A TEST OF THE HYPOTHESIS THAT CRONBACH'S ALPHA
RELIABILITY COEFFICIENT IS THE SAME FOR TWO TESTS
ADMINISTERED TO THE SAME SAMPLE

LEONARD S. FELDT
UNIVERSITY OF IOWA

In measurement studies the researcher may wish to test the hypothesis that Cronbach's alpha reliability coefficient is the same for two measurement procedures. A statistical test exists for independent samples of subjects. In this paper three procedures are developed for the situation in which the coefficients are determined from the same sample. All three procedures are computationally simple and give tight control of Type I error when the sample size is 50 or greater.

Key words: coefficient alpha, reliability coefficients, tests of significance.

Measurement studies in education and psychology occasionally call for a test of the hypothesis that Cronbach's coefficient alpha is the same for two tests or measurement procedures. For example, a researcher might wish to determine if the reliability of one approach to the assessment of a particular trait differs from that of an alternative approach. An investigator might want to evaluate the effect on reliability of a training program for evaluators or study the reliability implications of variations in test directions. Such situations demand a statistical test of the hypothesis that the population values of the reliability coefficients are equal.

If the values of coefficient alpha are obtained from independent random samples, the technique proposed by Feldt [1969], and the extension to the k-sample situation by Hakstian and Whalen [1976], may be used. This approach has been found to control Type I error quite precisely even in the limiting case of dichotomously scored items, in which coefficient alpha reduces to Kuder-Richardson Formula 20. However, when the coefficients are determined from the same sample of examinees, the Feldt statistical test cannot be validly employed. If applied to coefficients obtained from the same sample, the test would be unduly conservative. The purpose of this paper is to derive several approximate tests that may be used in this situation and to report the results of sampling studies which bear on the Type I error control of these tests.

The sampling theory for independent groups draws on the analysis of variance approach to computation of coefficient alpha. Regardless of the length of the score scale for each test part or item, \( r_a \) may be computed as \( 1 - (m_{SP}/m_S) \), where \( m_{SP} \) is the mean square for the parts (items) by subjects interaction and \( m_S \) is the mean square for subjects. Kristof [1963] and Feldt [1965] demonstrated that if the scores on \( k \) parallel parts of a test conform to the assumptions of the two-factor random model, \( (1 - \rho_a)m_S/m_{SP} \) is distributed as a central \( F \) with \( N - 1 \) and \( (N - 1)(k - 1) \) degrees of freedom. In this expression \( N \) is the number of examinees and \( \rho_a \) is the population value of the coefficient alpha. Since \( m_S/m_{SP} \) equals \( 1/(1 - r_a) \), it follows that \( (1 - \rho_a)/(1 - r_a) \) is distributed as \( F_{N-1,(k-1)(N-1)} \) and \( (1 - r_a)/(1 - \rho_a) \) is distributed as \( F_{(k-1)(N-1),N-1} \).

Building on this theory, Feldt [1969] noted that

The author is indebted to Jerry S. Gilmer for development of the computer programs used in this study.

Requests for reprints should be sent to Leonard S. Feldt, College of Education, Division of Educational Psychology, Measurement & Statistics, The University of Iowa, Iowa City, Iowa 52242.

0033-3123/80/0300-2834 $00.75/0
© 1980 The Psychometric Society
where $r_1$ and $r_2$ are coefficients based on independent samples. He further showed that if $(k_1 - 1)(N_1 - 1)$ and $(k_2 - 1)(N_2 - 1)$ are fairly large, as they would be with at least 21 items and 51 subjects, the distribution of the product of the two independent $F$ variables is practically identical to that of $F_{N_1-1,N_2-1}$. This makes possible a relatively simple test of $H_0: \rho_1 = \rho_2$, since

$$W = \frac{1 - r_2}{1 - r_1} \sim F_{N_1-1,N_2-1} \left( \frac{1 - \rho_2}{1 - \rho_1} \right).$$

If $\rho_1 = \rho_2$, $W$ is distributed as a central $F$ with $N_1 - 1$ and $N_2 - 1$ degrees of freedom. If the population coefficients are not equal, $W$ is distributed as a constant greater than or less than 1.0 times a central $F$. If $W$ is too large or too small to be accepted as a randomly drawn central $F$, the implication is that the constant $(1 - \rho_2)/(1 - \rho_1)$ does not equal 1.0 and hence the hypothesis is false.

With repeated use of the same sample for both instruments, the foregoing theory breaks down. Crucial to its derivation is the stipulation that $m_{sss}/m_{SS}$ for the first instrument be independent of $m_{sss}/m_{SS}$ for the second. This condition cannot be met if the scores on the two instruments under study are correlated. Statistical tests for which such dependence is assumed are developed in the following section.

**An Approximate Test of the Hypothesis $\rho_1 = \rho_2$ with Related Samples**

Assume that Tests 1 and 2, whose reliabilities are to be compared, have been taken by the same sample of $N$ examinees. Test 1 is composed of $k_1$ scoreable units and Test 2 of $k_2$ scoreable units. The scores and score distributions for the two tests may not be directly comparable, since the instruments may be of different lengths and may contain different kinds of exercises.

For each test the following assumptions are made:

(i) The examinees are a random sample from the population of interest.

(ii) The $k_1$ and $k_2$ units are random samples from the populations of units included in the domains covered by the tests.

(iii) In the entire population of examinees, true scores are normally distributed on the two tests. For Tests 1 and 2 the correlation between true scores is greater than or equal to zero.

(iv) Throughout the entire examinees-by-units matrix for Test $j (j = 1, 2)$, the errors of measurement associated with the part-test scores are homogeneous in variance and normally distributed. However, the variance of these errors is not necessarily the same for Tests 1 and 2.

(v) Errors of measurement on the parts of the tests are independent of each other and of the true scores, both within and across tests.

These are the usual assumptions associated with the two-factor, random model of analysis of variance. Independence of errors of measurement, within and across tests, is generally assumed within classical test theory.

We first note that the following expected values hold for the various mean squares:

$$E[m_{ss}] = E\left( \frac{\hat{\sigma}_{x}^2}{k_1} \right) = \frac{\sigma_{x}^2}{k_1}$$

$$E[m_{s}] = E\left( \frac{\hat{\sigma}_{x}^2}{k_2} \right) = \frac{\sigma_{x}^2}{k_2}$$