performing hypothesis testing on a single reliability coefficient (as in Case 1). If a one factor model fits the items, then statistical inferences about the reliability of the test score can be performed using coefficient omega because this coefficient equals the reliability of the test. There is an extensive literature on using structural equation modeling to perform statistical inferences using model-based reliability estimates for a variety of measurement models (see for instance Kano & Azuma, 2003; Raykov, 2004; and Raykov & Shrout, 2002). However, implementing the model-based approach may prove difficult in applications. Often, all measurement models under consideration will be rejected. In this case, model-based reliability inferences can be based on the best fitting model found, which will fit the data only approximately. However, because the model does not fit the data exactly, the model-based reliability estimate will be biased and the direction and magnitude of the bias will be unknown. Also, in those cases where a model can be found that is not rejected by the exact goodness of fit test, the measurement model may be too complex for applied researchers to compute the appropriate reliability estimate. Finally, implementing a model-based approach becomes more difficult when interest lies in drawing reliability inferences across different populations or for test-retest situations (the Cases 2 and 3 discussed above) because different measurement models may be needed across populations or time points.

The alternative strategy, drawing inferences using population coefficient alpha, is easy to implement because it is model-free. However, researchers using this strategy should bear in mind that they are performing statistical inferences on population alpha and not on population reliability. In general, these are two different population parameters that are only equal when a tau-equivalent model fits the items. To claim inferences about population reliability using coefficient alpha, researchers need to test the adequacy of the tau-equivalent model. Further, coefficient alpha is not always a lower bound to population reliability. Coefficient alpha is a lower bound to the reliability of a test score whenever the item scores have the decomposition

$$Y_i = T_i + E_i\,, \qquad\qquad i = 1,\, \ldots,\, p, \qquad (6)$$

where the T_i's and E_i's are uncorrelated, and the E_i's are uncorrelated with each other (e.g., Novick & Lewis, 1967; Bentler, 2007). This condition is quite general. For instance, if a k-factor model fits the data, Equation (6) is satisfied, and coefficient alpha is a lower bound to the reliability of the test score. To see this, consider the k-factor model $Y_i = \mu_i + \boldsymbol{\lambda}_i'\boldsymbol{\eta} + \varepsilon_i$, where μ_i, $\boldsymbol{\lambda}_i'$ and ε_i denote the mean, the $1 \times k$ vector of factor loadings, and the unique factor for the ith item, respectively, and $\boldsymbol{\eta}$ denotes the $k \times 1$ vector of factors. Letting $\mu_i + \boldsymbol{\lambda}_i'\boldsymbol{\eta} = T_i$ and $\varepsilon_i = E_i$, the k-factor model is a special case of Equation (6). Thus, in many instances, such as when a k-factor model fits the data, coefficient alpha is a lower bound to population reliability. Nevertheless, it may be best to simply claim inferences about population alpha rather than population reliability. Claiming lower bound properties for coefficient alpha without fitting a measurement model should be avoided because when Equation (6) is not satisfied, such as when some of the errors, E_i, are correlated, population alpha may be larger than population reliability (see Komarov, 1997; Raykov, 2001; Green & Hershberger, 2000) [7].

7. Conclusion

In this paper we have shown that drawing statistical inferences for population alpha is quite straightforward. Statistical inferences for coefficient alpha are model-free and do not require assuming that the items composing the test score are normally distributed. Because of this computational ease, researchers interested in drawing statistical inferences for population reliability may want to consider drawing inferences for population alpha instead. We do believe that researchers should attempt to draw inferences for population reliability whenever possible. However, this requires that a good fitting measurement model can be found and the model-based reliability estimate is easy to compute. If a good fitting model can be found but the model-based reliability estimate is cumbersome to compute, researchers may consider drawing inferences for coefficient alpha instead. If the fitted model satisfies the conditions for coefficient alpha to be a lower bound to reliability, then drawing inferences for coefficient alpha becomes an attractive option to drawing inferences for population reliability from a computational viewpoint. When no good fitting measurement model can be found, researchers may still draw inferences for coefficient alpha, as this is a meaningful parameter per se. However, in this case, researchers drawing inferences about coefficient alpha should carefully avoid extrapolating their conclusions to population reliability or claiming that coefficient alpha is a lower bound of population reliability. These claims need be supported by model fitting. Finally, researchers drawing inferences about coefficient alpha should avoid claiming support for the unidimensionality of the items comprising the scale score. As has been shown, the computation of alpha is model-free, and in particular, it does not assume unidimensionality (see e.g. Cortina, 1993).

References

Agresti, A. (2002). *Categorical data analysis* (2nd Ed.). Hoboken, NJ: Wiley.

Barchard, K. A., & Hakstian, R. (1997). The robustness of confidence intervals for coefficient alpha under violation of the assumption of essential parallelism. *Multivariate Behavioral Research, 32*, 169–191.

Bentler, P.M. (2007). Covariance structure models for maximal reliability of unit-weighted composites. In S-Y. Lee (Ed.) *Handbook of latent variable and related models* (pp. 1-17). Amsterdam: Elsevier.

Browne, M.W. (1982). Covariance structures. In D.M. Hawkins (Ed.). *Topics in applied multivariate analysis* (pp. 72-141). Cambridge: Cambridge University Press.

Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37*, 62-83.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98–104.

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

Duhachek, A., & Iacobucci, D. (2004). Alpha's standard error (ASE): An accurate and precise confidence interval estimate. *Journal of Applied Psychology, 89*, 792-808.

D'Zurilla, T.J., Nezu, A.M., & Maydeu-Olivares, A. (2002). *Social Problem-Solving Inventory-Revised (SPSI-R).* North Tonawanda, NY: Multi-Health Systems, Inc.

Feldt, L. S. (1965). The approximate sampling distribution of Kuder– Richardson reliability coefficient twenty. *Psychometrika, 30*, 357–370.

Green, S.B. & Hershberger, S.L. (2000). Correlated errors in true score models and their effect on coefficient alpha. *Structural Equation* Modeling, *7*, 251-270.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika, 10*, 255-282.

Hakstian, A.R., & Whalen, T. E. (1976). A *k*-sample significance test for independent alpha coefficients. *Psychometrika, 41*, 219–231.

Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement, 60*, 523–531.

Kano, Y. & Axuma, Y. (2003). Use of SEM programs to precisely measure scale reliability. In H. Yanai, A. Okada, K. Shigematu, Y. Kano & J.J. Meulman (Eds.). *New Developments in Psychometrics* (pp. 141-148) Tokyo: Springer.

Kristof, W. (1963). The statistical theory of stepped-up reliability when a test has been divided into several equivalent parts. *Psychometrika, 28*, 221-238.

Komoroff, E. (1997). Effect of simultaneous violations of essential tau-equivalence and uncorrelated errors on coefficient alpha. *Applied Psychological Measurement*, *21*, 337-348.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.

Maydeu-Olivares, A., Coffman, D. L. & Hartmann, W.M. (2007). Asymptotically distribution free (ADF) interval estimation of coefficient alpha. *Psychological Methods*, 12, 157-176.

Miller, M.B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. *Structural Equation* Modeling, *2*, 255-273.

Muthén, L. & Muthén, B. (2006). *Mplus 4*. Los Angeles, CA: Muthén & Muthén.

Novick, M.R. & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, *32*, 1-13.

Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research, 32,* 329-353.

Raykov, T. (2001). Bias of coefficient α for fixed congeneric measures with correlated errors. *Applied Psychological Measurement*, *25*, 69-76.

Raykov, T. (2004). Point and interval estimation of reliability for multiple-component measuring instruments via linear constraint covariance structure modeling. *Structural Equation Modeling, 11*, 342-356.

Raykov, T. & Shrout, P.E. (2002). Reliability of scales with general structure: Point and interval estimation using a structural equation modeling approach. *Structural Equation Modeling, 9*, 195-212.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*, 350-353.

Shevlin, M., Miles, J.N.V., Davies, M.N.O., & Walker, S. (2000). Coefficient alpha: A useful indicator of reliability? *Personality and Individual Differences*, *28*, 229-237.

Ten Berge, J.M.F. (2000). Linking reliability and factor analysis: Recent developments in some classical psychometric problems. In S.E. Harmon (Ed.). *Advances in Personality*

Psychology, vol. 1. (pp. 138-156). Philadelphia: Psychology Press.

van Zyl, J. M., Neudecker, H., & Nel, D. G. (2000). On the distribution of the maximum likelihood estimator of Cronbach's alpha. *Psychometrika, 65*, 271-280.

Yuan, K.-H., Guarnaccia, C. A., & Hayslip, B. (2003). A study of the distribution of sample coefficient alpha with the Hopkins symptom checklist: Bootstrap versus asymptotics. *Educational and Psychological Measurement, 63*, 5-23.

Footnotes

[1] Only a positive definite covariance matrix is assumed. All previous derivations, which assumed particular models (e.g. tau equivalence) for the covariance matrix, can be treated as special cases of their result.

[2] In the *parallel items* model, in addition to the assumptions of the true-score equivalent model, the unique variances of the error terms in the factor model are assumed to be equal for all items. A more constrained version of the parallel items model is the *strictly parallel items* model, where additionally the item means are assumed to be equal across items.

[3] ADF estimation replaces the normality assumption by the milder assumption that eighth order moments of the distribution of the data are finite. This assumption is satisfied in the case of Likert-type items, where the distribution of each item is multinomial. The assumption ensures that the fourth order sample moments are consistent estimators of their population counterparts (Browne, 1984).

[4] In Mplus 4, GLS and ML denote GLS and ML estimation, respectively, under normality assumptions. ML estimation with robust standard errors is performed by using MLM or MLMV. MLM and MLMV yield the same parameter estimates and standard errors, and differ only in the goodness of fit statistics provided. For the models considered here, MLM and MLMV yield the same fit, a perfect fit, because the models are saturated.

[5] The correlations and standard deviations provided suffice for NT hypotheses testing involving coefficient α. For ADF hypotheses testing, the raw data is needed. The raw data is provided as supplementary materials.

[6] Mplus saves the estimated parameters to an external file. These parameters are then read by an additional Mplus input file to perform the simulation. The Mplus files used in the simulation are available from the authors upon request.

[7] In our experience, however, the situations where alpha can be larger than reliability are rather rare.

Table 1

Correlations and standard deviations among the items of the NPO

Males ($N = 100$)

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10
i1	1.00									
i2	.32	1.00								
i3	.23	.24	1.00							
i4	.28	.29	.31	1.00						
i5	.41	.44	.22	.28	1.00					
i6	.44	.35	.47	.38	.39	1.00				
i7	.21	.48	.37	.25	.33	.47	1.00			
i8	.29	.32	.50	.26	.33	.32	.28	1.00		
i9	.24	.35	.29	.29	.35	.37	.46	.33	1.00	
i10	.20	.23	.46	.26	.25	.47	.50	.28	.55	1.00
SD	1.09	1.15	1.19	1.02	.97	1.19	1.02	1.36	1.19	1.08

Females ($N = 100$)

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10
i1	1.00									
i2	.39	1.00								
i3	.39	.42	1.00							
i4	.37	.39	.46	1.00						
i5	.37	.46	.18	.43	1.00					
i6	.26	.41	.37	.51	.45	1.00				
i7	.43	.54	.47	.62	.50	.50	1.00			
i8	.20	.24	.25	.41	.33	.29	.50	1.00		
i9	.29	.42	.40	.42	.37	.42	.65	.42	1.00	
i10	.46	.50	.60	.55	.38	.39	.64	.44	.63	1.00
SD	1.13	1.14	1.19	1.30	1.17	1.21	1.23	1.20	1.34	1.31

Table 2

Correlations and standard deviations among the items of the short scale of the NPO measured at two time points ($N = 138$). Items measured at time 1 are denoted as i1 to i5 and items measured at time 2 are denoted as r1 to r5.

	i1	i2	i3	i4	i5	r1	r2	r3	r4	r5
i1	1.00									
i2	.53	1.00								
i3	.41	.33	1.00							
i4	.37	.37	.35	1.00						
i5	.47	.31	.42	.24	1.00					
r1	.52	.39	.35	.34	.45	1.00				
r2	.45	.58	.37	.36	.45	.57	1.00			
r3	.45	.41	.69	.35	.59	.51	.51	1.00		
r4	.34	.35	.28	.49	.34	.45	.36	.50	1.00	
r5	.48	.48	.37	.36	.64	.61	.59	.56	.44	1.00
SD	1.16	1.16	1.22	1.10	1.25	1.12	1.19	1.20	1.02	1.29

Table 3

Empirical rejection rates of the test statistic at the exact settings for each of our examples. In each case, 1000 replications were generated using the actual example's sample size.

Example	Sample size	Empirical rejection rates 1%	5%	10%	20%	30%	40%	50%	60%	70%	80%	90%
1	100	2.3	7.3	11.6	19.9	29.4	39.0	48.1	60.2	71.1	81.31	90.7
2	200	1.5	5.4	10.5	20.8	30.4	39.3	50.2	60.2	69.6	79.9	90.5
3	100	0.1	3.1	7.7	19.8	30.3	40.0	48.8	57.8	67.0	76.0	86.7
4	138	2.2	6.8	12.1	21.5	30.0	39.1	49.8	61.2	72.1	83.1	92.5

```
! Mplus input file to illustrate how to compute coefficient
! alpha and its confidence interval without normality assumptions
! The program reads the file npo.dat containing
! 200 observations on the items of the NPO scale collected using
! 5 response alternatives
! There are 100 respondents from each gender
Title: Testing alpha = .9 for females, NPO full scale;
DATA: FILE IS 'npo.dat';
! The following lines provide names for the variables
VARIABLE: NAMES ARE i1-i10 gender;
! only responses from male respondents are used
        USEOBSERVATIONS ARE (gender EQ 2);
        USEVARIABLES ARE i1-i10;
! For an ADF CI for coefficient alpha use MLM estimation
! For a NT CI for coefficient alpha use ML estimation
ANALYSIS: ESTIMATOR=ML;
MODEL:
! All the covariances among the items are free to be estimated
! Values in parenthesis assign names to the estimated parameters
    i1 WITH i2-i10*(cov12-cov20);
    i2 WITH i3-i10*(cov23-cov30);
    i3 WITH i4-i10*(cov34-cov40);
    i4 WITH i5-i10*(cov45-cov50);
    i5 WITH i6-i10*(cov56-cov60);
    i6 WITH i7-i10*(cov67-cov70);
    i7 WITH i8-i10*(cov78-cov80);
    i8 WITH i9-i10*(cov89-cov90);
    i9 WITH i10*(cov910);
! All the items' variances are free to be estimated
    i1-i10*(var1-var10);
MODEL CONSTRAINT:
! Declares new variables which are functions of
! previous variables
    NEW(sumvars sumcovs summat alpha alphadif);
! Defines sumvars as the sum of the items' variances
    sumvars = var1 + var2 + var3 + var4 + var5
    + var6 + var7 + var8 + var9 + var10;
! Defines sumcovs as the sum of the items' covariances
    sumcovs =
cov12 + cov13 + cov14 + cov15 + cov16 + cov17 + cov18 + cov19 + cov20
+ cov23 + cov24 + cov25 + cov26 + cov27 + cov28 + cov29 + cov30
+ cov34 + cov35 + cov36 + cov37 + cov38 + cov39 + cov40
+ cov45 + cov46 + cov47 + cov48 + cov49 + cov50
+ cov56 + cov57 + cov58 + cov59 + cov60
+ cov67 + cov68 + cov69 + cov70
+ cov78 + cov79 + cov80
+ cov89 + cov90
+ cov910;
! Defines summat as the sum of the variances and covariances
    summat = sumvars + 2 * sumcovs;
! Defines alpha
! 10/9 is number of items/(number of items-1)
    alpha = (10/9) * (1 - sumvars/summat);
    alphadif = alpha - .9;
! use CINT to obtain confidence intervals
OUTPUT: CINT;
```

```
! Mplus input file to illustrate how to test the equality of
! coefficient alpha between two populations
! without normality assumptions
! The program reads the file npo.dat containing
! 200 observations on the items of the NPO scale collected using
! 5 response alternatives
! There are 100 respondents from each gender
! The equality of coefficient alpha between males and females is
! tested
Title: Testing equality of alpha across gender, NPO full scale;
DATA: FILE IS 'npo.dat';
! The following lines provide names for the variables
VARIABLE:
        NAMES ARE i1-i10 gender;
        GROUPING IS gender (1=males 2=females);
        USEVARIABLES ARE i1-i10;
! For a ADF CI for coefficient alpha use MLM estimation
! For a NT CI for coefficient alpha use ML estimation
ANALYSIS: ESTIMATOR=ML;
MODEL males:
! All the covariances among the items are free to be estimated
! Values in parenthesis assign names to the estimated parameters
    i1 WITH i2-i10*(cov1_12-cov1_20);
    i2 WITH i3-i10*(cov1_23-cov1_30);
    i3 WITH i4-i10*(cov1_34-cov1_40);
    i4 WITH i5-i10*(cov1_45-cov1_50);
    i5 WITH i6-i10*(cov1_56-cov1_60);
    i6 WITH i7-i10*(cov1_67-cov1_70);
    i7 WITH i8-i10*(cov1_78-cov1_80);
    i8 WITH i9-i10*(cov1_89-cov1_90);
    i9 WITH i10*(cov1_910);
! All the items' variances are free to be estimated
    i1-i10*(var1_1-var1_10);
MODEL females:
! All the covariances among the items are free to be estimated
! Values in parenthesis assign names to the estimated parameters
    i1 WITH i2-i10*(cov2_12-cov2_20);
    i2 WITH i3-i10*(cov2_23-cov2_30);
    i3 WITH i4-i10*(cov2_34-cov2_40);
    i4 WITH i5-i10*(cov2_45-cov2_50);
    i5 WITH i6-i10*(cov2_56-cov2_60);
    i6 WITH i7-i10*(cov2_67-cov2_70);
    i7 WITH i8-i10*(cov2_78-cov2_80);
    i8 WITH i9-i10*(cov2_89-cov2_90);
    i9 WITH i10*(cov2_910);
! All the items' variances are free to be estimated
    i1-i10*(var2_1-var2_10);
MODEL CONSTRAINT:
! Declares new variables which are functions of
! previous variables
! subscript 1 refers to males, and 2 to females
    NEW(sumvars1 sumcovs1 summat1 alpha1
    sumvars2 sumcovs2 summat2 alpha2
    alphadif);

! COMPUTATION OF ALPHA FOR MALES
```

```
! Defines sumvars as the sum of the items' variances
    sumvars1 = var1_1 + var1_2 + var1_3 + var1_4 + var1_5
    + var1_6 + var1_7 + var1_8 + var1_9 + var1_10;
! Defines sumcovs as the sum of the items' covariances
    sumcovs1 =
    cov1_12 + cov1_13 + cov1_14 + cov1_15 +
    cov1_16 + cov1_17 + cov1_18 + cov1_19 + cov1_20
    + cov1_23 + cov1_24 + cov1_25 + cov1_26 + cov1_27 +
    cov1_28 + cov1_29 + cov1_30
    + cov1_34 + cov1_35 + cov1_36 + cov1_37 + cov1_38 +
    cov1_39 + cov1_40
    + cov1_45 + cov1_46 + cov1_47 + cov1_48 + cov1_49 + cov1_50
    + cov1_56 + cov1_57 + cov1_58 + cov1_59 + cov1_60
    + cov1_67 + cov1_68 + cov1_69 + cov1_70
    + cov1_78 + cov1_79 + cov1_80
    + cov1_89 + cov1_90
    + cov1_910;
! Defines summat as the sum of the variances and covariances
    summat1 = sumvars1 + 2 * sumcovs1;
! Defines alpha
! 10/9 is number of items/(number of items-1)
    alpha1 = (10/9) * (1 - sumvars1/summat1);

! COMPUTATION OF ALPHA FOR FEMALES

! Defines sumvars as the sum of the items' variances
    sumvars2 = var2_1 + var2_2 + var2_3 + var2_4 + var2_5
    + var2_6 + var2_7 + var2_8 + var2_9 + var2_10;
! Defines sumcovs as the sum of the items' covariances
    sumcovs2 =
    cov2_12 + cov2_13 + cov2_14 + cov2_15 +
    cov2_16 + cov2_17 + cov2_18 + cov2_19 + cov2_20
    + cov2_23 + cov2_24 + cov2_25 + cov2_26 + cov2_27 +
    cov2_28 + cov2_29 + cov2_30
    + cov2_34 + cov2_35 + cov2_36 + cov2_37 + cov2_38 +
    cov2_39 + cov2_40
    + cov2_45 + cov2_46 + cov2_47 + cov2_48 + cov2_49 + cov2_50
    + cov2_56 + cov2_57 + cov2_58 + cov2_59 + cov2_60
    + cov2_67 + cov2_68 + cov2_69 + cov2_70
    + cov2_78 + cov2_79 + cov2_80
    + cov2_89 + cov2_90
    + cov2_910;
! Defines summat as the sum of the variances and covariances
    summat2 = sumvars2 + 2 * sumcovs2;
! Defines alpha
! 10/9 is number of items/(number of items-1)
    alpha2 = (10/9) * (1 - sumvars2/summat2);

! Defines alphadif = difference between alpha for males and females
    alphadif = alpha1 - alpha2;

! use CINT to obtain confidence intervals for each estimated parameter
! and estimated alpha
OUTPUT: CINT;
```

```
! Mplus input file to illustrate how to test the equality of
! coefficient alpha between two subsets of items computed
! in the same sample
!
! The program reads the file npo.dat containing
! 200 observations on the items of the NPO scale collected using
! 5 response alternatives
! There are 100 respondents from each gender
!
! The equality of coefficient alpha between the reduced scale and the
! full scale is tested
! The reduced scale is composed of items i1 to i5
! The full scale is composed of items i1 to i10
! Inputing a data file to Mplus where the items in the reduced scale are
! together in the first columns of the file simplifies the programming

Title: Testing equality of alpha between the full and short form of NPO;
DATA: FILE IS 'npo.dat';

! The following lines provide names for the variables
VARIABLE:
   NAMES ARE i1-i10 gender;
! only responses from male respondents are used
   USEOBSERVATIONS ARE (gender EQ 1);
   USEVARIABLES ARE i1-i10;

! For an ADF CI for coefficient alpha use MLM estimation
! For a NT CI for coefficient alpha use ML estimation
ANALYSIS: ESTIMATOR=ML;

MODEL:
! All the covariances among the items are free to be estimated
! Values in parenthesis assign names to the estimated parameters
    i1 WITH i2-i10*(cov12-cov20);
    i2 WITH i3-i10*(cov23-cov30);
    i3 WITH i4-i10*(cov34-cov40);
    i4 WITH i5-i10*(cov45-cov50);
    i5 WITH i6-i10*(cov56-cov60);
    i6 WITH i7-i10*(cov67-cov70);
    i7 WITH i8-i10*(cov78-cov80);
    i8 WITH i9-i10*(cov89-cov90);
    i9 WITH i10*(cov910);

! All the items' variances are free to be estimated
    i1-i10*(var1-var10);

MODEL CONSTRAINT:
! Declares new variables which are functions of
! previously defined parameters
! Variables with subscript 1 refer to the full scale (10 items)
! Variables with subscript 2 refer to the short scale (5 items)

    NEW(sumvars1 sumcovs1 summat1 alpha1
    sumvars2 sumcovs2 summat2 alpha2
    alphadif);
```

```
! COMPUTATION OF ALPHA FOR THE FULL SCALE
!
! Defines sumvars as the sum of the items' variances
    sumvars1 = var1 + var2 + var3 + var4 + var5
    + var6 + var7 + var8 + var9 + var10;
! Defines sumcovs as the sum of the items' covariances
    sumcovs1 =
    cov12 + cov13 + cov14 + cov15 + cov16 + cov17 + cov18 + cov19 + cov20
    + cov23 + cov24 + cov25 + cov26 + cov27 + cov28 + cov29 + cov30
    + cov34 + cov35 + cov36 + cov37 + cov38 + cov39 + cov40
    + cov45 + cov46 + cov47 + cov48 + cov49 + cov50
    + cov56 + cov57 + cov58 + cov59 + cov60
    + cov67 + cov68 + cov69 + cov70
    + cov78 + cov79 + cov80
    + cov89 + cov90
    + cov910;

! Defines summat as the sum of the variances and covariances
    summat1 = sumvars1 + 2 * sumcovs1;
! Defines alpha
! 10/9 is number of items/(number of items-1)
    alpha1 = (10/9) * (1 - sumvars1/summat1);

! COMPUTATION OF ALPHA FOR THE SHORT SCALE
!
! Defines sumvars as the sum of the items' variances
    sumvars2 = var1 + var2 + var3 + var4 + var5;
! Defines sumcovs as the sum of the items' covariances
    sumcovs2=
    cov12 + cov13 + cov14 + cov15
    + cov23 + cov24 + cov25
    + cov34 + cov35
    + cov45 ;

! Defines summat as the sum of the variances and covariances
    summat2 = sumvars2 + 2 * sumcovs2;
! Defines alpha
! 5/4 is number of items/(number of items-1) in the reduced scale
    alpha2 = (5/4) * (1 - sumvars2/summat2);

! Define alphadif as the difference between alpha for the full scale
! and alpha for the short scale
    alphadif = alpha1 - alpha2;

! use CINT to obtain confidence intervals for each estimated parameter
! and estimated alpha
!OUTPUT: CINT;
```

```
! Mplus input file to illustrate how to test the equality of
! the coefficient alpha for a test score measured at two time points
! in the same sample
!
! The program reads the file retest npo.dat containing
! 138 observations on the items of the short NPO scale, measured twice.
! The scale consists of 5 items, each with 5 response alternatives
! The file contains 10 columns. The first 5 correspond to the items at time
! 1, and the last 5 correspond to the items at time 2.

Title: Test of the equality of alpha at two time points;
DATA: FILE IS 'npo retest.dat';

! The following lines provide names for the variables
VARIABLE:
   NAMES ARE i1-i10;

! For an ADF CI for coefficient alpha use MLM estimation
! For a NT CI for coefficient alpha use ML estimation
ANALYSIS:  ESTIMATOR=ML;

MODEL:
! All the covariances among the items are free to be estimated
! Values in parenthesis assign names to the estimated parameters
    i1 WITH i2-i10*(cov12-cov20);
    i2 WITH i3-i10*(cov23-cov30);
    i3 WITH i4-i10*(cov34-cov40);
    i4 WITH i5-i10*(cov45-cov50);
    i5 WITH i6-i10*(cov56-cov60);
    i6 WITH i7-i10*(cov67-cov70);
    i7 WITH i8-i10*(cov78-cov80);
    i8 WITH i9-i10*(cov89-cov90);
    i9 WITH i10*(cov910);

! All the items' variances are free to be estimated
    i1-i10*(var1-var10);

MODEL CONSTRAINT:
! Declares new variables which are functions of
! previously defined parameters
! Variables with subscript 1 refer to the scale at time 1 (5 items)
! Variables with subscript 2 refer to the scale at time 2 (5 items)

    NEW(sumvars1 sumcovs1 summat1 alpha1
    sumvars2 sumcovs2 summat2 alpha2
    alphadif);

! COMPUTATION OF ALPHA FOR THE TIME 1 SCALE
! Defines sumvars as the sum of the items' variances
    sumvars1 = var1 + var2 + var3 + var4 + var5;
! Defines sumcovs as the sum of the items' covariances
    sumcovs1 =
    cov12 + cov13 + cov14 + cov15
    + cov23 + cov24 + cov25
    + cov34 + cov35
    + cov45;
```

```
! Defines summat as the sum of the variances and covariances
    summat1 = sumvars1 + 2 * sumcovs1;
! Defines alpha
! 5/4 is number of items/(number of items-1)
    alpha1 = (5/4) * (1 - sumvars1/summat1);

! COMPUTATION OF ALPHA FOR THE TIME 2 SCALE
!
! Defines sumvars as the sum of the items' variances
    sumvars2 = var6 + var7 + var8 + var9+ var10;
! Defines sumcovs as the sum of the items' covariances
    sumcovs2=
    cov67 + cov68 + cov69 + cov70
    + cov78 + cov79 + cov80
    + cov89 + cov90
    + cov910;
! Defines summat as the sum of the variances and covariances
    summat2 = sumvars2 + 2 * sumcovs2;
! Defines alpha
! 5/4 is number of items/(number of items-1)
    alpha2 = (5/4) * (1 - sumvars2/summat2);

! Define alphadif= difference between alpha at times 1 and 2
    alphadif = alpha1 - alpha2;

! use CINT to obtain confidence intervals for each estimated alpha
OUTPUT: CINT;
```

NOTAS

NOTAS

NOTAS

Depósito Legal: M-20073   I.S.S.N.: 1579-4873