King Fahd University of Petroleum & Minerals

DEPARTMENT OF MATHEMATICAL SCIENCES

Technical Report Series

TR 421

May 2011

Sequential Testing of Two Independent Samples of Cronbach Alpha Coefficient

M.H. Omar

DHAHRAN 31261 ● SAUDI ARABIA ● www.kfupm.edu.sa/math/ ● E-mail: c-math@kfupm.edu.sa

Sequential Testing of Two Independent Samples of Cronbach Alpha Coefficient

Md Hafidz Omar*

May 15, 2011.

* *Corresponding author's* Postal address: Department of Mathematics and Statistics, KFUPM, Box 1161, Dhahran 31261, KSA. E-mail address: omarmh@kfupm.edu.sa

Abstract

Cronbach's coefficient alpha is an important measure of internal consistency of tests administered. In this paper, we propose and construct sequential testing procedures for hypotheses about Cronbach's coefficient alpha based on two independent samples. The proposed method is expected to reduce sample size for testing up to 50% less than non-sequential counterparts. To illustrate the application of the proposed methodology, some real data are explored as examples.

[**2000 AMS Classifications**] **Primary** *62E20,65C05*; **Secondary** 6*2F12, 62J07, 65C60,62F03,62F05*

Keywords: *Cronbach alpha, sequential analysis, asymptotic properties, relative precision, Monte-Carlo simulation.*

1 Introduction and Preliminaries

Sequential statistical methods were originally developed in order to obtain economic benefits in the context of industrial processes. For a trial with a positive result, early stopping means that a new product can be exploited sooner. If a negative result is indicated, early stopping ensures that resources are not wasted. Sequential methods typically lead to saving sample size, time and cost when compared with standard fixed sample procedures.

Recently, sequential estimation and testing methods have been extensively used in medical clinical trials in order to attain ethical, economical and administrative benefits. The most important benefit is the ethical one wherein, for instance, decision regarding which of two competitive treatments is better can be made as early as possible and therefore, the patients can be removed from the inferior treatment arm in due time.

From an educational testing context, when one new field test form is tested along with an already established form in a single administration, it is important to check if the new form is equally reliable as the old form. If the new form is not as reliable as the old, then more developmental work may be needed. As such, detecting any real difference as early as possible is better than later as corrective actions may be entertained to rectify the problem before the new form can be used for official student test scores.

In general, a sequential statistical testing procedure is conducted by defining an appropriate test statistic process and collecting observations overtime. Whenever an observation is accrued, the investigator would compute the value of the test statistic and compare it to design boundaries. If the statistic crosses the boundaries, then the testing would stop and a significant result would be announced about the parameter being tested. If, on the other hand, the statistic does not cross the boundaries, the collection of samples and testing procedure would continue until a pre-specified sample size has been reached, where the investigator will stop the trial. If by then no boundary has been crossed, then the investigator would announce that there is no evidence of significant result. The computations of the boundary as well as maximum sample sizes, type I and type II errors of the procedure, all depend on distributional properties of the test statistic process. Often, such test statistic process can be, under some regularity conditions, approximated by Brownian motion process whose properties are well documented.

The sequential analysis methodology was first introduced by Abraham Wald in 1947,who developed what is known as the Sequential Probability Ratio Test (SPRT) which dealt with simple null hypothesis versus a simple alternative. Wald and Wolfowitz (1948) proved some optimality properties of such SPRTs. Wald's procedure was to continue sampling unless one of two parallel boundaries are crossed by the likelihood ratio test statistic. The eventuality that the SPRT would continue indefinitely could happen, although with probability approaching 0. In addition, Wald's initial procedure has two shortcomings. The first was that it did not deal with the presence of other (nuisance) parameters other than the parameter being tests and secondly, it did not deal with composite hypotheses where, for instance, the alternative hypothesis space is not a single point. These shortcomings were partially rectified by authors such as Bartlett (1949) and Cox (1963), who both however required some stringent regularity condition that the null and alternative hypotheses spaces are contiguous to each other. Alternatively, Gombay (1996, 1997) and Gombay and Hussein (2006) relaxed the contiguity assumption and provided some tests based on the sequential likelihood ratio along with their asymptotic critical values at given significance levels.

Another approach to sequential testing is the so called "group sequential" method. This approach is suitable in particular in clinical trials involving humans where a data monitoring committee has to meet periodically to review the data. In such situations, the data need to be looked at only periodically and not after each observation has been collected due to ethical isssues. Thus, group sequential methods perform the testing at few interim analysis using batches of observations recruited in between the interim analysis. Recently, sequential and group sequential methods have been extended and applied in various scientific areas.

In this article, we propose and investigate sequential testing of hypotheses about 2-independent sample Cronbach's alpha reliability coefficients. Comparison of reliability coefficients may be important and needed in many situations. Feldt and Brennan (1989) described several of these situations. In particular, modifications of testing methods and efficacy of various grading formulas are generally assessed partly by their effects on reliability. Experimentations or pilot testing in the areas of question formats and question-selection techniques often leads to different reliability coefficients. Comparisons among several competing forms of the same instrument or among several competing instruments such as SAT-verbal and TOEFL and among distinct examinee populations necessitate the use of statistical inferences on the reliability coefficients.

If forms of an instrument are completely parallel and can be administered, independent reliability coefficients may be compared through standard techniques for the Pearson product moment correlations (Feldt and Brennan, 1989). Hays (1981), on page 467, pointed out that Fisher's transformations can be applied to each reliability coefficient. Hence, the chi-square test can be used to test the hypothesis of equality of reliability coefficients.

However, Feldt and Brennan (1989) pointed out that when test or questionnaire reliability is estimated via Cronbach's coefficient alpha, the sampling theory is not the same as that which applies to the product moment coefficients. Both Feldt (1965) and Kristoff (1963) independently developed the foundations of this sampling theory. Many researchers such as Bay (1973), Kristoff (1970), and Pandey and Hubert (1975) have also provided useful refinements to the theory. The foundations of the sampling theory are the assumptions of normality and independence of errors. In addition, it has been shown in the literature that the sampling theory also applies to test questions or questionnaire items that are dichotomously scored and as such also applies to the Kuder-Richardson formula 20, which is a special case of the Cronbach's alpha.

Testing independent alpha coefficients have been explored by Feldt (1969) and Hakstian and Whalen (1976). Feldt's procedure is limited to two alpha coefficients and employs a ratio W that is approximately distributed under the null hypothesis as an F statistic with $n_1 - 1$ and $n_2 - 1$ degrees of freedoms. Hakstian-Whalen test permits comparison of k alpha coefficients and employs a test statistic M which is distributed as a chi-square with $k - 1$ degrees of freedom when the null hypothesis of equality of coefficients is true. Bonett (2003) proposed a test known as Fisher-Bonett test, based on log transformation of the Cronbach's alpha. The test was later on extended by Feldt and Kim (2008) to the case of more than two samples. The Fisher-Bonett test or its extension by Feldt and Kim is simpler than tests based on F-statistic with approximate degrees of freedom. This is because the former has, asymptotically, standard normal distribution and the latter may be under powered because of the approximate degrees of freedom, which is often rounded to the nearest integer by many investigators. Similar behaviour was noticed in Hussein et. al. (2010) where a test of equality of two capability indices based on log transformation was proposed which, when contrasted to an approximate F-test proposed earlier for the same situation, showed a better or equal performance.

For testing equality of two alpha coefficients based on two dependent samples, Feldt (1980) and AlSawalmeh and Feldt (1994) derived three procedures. Woodruff and Feldt (1986) extended the methodology to the case of $m \geq 2$.

Although testing of Cronbach's coefficient alpha has been investigated in the literature as early as 1969 by Feldt, and several variations of tests have been investigated since then, to the best of our knowledge, no hypothesis testing under a sequential analyses scheme has ever been attempted. Therefore, investigation of sequential analysis for hypothesis testing of Cronbach's coefficient alpha is of special interest. By developing the sequential analysis scheme, hypothesis test on the Cronbach's coefficient alpha can be developed. The results of this investigation are expected to be utilized in many areas that utilize Cronbach's coefficient alpha and statistical methods for inference.

In this paper, we intend to develop a sequential procedure for testing the hypothesis that two processes have the same reliability indices alpha. In section 2, we propose and study a general Wald-type sequential testing procedure for a null hypothesis of the form H_0: $h(\theta) = \mathbf{0}$ against a two-sided alternative hypothesis, where θ is a vector of parameters and $h(.)$ is some function. In section 3, we derive a Wald-type test statistic based on the log-transformed and squared ratio of two alpha indices. By using the results of section 2, we give the sequential version of such Wald-type test statistic. In section 4, we conduct some numerical studies to examine the performance of the proposed sequential procedure, both on simulated data and real process data. Furthermore, we compare the sequential to the non-sequential Wald-type tests. Finally, in section 5 we give conclusions and avenues for further extensions.

2 A General Sequential Analyses Method

In this section, we introduce sequential testing procedures based on Wald's statistic process in a general setup. Then, we particularize the procedures to the case of testing equality of two Cronbach's coefficient alpha indices as discussed above. In the following subsections, we set up a general hypothesis of interest in the form $h(\theta) = 0$, and we state some regularity conditions needed for establishing the asymptotic properties of the sequential Wald statistic process for testing such hypothesis. We then establish a Brownian motion approximation for the statistic through which we can define the boundaries for the sequential testing of the general hypothesis.

2.1 Hypotheses and some regularity conditions

Suppose that a sequence of multivariate independent observations, $x_1, x_2, \cdots, x_k; \cdots$, is being collected over time. Assume that these observations are from a common distribution with probability density function $f(x;\theta)$, where θ is an unknown vector of parameters. Suppose that the following hypotheses are of interest;

$$H_0 : h(\theta) = \mathbf{0}, \quad \text{vs} \quad H_A : h(\theta) \neq \mathbf{0}$$

where θ is a vector of parameters of interest, and $h(\theta) : \mathbb{R}^d \to \mathbb{R}^q$: with $q \leq d$ is a vector valued function with first order derivative matrix denoted by $H(\theta)$.

For $\theta \in \Omega \subset \mathbb{R}^d, d \geq 1$,the following regularity conditions are needed.

- C1. The distribution function $F(x;\theta)$, of the vector random variable X is identifable over Ω.

- C2. There exists an open subset, $\Omega_0 \subset \Omega$, containing the true value of the parameter under the null hypothesis, such that the partial derivatives,

$$\frac{\partial}{\partial\theta_i} lnf(x;\theta), \quad \frac{\partial^2}{\partial\theta_i\partial\theta_j} lnf(x;\theta), \quad \frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k} lnf(x;\theta),$$

  exist and are continuous for all $x \in \mathbb{R}^l, \theta \in \Omega_0$.

- C3. For each $\theta \in \Omega_0$ and $k = 1, 2, 3, \cdots$, the score equation

$$\sum_{i=1}^{k} \frac{\partial}{\partial\theta} lnf(x;\theta)$$

  has unique solution.

- C4. There are functions, $M_1(x)$ and $M_2(x)$, that have finite expectations under any of the parameter values, $\theta \in \Omega_0$, such that

$$\left|\frac{\partial}{\partial\theta_i} lnf(x;\theta)\right| \leq M_1(x), \quad \left|\frac{\partial^2}{\partial\theta_i\partial\theta_j} lnf(x;\theta)\right| \leq M_2(x),$$
$$\left|\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k} lnf(x;\theta)\right| \leq M_2(x),$$

  for all $\theta \in \Omega_0$ and $1 \leq i, j, k \leq d$.

- C5. $E_\theta \frac{\partial}{\partial\theta_i} lnf(x;\theta) = 0, 1 \le i \le d, \theta \in \Omega_0$.
- C6. $I_{ij}(\theta) = -E[\frac{\partial^2}{\partial\theta_i\partial\theta_j} lnf(x;\theta)]$, and $I^{-1}(\theta)$ exist and are continuous for all $\theta \in \Omega_0$ and $1 \le i, j \le d$.
- C7. $Var[\frac{\partial^2}{\partial\theta_i\partial\theta_j} lnf(x;\theta)] < \infty$ for $1 \le i, j \le d$.
- C8. $E_\theta \left| \frac{\partial}{\partial\theta_i} lnf(x;\theta) \right|^{2+\delta} < \infty$ for $i = 1, 2, \cdots, d$, and for some $\delta > 0$.
- C9. The function $h(\theta)$ is continuously differentiable over Ω_0 and its first-order derivative matrix, $H(\theta)$, is bounded and of rank q.

Conditions C1-C7 are standard for the existence of multivariate maximum likelihood estimators (c.f., Serfling, 1980), whereas C8 is necessary for the strong invariance principles used in proving our results. In the following section, we will give large sample approximations for the sequential counterpart of this quantity.

2.2 Asymptotic results and test procedures

In a non-sequential testing situation with sample of size k, one would consider the quantity

$$W_k = kh(\hat{\theta})[H'(\theta)I^{-1}(\theta)H(\theta)]^{-1}h(\hat{\theta})^t \tag{2.1}$$

and then replace the parameter θ by some consistent estimator in order to obtain a Wald-type test statistic. To prepare the way for the sequential testing procedures, we first show in the following theorem that this quantity can be approximated by a functional of Brownian motions.

Theorem 2.1. Under H_0, if conditions C1-C9 hold, then there exist independent Wiener processes, $B_j(t), j = 1, 2, \cdots, q$, such that for $\alpha \le \frac{1}{2} - \frac{1}{2+\delta}, \delta > 0$,

$$\sup_{1 \le t < \infty} \left| W_{[kt]} - U(kt) \right| = O(k^{-\alpha}(\log\log k)^{1/2}) \qquad a.s., \tag{2.2}$$

where

$$U(x) = \frac{1}{x}\sum_{j=1}^{q} B_j^2(x)$$

$$W_{[kt]} = [kt]\left\{ h(\hat{\theta}_{[kt]})[H'(\theta)I^{-1}(\theta)H(\theta)]^{-1}h(\hat{\theta}_{[kt]})^t \right\} \tag{2.3}$$

as [.] *denotes the integer part of its argument.*

Proof. As stated in Hussein (2005), by using a two-term Taylor expansion, the Law of Iterated Logarithm and Lemma 2.1 of Gombay and Horvath (1994) which states that if θ is the true parameter of the model then the Euclidean distance $\left\|\hat{\theta}_k - \theta\right\|$ is almost surely $O(((\log\log k)/k))$, we have

$$-\sum_{i=1}^{k} \frac{\partial}{\partial\theta} \ln f(x;\theta) = k(\hat{\theta}_k - \theta)\frac{1}{k}\sum_{i=1}^{k} \nabla_\theta^2 \ln f(x_i;\theta) + O(k^{-\alpha}(\log\log k)^{3/2}k^{-1/2}) \qquad a.s. \tag{2.4}$$

Using C4, C6, and the Law of Iterated Logarithm,

$$\sqrt{k}(\hat{\theta}_k - \theta) = \frac{1}{\sqrt{k}}\sum_{i=1}^{k} \frac{\partial}{\partial\theta} \ln f(x;\theta) I(\theta)^{-1} + O(k^{-\alpha}(\log\log k)^{3/2}k^{-1/2}) \qquad a.s. \tag{2.5}$$

The first term in the right hand side of the above expression is the usual normalized score vector, and its covariance matrix is the Fisher information matrix, $I(\theta)$. Next, following Sen and Singer (1993, p. 220), let $\mathbf{u}$ be vector in $\mathbb{R}^d$ and consider the Taylor expansion

$$h(\theta + k^{-1/2}\mathbf{u}) = h(\theta) + k^{-1/2}H(\theta)\mathbf{u} + k^{-1/2}\left\{H(\theta^*) - H(\theta)\right\}\mathbf{u}$$

where $\theta^* = \theta + k^{-1/2}\gamma\mathbf{u}$, $0 \le \gamma \le 1$. Now by setting $\mathbf{u} = \sqrt{k}(\hat{\theta}_k - \theta)$ and exploiting the facts that $H(\theta)$ is continuous, $h(\theta) = \mathbf{0}$ under the null hypothesis and that $\left\|\hat{\theta}_k - \theta\right\|$ is almost surely $O(((\log\log k)/k))$, we get that

$$\begin{aligned}\sqrt{k}h(\hat{\theta}_k) &= H(\theta)\sqrt{k}(\hat{\theta}_k - \theta) + \left\{H(\theta^*) - H(\theta)\right\}(\hat{\theta}_k - \theta) \\ &= H(\theta)\sqrt{k}(\hat{\theta}_k - \theta) + O(((\log\log k)/k))\end{aligned} \tag{2.6}$$

By replacing $\mathbf{u} = \sqrt{k}(\hat{\theta}_k - \theta)$ in (2.6) by its expression given in (2.5) and noticing that we have assumed boundedness of $H(\theta)$ so that its product by error terms that are $O(.)$ will be intact, we get

$$\begin{aligned}\sqrt{k}h(\hat{\theta}_k) = H(\theta)&\left\{\frac{1}{\sqrt{k}}\sum_{i=1}^{k} \frac{\partial}{\partial\theta} \ln f(x;\theta) I(\theta)^{-1}\right\} \\ &+ O(((\log\log k)^{3/2}/k^{-1/2})) \qquad a.s.\end{aligned} \tag{2.7}$$

The covariance matrix of the main term in the right hand side of the above equation is $H'(\theta)I^{-1}(\theta)H(\theta)$. Now (2.2) follows from the invariance principles in Einmahl (1987) and Lemma 3.3 of Gombay (1996).

Corollary. *Under the conditions of Theorem 1,*

$$\max_{1\le k\le n} [(k/n)W_k]^{1/2} \xrightarrow{D} \sup_{1\le t\le n} \left(\sum_{j=1}^{q} B_j^2(t)\right)^{1/2}$$

where W_k is as defined in (2.1) and $\xrightarrow{D}$ denotes convergence in distribution (Serfling, 1980).

Proof. The proof is similar to that of corollary 3.1 in Hussein (2005).

The quantity W_k cannot be used directly for testing purposes, as it contains the unknown parameter θ. If we replace the parameter by any almost surely convergent estimator, we can show that, as in Gombay et al. (2007), the above corollary will still remain valid. Using the MLE $\hat{\theta}_k$, we thus can define a valid test statistic as

$$W_k^* = kh(\hat{\theta}_k)[H'(\hat{\theta}_k)I^{-1}(\hat{\theta}_k)H(\hat{\theta}_k)]^{-1}h(\hat{\theta}_k)^t.$$

Now, the Corollary allows us to define the following α-level sequential test, truncated at the maximal allowable sample size n_0.

TEST. *For $k = 2, 3, \cdots, n_0$, monitor $[(k/n_0)W_k^*]^{1/2}$, and reject H_0 the first time it exceeds $CV(\alpha)$. If it doesn't exceed $CV(\alpha)$ by n_0 then do not reject H_0.*

The maximal sample size n_0 is usually decided by the investigators either for financial reasons or for statistical reasons such as attaining a desired power. The critical value $CV(\alpha)$ is available from Borodin and Salminen (1996). For $q = 1$ it is obtained from the well known distribution of $\sup_{0<t<1} |B(t)|$. For instance, when $q = 1$, the critical values are 1.96, 2.24, and 2.80 for $\alpha = 0.1, 0.05$, and 0.01 respectively. For significant levels other than these and for $q = 1$, one can compute the critical values from the formula

$$1 - \alpha = \frac{4}{\pi}\sum \frac{(-1)^k}{2k+1} \exp\left(\frac{-\pi^2(2k+1)^2}{8CV^2}\right). \tag{2.8}$$

Often, the first five terms of this formula are sufficient to give accurate results.

3 Testing the Cronbach α_c indices

3.1 Sequential Analyses Method

Suppose that we are interested in sequential testing of hypotheses of the form H_o: $\alpha_1 = \alpha_2$ vs H_A: $\alpha_1 \neq \alpha_2$where α_i is Cronbach's coefficient alpha. We assume first that the tests are tau-equivalent (see Lord & Novick, 1968), and that the covariance matrix of the test parts have compound symmetry. That is, the covariance matrix Σ, has elements $\sigma_{i\neq j} = \sigma^2\rho$ and $\sigma_{ii} = \sigma^2$ for $i = 1, 2, \ldots m$ where σ^2 is a constant and m is the number of parts of the test administered. In such case, the Cronbach's alpha is given by

$$\alpha_c = \frac{m}{m-1}(1 - \frac{tr(\Sigma)}{1_m \Sigma 1_m}) = \frac{m\rho}{1+(m-1)\rho}$$

where 1_m is an m x 1 vector of 1's, ρ is known as intra-class correlation parameter.

To adopt the general sequential testing procedure described in the previous section to the situation of testing the equality of two Cronbach's α_c indices, we need to assume that the samples from the two processes arrive independently in pairs. Therefore, assume that we have a sequence of independent and identically distributed 1x2m-dimensional vector of data of the form $\mathbf{z}_i = (\mathbf{x}_i; \mathbf{y}_i)$ for $i = 1, 2, 3, \cdots, k, \cdots, n$ with pdf given by $f(\mathbf{z}_i; \theta) = f(\mathbf{x}_i; \mu_1, \sigma_1, \alpha_{c_1}) f(\mathbf{y}_i; \mu_2, \sigma_2, \alpha_{c_2})$, where $\theta = (\mu_1, \sigma_1, \alpha_{c_1}, \mu_2, \sigma_2, \alpha_{c_2})$. We denote the sample mean vectors, sample covariance matrices, and sample coefficient alphas based on the data up to the kth pair of observations from the two processes by

$$\begin{aligned}
\bar{\mathbf{x}}_k &= \frac{1}{k}\sum_i^k \mathbf{x}_i., \quad \mathbf{S}_{xk} = \frac{1}{k-1}\sum_i^k (\mathbf{x}_i - \bar{\mathbf{x}}_k)(\mathbf{x}_i - \bar{\mathbf{x}}_k)' \\
\bar{\mathbf{y}}_k &= \frac{1}{k}\sum_i^k \mathbf{y}_i., \quad \mathbf{S}_{yk} = \frac{1}{k-1}\sum_i^k (\mathbf{y}_i - \bar{\mathbf{y}}_k)(\mathbf{y}_i - \bar{\mathbf{y}}_k)' \\
\hat{\alpha}_{c_{1k}} &= \frac{m}{m-1}(1 - \frac{tr(\mathbf{S}_{xk})}{1_m \mathbf{S}_{xk} 1_m}) = \frac{m\hat{\rho}}{1+(m-1)\hat{\rho}} \\
\hat{\alpha}_{c_{2k}} &= \frac{m}{m-1}(1 - \frac{tr(\mathbf{S}_{yk})}{1_m \mathbf{S}_{yk} 1_m}) = \frac{m\hat{\rho}}{1+(m-1)\hat{\rho}} \qquad (3.1)
\end{aligned}$$

van Zyl, Neudecker, & Nel (2000) derived the asymptotic distribution of $(1/2)\ln(1-\hat{\alpha}_c)$ under the assumption of normality and compound symmetry to be as follows:

$$\sqrt{n}[(1/2)\ln(1-\hat{\alpha}_c) - (1/2)\ln(1-\alpha_c)] \xrightarrow{D} N(0, \frac{m}{2(m-1)}) \qquad (3.2)$$

where $\xrightarrow{D}$ represents convergence in distribution. Thus, it can be shown that for two independent samples that

$$\begin{aligned}
&\sqrt{n}[(1/2)\ln(1-\hat{\alpha}_{c_{1k}}) - \sqrt{n}[(1/2)\ln(1-\hat{\alpha}_{c_{2k}})] \\
&\qquad \xrightarrow{D} N((1/2)\ln(1-\alpha_1) - (1/2)\ln(1-\alpha_2), \frac{m}{m-1})
\end{aligned}$$

Testing equality of two Cronbach's alpha coefficients based on two independent samples can be carried out but will require re-parametrization of the problem so that the two density functions have common parameter of interest. Since a log function is an increasing one-to-one function, testing the hypothesis H_o: $\alpha_1 = \alpha_2$ is equivalent to H_o: $\ln(1-\alpha_1) - \ln(1-\alpha_2) = 0$ and we can take advantage of (**??**) in the hypothesis testing plan. We consider the two-sided alternative hypothesis of the form

$$H_A : (1/2)\ln(1-\alpha_{c_1})] - (1/2)\ln(1-\alpha_{c_2})] \neq 0.$$

Therefore, the function $h(.)$ used in the general theory of section 2 becomes, in this case,

$$h(\theta) = h(\alpha_{c_1}, \alpha_{c_2}) = (1/2)\ln(\frac{1-\alpha_{c_1}}{1-\alpha_{c_2}}).$$

Notice that in this case the dimension of the problem are $d = 2$ and $q = 1$, thus $h(\theta) = 0$. Furthermore, under the null hypothesis $h(\theta) = 0$.

Now, the Wald statistic is given by

$$\begin{aligned} W_k^* &= kh(\hat{\theta}_k)[H'(\hat{\theta}_k)I^{-1}(\hat{\theta}_k)H(\hat{\theta}_k)]^{-1}h(\hat{\theta}_k)^t \\ &= kh^2(\hat{\theta}_k)[H'(\hat{\theta}_k)I^{-1}(\hat{\theta}_k)H(\hat{\theta}_k)]^{-1}. \end{aligned}$$

where $h(\hat{\theta}_k) = h(\hat{\alpha}_{c_1}, \hat{\alpha}_{c_2}) = (1/2)\ln\left(\frac{1-\hat{\alpha}_{c_1}}{1-\hat{\alpha}_{c_2}}\right)$ and $H(\hat{\theta}_k)$ is a partial derivative matrix of $h(\theta)$ computed at $\hat{\theta}_k$.

After some simple algebra, we obtain the Wald test statistic process in the following closed form

$$W_k^* = k\left(\frac{1}{2}\ln\left(\frac{1-\hat{\alpha}_{c_1}}{1-\hat{\alpha}_{c_2}}\right)\right)^2 / \left(\frac{m}{m-1}\right). \tag{3.3}$$

For the sequential testing, one could adopt the following algorithm:

Step 1. Set a desired maximum sample size n_0 and significance level α. Find the critical boundary CV , corresponding to the desired significance level α, from the formula (2.8). If the desired significance level is either $\alpha = 0.1$, 0.05, or 0.01, then the corresponding critical values are 1.96, 2.24, and 2.80, respectively.

Step 2. Collect an initial sample of size $k \geq 2$ pairs of observations $(\mathbf{x}_i; \mathbf{y}_i)$, $i = 1, ..., m$, where $\mathbf{x}_i$ are measurements from an instrument with cronbach reliability coefficient α_{c_1}, and $\mathbf{y}_i$ are measurements from an instrument with cronbach reliability coefficient α_{c_2}.

Step 3. Compute sample means and sample variances based on the k pairs of observations using equations (3.1) and then compute W_k^* from equation (3.3).

Step 4. If $k \leq n_0$ and $(\frac{k}{n_0}W_k^*)^{1/2} \geq CV$, the critical value CV of Step 1, then reject the hypothesis that states the two processes have the same cronbach's α_c reliability indices and favor the two-sided alternative hypothesis that states they do not have equal cronbach's α_c reliability indices and stop testing.

Step 5. If $k < n_0$, and $(\frac{k}{n_0} W_k^*)^{1/2} < CV$, then set $k = k + 1$ and collect one more pair of observations from the two processes and go to Step 3. However, if $k = n_0$ and $(\frac{k}{n_0} W_k^*)^{1/2} < CV$, then stop testing and fail to reject the null hypothesis.

3.1.1 Simulation Data

In this simulation study, we explore the sequential testing of 2-sample Cronbach α_c coefficients under various conditions as follows:

1. Number of parts, $m = 40$.
2. Number of examinees $n_0 = 500$, 750, and 1000
3. The sequential anayses are done for $k = 1, 2, \cdots, 25$ data collection periods
4. The analyses in (1) through (3) above are replicated 10000 times.

With these simulation conditions, we are able to examine the empirical power of the sequential testing for the equivalence of 2-sample Cronbach α_c coefficients under the aforementioned various conditions. In addition, we are able to examine the summary statistics of the stopping time periods of the sequential testing. The results are compiled and discussed next.

3.2 Results and Discussion

In this section, results of the simulation are compiled and summarized. Tables 1 provides a summary of the empirical powers of the sequential testing while Tables 2 provides summary information for the stopping times for the sequential testing.

Table 1 below provides the empirical power of the sequential test analyses for various values of Cronbach α_c and n when $m = 40$. Several observations can be made. First, when the true difference between the 2-sample Cronbach α_c coefficients is **zero**, the empirical power observed are the observed significance levels which we expect to be around 0.05. However, there appear to be some bias in the sequential analysis where the significance level is understated. Second, when the true difference between the 2-sample Cronbach α_c coefficients is **not zero**, it matters at what Cronbach α_c reliability level the test is done. The empirical power is larger if the first Cronbach α_c reliability is 0.80 compared to when it is 0.75 even when the true difference is the same (0.05). Third, the empirical power increases to almost 1 when the true difference between the 2-sample Cronbach α_c coefficients reaches 0.10. Fourth, empirical power is higher with larger examinee sample size, n.

$\alpha_{C_2} - \alpha_{C_1}$	α_{C_1} 0.75	0.80	0.85
		$n = 500$	
0	0.0425	0.0410	0.0422
0.05	0.6299	0.8516	
0.10	0.9999		
		$n = 750$	
0	0.0396	0.4070	0.4010
0.05	0.8074	0.9598	
0.10	0.9999		
		$n = 1000$	
0	0.0402	0.0379	0.0405
0.05	0.9108	0.9900	
0.10	1.0000		

Table 1. Empirical power of the Sequential test for various values of Cronbach α_c and n_0 when $m = 40$.

Table 2 below provides summary of the stopping periods of the sequential test analyses for various values of Cronbach α_c and n when $m = 20$. Several observations can be made. First, when the true difference between the 2-sample Cronbach α_c coefficients **is zero**, we expect that the stopping periods to average more than the 25 periods or no stopping at all. However, some type I errors occur which result in the average stopping periods in the 20s in all conditions considered. Second, when the true difference between the 2-sample Cronbach α_c coefficients is **not zero**, it matters at what Cronbach α_c reliability level the test is done. The average stopping time is larger if the first Cronbach α_c reliability is 0.75 compared to when it is 0.80 even when the true difference is the same (0.05). Third, the average stopping times decreases to almost 31% to 40% of the total sample n_0 when the true difference between the 2-sample Cronbach α_c coefficients reaches 0.10. Fourth, the average stopping time decreases faster with larger examinee sample size, n_0. Fifth, the standard error of stopping times are smallest when the true difference between the 2-sample Cronbach α_c coefficients is 0.10.

	α_{C_1}					
	0.75		0.80		0.85	
$\alpha_{C_2} - \alpha_{C_1}$	M	SE	M	SE	M	SE
			$n = 500$			
0	20.01	3.97	20.34	3.83	20.05	4.00
0.05	18.51	4.29	17.04	4.38		
0.10	10.86	2.95				
			$n = 750$			
0	20.87	3.47	20.33	3.93	20.94	3.65
0.05	17.38	4.40	15.10	4.21		
0.10	8.88	2.17				
			$n = 1000$			
0	20.55	3.82	20.26	3.80	20.84	3.56
0.05	16.17	4.36	13.45	3.80		
0.10	7.72	1.73				

Table 2. Summary of Stopping periods of the Sequential test for various values of Cronbach α_c and n_0 when $m = 40$.

To provide more context to the results in Tables 2, we can convert all the average stopping times by dividing with the last stopping period of 25 which would be the case if no sequential testing is done. By doing this, we effectively compare the efficiency of the sequential testing procedure to the regular one-time testing procedure for the 2-sample Cronbach α_c. In addition, this ratio also provides us how much saving in sample size is produced on average by employing the sequential testing procedure. Overall, we find that the sample sizes needed for the sequential analyses can decrease from only around 31 to 40% of the total sample sizes needed for the regular hypothesis testing for 2-sample Cronbach α_c.

In a nutshell, the simulation study corroborates the theoretical developments described in the previous section of the paper. And we can clearly see the advantage of employing a sequential testing procedure in this case.

However, the theoretical and empirical advantages shared in this paper will not be very meaningful without examining the practical impact of the procedure on real data. This work is presented next.

4 Example

In this section, we share an example involving real data. Kolen and Brennan (2004) described two forms of a test of equal lengths of 36-items each as an illustrative example in an equating design. These forms, designated as forms X and Y have reliability of 0.8421 and 0.8597 with large sample sizes of 1655 and 1638 examinees respectively. The van Zyl's Z-test (van Zyl et. al, 2003) for 2-independent sample Cronbach alpha provides a Z statistic of 3.3413 with a 2-tailed p-value of 0.00083. That is, the

Cronbach alpha coefficients for these two independent forms are statisticaly different from one another. Now, the question we pose in this paper is at what point in the data monitoring stream can we actually detect a difference. For this, we need to do a sequential analysis.

For the sequential analysis, we will treat n_0 to be $n_0 = 1655 + 1638 = 3293$.To detect a significant difference of this magnitude with power of 0.90 and a two sided hypothesis at 0.05 significance level, the transformed difference of 0.058263 required a sample size of 3095.354 which rounds up to 3096. This minimum sample size is about 94% of the sample size n_0 in the real data.

Figures 1 and 2 below provide the results of monitoring Kolen and Brennan's data. Figure 1 shows the Cronbach α_c estimates for both forms X and Y where form Y values are denoted by dashed lines..It appears that at all monitoring periods, form Y is more reliable than form X as estimated by their Cronbach alpha coefficients. But, at what point in the data monitoring stream are these forms actually statistically different in Cronbach reliability α_c.

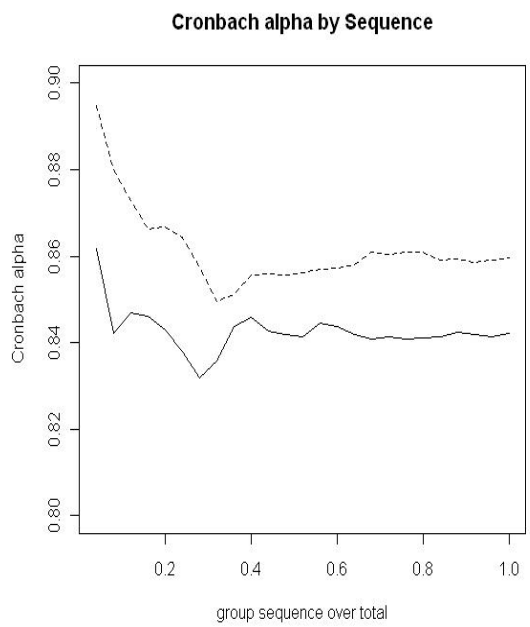

Figure 1. Cronbach α_c for forms X and Y (dashed) over monitoring periods

Figure 2 provides the root of the Wald test statistic function for every monitoring period. With the sequential analysis, we first detected a significance difference at sample size of 3038. This is about 3038/3293 or 92.26% of the original sample size.

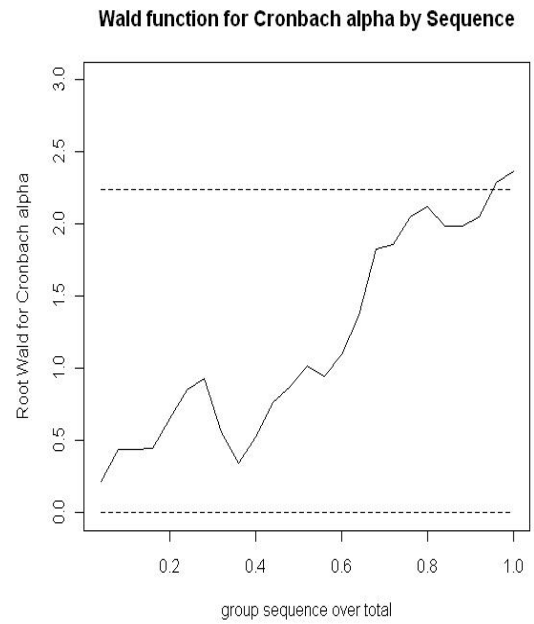

Figure 2. Wald test statistics function over monitoring periods

5 Conclusion and Further Research

In this paper, we have studied the possibility of testing Cronbach coefficient α_c using a sequential analysis scheme. We provided some theoretical development of the sequential testing. We then provided some simulation analyses and reported their results. And finally, we shared an example on sequential testing using real life data.

We found in this paper that generally for testing 2 independent samples Cronbach coefficient α_c, the sample sizes needed for the sequential analyses can decrease to only around 31 to 40% of the total sample sizes needed for the regular hypothesis testing for 2-sample Cronbach α_c. However, as can be seen in our example, this same percentage may not always be obtained for real data. Nevertheless, there is still some amount of savings involved.

In addition, we found that the empirical power for the sequential test increases quite rapidly when the true difference between Cronbach α_c reaches 0.10. This effect is further amplified if one of the Cronbach α_c is already above 0.75.

In conclusion, the theoretical development, the simulation study and the analyses on real data collectively purport the advantages of using a sequential testing of 2 independant samples Cronbach coefficient α_c. And the larger the expected difference between the Cronbach coefficient α_c, the more useful the procedure is for the practitioner if one of the main concern is saving sample sizes.

6 Acknowledgements

The research work of Dr. Omar is supported by a grant from King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, KSA under project #FT100025. Dr. Omar would like to express gratitude to Dr. AbdulKadir Hussein for consultations on this project when he was at KFUPM as a visiting professor of the Department of Mathematics and Statistics

7 References

AlSawalmeh, Yousef M. and Feldt, L. S. (1999). Testing the Equality of Two Independent Coefficients Adjusted by the Spearman-Brown Formula. *Applied Psychological Measurement*, 23(4), 363–370.

Bartlett, M. S. (1949), 'The large sample theory of sequential tests', *Proc. Camb. Phil. Soc.* 42, 239–244.

Bay, K. S. (1973). The effect of non-normality on the sampling distribution and standard error of reliability coefficient estimates under an analysis of variance model. *British Journal of Mathematical and Statistical Psychology*, 26, 45-47.

Bonett, D. G. (2003). Sample size requirements for comparing two alpha coefficients. *Applied Psychological Measurement*, 27, 72–74.

Cox, D. R. (1963).Large sample sequential tests for composite hypothesis. *Sankya A* 25, 5–12.

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

Einmahl, U. (1987). Strong invariance principles for partial sums of independent random vectors. *Ann. Probab.* 15, 1419–1440.

Einmahl, U. (1989). Extensions of results of kolm´os and tusn´ady to the multivariate case. *J. Multivariate Anal.* 28, 20–68.

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

Feldt, L. S. (1969). A test of the hypothesis that Cronbach's alpha or Kuder-Richardson reliability coefficient twenty is the same for two tests. *Psychometrika*, 40, 557-561.

Feldt (1980). A test of the hypothesis that Cronbach's alpha reliability coefficient is the same for two tests administered to the same sample. *Psychometrika*, 45(1), 99-105.

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

Feldt, L. S. and Kim, S. (2006). Testing the Difference Between Two Alpha Coefficients With Small Samples of Subjects and Raters. *Educational and Psychological Measurement*, 66(4), 589-600.

Feldt, L.S. and Kim, S. (2008). A Comparison of Tests for Equality of Two or More Independent Alpha Coefficients. *Journal of Educational Measurement*, 45(2) 179–193.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R.L. Linn (Ed)'s *Educational Measurement* (3rd ed.). Phoenix: ACE-Oryx Express.

Fisher, R.A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population, *Biometrika*, 10, 507-521.

Gombay, E and Hussein, A.(2006). A class of sequential tests for two-sample composite hypotheses., *Canadian Journal of Statistics*, 34(2): 217-232.

Gombay, E. (1996). The weighted sequential likelihood ratio. *Canadian J. Statist.* 24, 229–239.

Gombay, E. (1997). The likelihood ratio under noncontiguous alternatives. *Canadian J. Statist.* 25, 417–423.

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

Hays, W. L. (1981). *Statistics* (3rd ed.). New York: Holt, Rinehart, and Winston.

Hussein, A. and Carriere, K. C. (2005). Group Sequential Procedures under Variance Heterogeneity. *Statistical Methods in Medical Research.*, 14 (2): 121-128.

Hussein, A. (2005). Sequential comparison of two treatments using weighted Wald type statistics. *Communications in Statistics-Theory & Methods*,34(7)

Kolen, M.J. and Brennan, R.L. (2004). *Test Equating, Scaling, and Linking: Methods and Practices* (*2nd ed*). Springer Science + Business Media, Inc: New York.

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

Kristoff, W. (1970). On the sampling theory of reliability estimation. *Journal of Mathematical Psychology*, 7, 371-377.

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

Pandey, T. N. and Hubert, L. (1975). An empirical comparison of several interval estimation procedures for coefficient alpha. *Psychometrika*, 40, 169-181.

Bhatti, S. and Hussein, A. (2010). Sequential testing of equality of two process capability indices. *Technometrics*, (To be submitted soon)

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.

Wald, A. & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. statist.* 19, 326–339.

Yin Cui, Yuejiao Fu and Hussein, A. (2009). Group sequential testing of homogeneity in genetic linkage analysis, *Computational Statistics and Data Analysis*, 53(10), 3630-3639.