

# Journal of Educational and Behavioral Statistics

<http://jebbs.aera.net>

---

## Estimation of the Box Correction for Degrees of Freedom from Sample Data in Randomized Block and Split-Plot Designs

Huynh Huynh and Leonard S. Feldt

*JOURNAL OF EDUCATIONAL AND BEHAVIORAL STATISTICS* 1976 1: 69

DOI: 10.3102/10769986001001069

The online version of this article can be found at:

<http://jeb.sagepub.com/content/1/1/69>

---

Published on behalf of



American Educational Research Association

and



<http://www.sagepublications.com>

Additional services and information for *Journal of Educational and Behavioral Statistics* can be found at:

**Email Alerts:** <http://jebbs.aera.net/alerts>

**Subscriptions:** <http://jebbs.aera.net/subscriptions>

**Reprints:** <http://www.aera.net/reprints>

**Permissions:** <http://www.aera.net/permissions>

**Citations:** <http://jeb.sagepub.com/content/1/1/69.refs.html>

ESTIMATION OF THE BOX CORRECTION FOR DEGREES OF FREEDOM FROM  
SAMPLE DATA IN RANDOMIZED BLOCK AND SPLIT-PLOT DESIGNS

Huynh Huynh

Leonard S. Feldt

Boston College

University of Iowa

*Key words: Box Correction; ANOVA Assumptions; Block Design*

ABSTRACT

It has been suggested that when the variance assumptions of a repeated measures ANOVA are not met, the df of the mean square ratio should be adjusted by the sample estimate of the Box correction factor,  $\epsilon$ . This procedure works well when  $\epsilon$  is low, but the estimate is seriously biased when this is not the case. An alternate estimate is proposed which is shown by Monte Carlo methods to be less biased for moderately large  $\epsilon$ .

INTRODUCTION

In the randomized block ANOVA design with  $k$  treatments and  $n$  blocks, the population covariance matrix must satisfy certain conditions if the mean square ratio is to be distributed as  $F$  with  $\nu_1 = k - 1$ ,  $\nu_2 = (k - 1)(n - 1)$ . Assuming blocks are randomly drawn and interaction effects are normally distributed, the matrix requirement will be met if, for all pairs of treatments, the variances of differences are homogeneous. Symbolically, the last requirement demands that  $\sigma_{x_i}^2 + \sigma_{x_j}^2 - 2\sigma_{x_i x_j} = \sigma_{x_i - x_j}^2$  must be constant for  $i, j = 1, \dots, k (i \neq j)$ . (Huynh and Feldt, 1970). In applied work, this requirement is generally interpreted to demand that the variances within the treatments must be

equal, and the correlations between the measures under all pairs of treatments must be equal.

Data from the Iowa Tests of Basic Skills (1974) may be used to illustrate these conditions. The five principal scores--Vocabulary, Reading, Language, Work Study Skills, and Mathematics--are reported to have the following variances and correlations in grade 6 of the national norming sample. (Variances are presented in parentheses.)

	V	R	L	W	M
V	(261.1)	.81	.74	.72	.69
R		(230.4)	.75	.78	.73
L			(264.7)	.78	.74
W				(182.2)	.82
M					(165.9)

In this sample ( $n = 2548$ ) the correlations are reasonably homogeneous, but the variances exhibit a moderate degree of heterogeneity. When an analysis of variance was considered to identify the areas in which boys and girls differed significantly in average achievement, this matrix indicated that the ANOVA assumptions were almost certainly violated.

Research on ethnic group differences may pose the same problem. For example, Lesser, Fifer, and Clark (1965) compared the profiles of average scores of four ethnic groups on a battery of four ability tests. The correlation matrices strongly suggest that the data for these populations failed to satisfy the assumptions of the ANOVA model.

The effect of such violation of assumptions was clearly shown by Box (1954). For any arbitrary covariance matrix which deviates from the required form, the mean square ratio is approximately distributed as  $F$  with reduced degrees of freedom. The amount of the reduction is dictated by a multiplicative factor,  $\epsilon$ , which depends on the population covariance matrix. When the matrix fulfills the condition noted above,  $\epsilon = 1.0$ ; otherwise,  $\epsilon < 1.0$ , with a minimum

of  $\frac{1}{k-1}$ . This result was extended to the split-plot

design by Geisser and Greenhouse (1958). Application to the problem of comparing the test profiles of several groups was presented by the same authors (1959).

Since  $\epsilon$  depends on the population matrix and is almost always unknown, Greenhouse and Geisser (1959) suggest a three-step approach to testing significance. The mean square ratio is first compared to the critical value of  $F$  with  $k - 1$  and  $(k - 1)(n - 1)$  degrees of freedom. If the ratio does not exceed this value, the analysis ends, since adjustment of the  $df$  would result in a larger critical value. If the ratio surpasses this critical value, it is compared to the critical value for  $F$  with 1 and  $n - 1$  degrees of freedom. A ratio larger than this second critical value may be unequivocally declared significant, since the  $df$  have been maximally reduced. When the mean square ratio exceeds the first critical point but not the second, the issue is in doubt. In this situation Greenhouse and Geisser suggest two alternatives: use the exact test procedures (Hotelling, 1931; Winer, 1971) or adjust the  $df$  with  $\epsilon$  estimated from the sample matrix. Analogous steps are suggested for the interaction test in the split-plot design.

The second suggestion, adjusting the  $df$ , may be more appealing than the first under some circumstances. Imhof (1962) has shown that when  $n$  is close to  $k$ , the power of the  $T^2$  test compares very unfavorably even to the conservative  $F$  test. Also, the statistic  $T^2$  is difficult to compute by hand. Prompted by these considerations, Collier et al. (1967) and Stoloff (1970) investigated the effect on Type I error of adjusting  $df$  on the basis of  $\epsilon$  computed from the sample matrix. Their data suggest that when  $n$  is less than twice the number of treatments,  $\hat{\epsilon}$  may be seriously biased if  $\epsilon$  is near or a little above .75. (As noted later, such values of  $\epsilon$  are fairly common in the case of standardized test batteries.) The estimate then tends to over-correct the degrees of freedom and produces a more stringent significance level than the nominal level being employed.

The basic idea of using a sample estimate seems sound. But it would be desirable to find an estimate of  $\epsilon$  which is less biased and less dependent on large sample size when the covariance matrix deviates only moderately from the classical model. The derivation of such an estimate and its validation by Monte Carlo methods was the subject of this study.

#### DERIVATION OF THE STATISTIC $\hat{\epsilon}$

Let  $\Sigma$  be the covariance matrix of a randomized block design. The matrix includes diagonal elements  $\sigma_{ii}$  ( $i=1, \dots, k$ ), the population variances within treatments,

and off-diagonal elements  $\sigma_{ij}$  ( $i \neq j$ ;  $i, j=1, \dots, k$ ), the population covariances between treatments. In the split-plot design it is assumed that each of the independent populations under the second treatment dimension has the same covariance matrix  $\Sigma$ . The correction factor  $\epsilon$  is defined by the formula

$$\epsilon = k^2 (\bar{\sigma}_{ii} - \sigma_{..})^2 / (k - 1) (\sum_{ij} \sigma_{ij}^2 - 2k \sum_i \sigma_{i.}^2 + k^2 \sigma_{..}^2) \quad (1)$$

In this expression  $\bar{\sigma}_{ii}$  is the mean of the variances,  $\sigma_{i.}$  the mean of the  $i$ th row or column, and  $\sigma_{..}$  the mean of all elements in the population covariance matrix. The quantity  $k$  represents the order of the matrix, that is, the number of levels of the treatment which gives rise to repeated measures on the blocks or the subjects.

An alternate expression for  $\epsilon$  is as follows:

$$\epsilon = \frac{(\sum \lambda_i)^2}{(k - 1) \sum \lambda_i^2} \quad (2)$$

In this expression  $\lambda_i$  ( $i=1, \dots, k-1$ ) are the positive eigenvalues of the matrix  $A\Delta$ , where  $A = \{\delta_{ij} - 1/k\}$ . The quantity  $\delta_{ij}$  is Kronecker's delta; that is,  $\delta_{ij} = 1$  if  $i = j$  and zero otherwise.

The estimator  $\hat{\epsilon}$  is obtained by substitution of sample values for parameters in (1). (These are pooled estimates in the case of the split-plot design.) Thus, the formula for  $\hat{\epsilon}$  becomes

$$\hat{\epsilon} = k^2 (\bar{S}_{ii} - S_{..})^2 / (k - 1) [\sum_{ij} S_{ij}^2 - 2k \sum_i S_{i.}^2 + k^2 S_{..}^2] \quad (3)$$

If the parent population is multivariate normal,  $\hat{\epsilon}$  is the maximum likelihood estimator for  $\epsilon$  (Anderson, 1958).

It is probably intuitively obvious that  $\hat{\epsilon}$  is biased when the population matrix is, in fact, homogeneous. For any covariance matrix, be it a sample or population matrix, the numerical maximum of  $\epsilon$  is 1.0. If there are any deviations from the condition that  $\sigma_{x_i}^2 - x_j$  is a constant, that is, if any pair of treatments has a variance of differences not equal to the variance of differences for any

other pair,  $\epsilon$  will be less than 1.0. A sample matrix can always be expected to evidence some heterogeneity of this kind, even though the population matrix does not. Thus, when  $\hat{\epsilon}$  is applied to the df, some reduction will occur even though none is truly called for.

It would be desirable to have an unbiased estimator for  $\epsilon$ . Such an estimator, unfortunately, is not known. When the numerator and denominator of a fraction involve statistics that are not independent, derivation of the expected value of the fraction may prove extremely difficult. With unknown expected value, correction for bias is impossible. Even if the expected value were known, correction of bias when  $\epsilon \neq 1$  would probably be impossible unless one is prepared to accept estimates greater than 1.

The expected values of the numerator and the denominator of a ratio are often considerably less difficult to derive than the expected value of the ratio itself. Thus, statisticians sometimes use the ratio of unbiased estimators in situations involving intractable ratios. Hajek (1962), for example, used this method in estimating the effective degrees of freedom for the t-test in a Behrens-Fisher problem. This approach was used in the present study. It was not obvious without Monte Carlo study, however, that a ratio of unbiased estimators would prove less biased than the maximum likelihood estimator,  $\hat{\epsilon}$ .

For the sake of simplicity the randomized block design is considered first. A generalization is made later to the split-plot design with common covariance matrix. Let A be the numerator of the right side of the equation (3), and B the denominator, deprived of k-1. Thus

$$A = k^2 (\bar{S}_{ii} - S_{..})^2$$

$$B = \sum_{ij} S_{ij}^2 - 2k \sum_{i.} S_{i.}^2 + k^2 S_{..}^2$$

It may be shown that

$$A = (k - 1)^2 (\bar{S}_{ii} - \bar{S}_{ij})^2,$$

in which  $\bar{S}_{ii}$  denotes the average variance and  $\bar{S}_{ij}$  the average covariance in the sample matrix. Since the treatments by blocks or error mean square equals  $(\bar{S}_{ii} - \bar{S}_{ij})$ ,

$$A = (k - 1)^2 MS_{\text{error}}^2 = \left[ \frac{SS_{\text{error}}}{n - 1} \right]^2$$

Box (1954) showed that under normality

$$SS_{\text{error}} = \sum_{i=1}^{k-1} \lambda_i \chi_i^2$$

where the  $k - 1$  chi-square variates, each with  $n - 1$  degrees of freedom are independent. Thus, the expected value of  $A$  is

$$E[A] = \frac{1}{(n - 1)^2} E \left[ \sum \lambda_i \chi_i^2 \right]^2 .$$

Using  $E[\chi_i^2]^2 = 2(n - 1) + (n - 1)^2$  and  $E[(\chi_i^2)(\chi_j^2)] = (n - 1)^2$ , one may ultimately derive

$$E[A] = \frac{2 \sum \lambda_i^2}{n - 1} + (\sum \lambda_i)^2 \quad (4)$$

To find the expected value for  $B$  in terms of  $\lambda_i$ , it is sufficient to consider the expectations of  $S_{ij}^2$ ,  $S_{i.}^2$  and  $k^2 S_{..}^2$ . The expectation of  $k^2 S_{..}^2$  can be obtained by noting that

$$k^2 S_{..}^2 = [\bar{S}_{i.} + (k - 1) \bar{S}_{ij}]^2 .$$

The expression within the brackets is the block mean square which is distributed as  $k \sigma_{..} \chi_{(n-1)}^2 / (n - 1)$ . Thus

$$E[k^2 S_{..}^2] = E[k \sigma_{..} \chi_{(n-1)}^2 / (n - 1)]^2 ,$$

and hence

$$E[k^2 S_{..}^2] = k^2 \sigma_{..}^2 + 2k^2 \sigma_{..}^2 / (n - 1) .$$

The expectations of  $S_{ij}^2$  and  $S_{i.}^2$  can be obtained by noting that the matrix  $S = (S_{ij})$  has the Wishart  $(\frac{\Sigma}{n - 1}, n - 1)$  distribution. From Anderson (1966, p. 161), the following expression can be obtained:

$$E(S_{ij}^2) = \sigma_{ij}^2 + (\sigma_{ij}^2 + \sigma_{ii} \sigma_{jj}) / (n - 1)$$

and

$$E(S_{i.}^2) = \sigma_{i.}^2 + (\sigma_{i.}^2 + \sigma_{ii}\sigma_{..})/(n - 1).$$

Summation of these two expected values over the appropriate indices and grouping the terms leads to the following expression:

$$E(B) = [n\sum_i \lambda_i^2 + (\sum_i \lambda_i)^2]/(n - 1). \quad (5)$$

From (4) and (5) it may be shown that

$$(\sum_i \lambda_i)^2 = \frac{(n - 1)}{(n + 1)(n - 2)} E[nA - 2B]$$

$$\sum_i \lambda_i^2 = \frac{(n - 1)}{(n + 1)(n - 2)} E[(n - 1)B - A]$$

Thus, an alternative estimator for  $\epsilon$ , as defined in (2), is

$$\tilde{\epsilon} = \frac{nA - 2B}{[k - 1][(n - 1)B - A]}$$

In terms of  $\hat{\epsilon}$ , the estimator is

$$\tilde{\epsilon} = \frac{n(k - 1)\hat{\epsilon} - 2}{[k - 1][n - 1 - (k - 1)\hat{\epsilon}]}$$

It may be verified that for any value of  $n$  and  $k$ ,  $\tilde{\epsilon} \geq \hat{\epsilon}$  with the equality holding when  $\hat{\epsilon} = 1/(k - 1)$ . Since it would not be reasonable to estimate  $\epsilon$  by a quantity larger than 1.0, whenever  $\tilde{\epsilon}$  exceeds this upper bound the estimator is equated to 1.0. It may also be noted that the difference between  $\tilde{\epsilon}$  and  $\hat{\epsilon}$  decreases with increasing  $n$ .

The foregoing development may be generalized to the split-plot design. This ANOVA model is appropriate when the researcher is interested in evaluating the similarity of profiles of average scores for several populations. For example, studies of the differences between boys and girls in their achievement, aptitude, or interest profiles would employ this design. On the assumption that the  $g$  independent populations have identical variance-covariance matrices, each element is estimated by pooling the corresponding sums of squared deviations or cross-products from the several groups. The statistic  $\hat{\epsilon}$  is then computed as before. With



g groups and a total of N subjects, the statistic  $\tilde{\epsilon}$  is obtained by the formula

$$\tilde{\epsilon} = \frac{N(k-1)\hat{\epsilon} - 2}{(k-1)[N-g-(k-1)\hat{\epsilon}]}$$

The same convention holds for the upper limit.

#### MONTE CARLO STUDY OF $\hat{\epsilon}$ AND $\tilde{\epsilon}$

Review of a number of test manuals suggests that the condition  $\epsilon > .750$  generally holds for the test batteries used in educational research. For example, the intercorrelation matrix for the original Primary Mental Abilities Tests (1938) has a value of  $\epsilon$  equal to .764. In their later factor studies of intelligence, the Thurstones (1941) reported a matrix for which  $\epsilon$  equals .852. For a five-variable matrix derived from the Wechsler Adult Intelligence Scale (1958) the value of  $\epsilon$  equals .752. Grade 6 of the Iowa Tests of Basic Skills (1974) has a dispersion matrix with  $\epsilon$  equal to .913. The manual for the Differential Aptitude Tests (1966) includes a variance-covariance matrix with  $\epsilon$  equal to .722 when the abilities are measured in stanine scores. The twelve subtests of the WISC-R (1974) have a matrix for which  $\epsilon$  equals .853.

On the basis of these well known test batteries and a sampling of research studies involving repeated measurements, five representative population matrices were devised with  $\epsilon \leq 1.0$ . For two matrices  $\epsilon < .750$ , for two others  $\epsilon > .750$ , and for the last  $\epsilon = 1.0$ . For convenience, all variables were assigned a population mean of zero and variance of 1.00. These matrices are presented in Table 1.

A computer program was devised to generate sample vectors of k scores, each vector simulating the score profile or repeated measurements of an experimental subject. These scores were generated from a k-variate normal distribution with one or another of the five population matrices in Table I. Each sample set of n profiles was analyzed via the randomized block model, the mean square ratio obtained, the estimators  $\hat{\epsilon}$  and  $\tilde{\epsilon}$  computed, and the approximate test carried out. The reference F distribution for each test was dictated by the product of  $\hat{\epsilon}$  or  $\tilde{\epsilon}$  times the conventional degrees of freedom. Eighteen thousand sets of sample data, leading to eighteen thousand significance tests by each approach, were generated for each combination of covariance matrix and block size (n = 10, 15, 20). Summary data for the estimators are presented in Table II, and empirical estimates of the incidence of Type I error are presented in Table III.

TABLE I

Covariance Matrices of Populations from which Samples Were Drawn in Monte Carlo Study of Statistics  $\hat{\epsilon}$  and  $\tilde{\epsilon}$

Value of $\epsilon$	Elements					
$\epsilon = .363$	1.0000					
	.8595	1.0000				
	.9602	.8575	1.0000			
	.6417	.8794	.6621	1.0000		
	.4368	.7705	.5952	.9107	1.0000	
						1.0000
$\epsilon = .522$	1.0000					
	.8000	1.0000				
	.6000	.8000	1.0000			
	.4000	.6000	.8000	1.0000		
	.3000	.4000	.6000	.8000	1.0000	
						1.0000
$\epsilon = .752$	1.0000					
	.8100	1.0000				
	.7400	.7000	1.0000			
	.5300	.5800	.5200	1.0000		
	.4300	.4500	.3900	.6100	1.0000	
						1.0000
$\epsilon = .831$	1.0000					
	.6200	1.0000				
	.6200	.6700	1.0000			
	.5400	.5300	.6200	1.0000		
	.2900	.3800	.4800	.6200	1.0000	
						1.0000
$\epsilon = 1.000$	1.0000					
	.5000	1.0000				
	.5000	.5000	1.0000			
	.5000	.5000	.5000	1.0000		
	.5000	.5000	.5000	.5000	1.0000	
						1.0000

In general,  $\hat{\epsilon}$  appears to be a better estimator than  $\tilde{\epsilon}$  (in terms of bias) when the parameter  $\epsilon$  is in the neighborhood of .5 or lower. However, when  $\epsilon$  is moderately large, say  $\epsilon > .75$ ,  $\tilde{\epsilon}$  is the less biased estimator. Increasing the block size results in a reduction of the bias for the two statistics. Whenever the bias is small, the improvement due to large block size is also small. In summary, it would appear that  $\tilde{\epsilon}$  is superior to  $\hat{\epsilon}$  in the range of  $\epsilon$  commonly found for batteries of standardized tests.

TABLE II

Estimated Means and Standard Deviations  
of the Sample Correction Factors ( $k = 5$ )

$\epsilon$	n	$\hat{\epsilon}$		$\tilde{\epsilon}$	
		Mean	S.D.	Mean	S.D.
.363	10	.363	.064	.418	.104
	15	.364	.055	.396	.074
	20	.361	.052	.384	.064
.522	10	.495	.097	.639	.170
	15	.515	.091	.609	.133
	20	.523	.091	.592	.120
.752	10	.596	.106	.809	.166
	15	.644	.101	.798	.146
	20	.666	.107	.766	.141
.831	10	.632	.105	.861	.150
	15	.689	.101	.858	.136
	20	.718	.107	.853	.136
1.000	10	.690	.093	.937	.102
	15	.770	.084	.953	.085
	20	.814	.091	.959	.097

What sort of power advantage accrues from the use of  $\tilde{\epsilon}$  rather than  $\hat{\epsilon}$  when  $1.0 \geq \epsilon \geq .75$ ? The answer to this question depends upon the degree of falsity in the null hypothesis, the value of  $\epsilon$ , the significance level and the sample size. The advantage of  $\tilde{\epsilon}$  is greatest when  $\epsilon$  is close to 1.0 and  $n$  is small. For purposes of illustration, we may take  $\epsilon = 1.0$ ,  $n = 10$ ,  $\alpha = .05$ , and  $k = 5$  (a five-treatment experiment). In this situation, the means of  $\hat{\epsilon}$  and  $\tilde{\epsilon}$  in the Monte Carlo portion of this study were .690 and .937, respectively. If these values are assumed representative, the use of  $\hat{\epsilon}$  typically leads to the use of a central F model with approximately  $(.690)(4) = 2.8$  and  $(.690)(4)(9) = 24.8$  degrees of freedom. The use of  $\tilde{\epsilon}$  leads to an F with 3.7 and 33.7 degrees of freedom.

The power of tests based on the two procedures depends on  $\phi^2$ , the non-centrality parameter of F. If  $\phi^2 = 2.8$ , the

TABLE III

Empirical Percents of Type I Error Associated with the  
Approximate Tests with Estimated Degrees of Freedom  
in Randomized Block Designs ( $k = 5$ )

		Correction Factor $\hat{\epsilon}$				Correction Factor $\tilde{\epsilon}$			
		Level of Significance (%)				Level of Significance			
$\epsilon$	$n$	10	5	2.5	1	10	5	2.5	1
.363	10	9.6	5.2	2.7	1.2	10.5	6.0	3.6	1.8
	15	9.6	5.1	2.9	1.2	10.1	5.4	3.1	1.4
	20	9.8	5.4	2.9	1.2	10.1	5.5	3.2	1.5
.522	10	8.6	4.4	2.3	1.1	10.2	5.7	3.5	1.8
	15	9.1	4.5	2.3	1.0	10.2	5.6	3.0	1.4
	20	9.5	5.1	2.4	1.2	10.3	5.5	3.1	1.5
.752	10	8.0	3.4	1.7	0.7	10.2	5.5	3.0	1.3
	15	8.2	3.8	1.9	0.8	9.6	5.1	2.7	1.3
	20	9.4	4.4	2.3	1.1	10.2	5.1	2.9	1.4
.831	10	7.8	3.6	1.7	0.6	10.1	5.3	2.9	1.3
	15	8.5	4.0	2.0	0.9	10.1	5.3	2.8	1.4
	20	9.1	4.6	2.1	0.8	10.3	5.4	2.8	1.2
1.000	10	7.1	2.9	1.2	0.3	9.5	4.6	2.3	0.9
	15	8.1	3.4	1.4	0.5	9.8	4.7	2.3	0.9
	20	9.3	4.2	1.8	0.6	10.5	5.0	2.5	1.0

power values associated with  $\hat{\epsilon}$  and  $\tilde{\epsilon}$  are approximately .74 and .81, respectively. The difference is not large, but it is achieved at practically no expense, and  $\hat{\epsilon}$  better controls Type I error. It is of interest to note that with this same degree of falsity and  $n = 11$ , Hotelling's  $T^2$  has a power of .57 (Huynh, 1969). When  $\epsilon = .831$  and the observed average values of  $\hat{\epsilon}$  and  $\tilde{\epsilon}$  are used, the power differential is .71 vs .78. With smaller values of  $n$ , the differences in power would be larger; with smaller values of  $\epsilon$  the differences would be smaller. If one preferred to summarize the results of an experiment in terms of confidence intervals for the contrasts of interest or with Bayesian credibility

intervals, the use of  $\tilde{\epsilon}$  rather than  $\hat{\epsilon}$  would result in analogous reductions in the length of the intervals when  $\epsilon > .75$ .

The characteristics and merits of the statistics  $\hat{\epsilon}$  and  $\tilde{\epsilon}$  manifest themselves again clearly in the estimated Type I error probabilities of the approximate tests based on each. From the data in Table III, it can be seen that the test based on  $\hat{\epsilon}$  is more satisfactory when the parameter  $\epsilon$  is relatively low or when the number of blocks or subjects is fairly large. The test based on  $\tilde{\epsilon}$ , on the other hand, behaves very well at the nominal 10 or 5 percent level in all of the situations considered. At the nominal 2.5 and 1 percent levels it gives somewhat more relaxed, but reasonably adequate, control over Type I error whenever the covariance matrix is not extremely heterogeneous. This test is less dependent on the number of blocks, and is fairly good even with a block size as small as twice the number of treatment levels. Of course, if  $n > 20$  and powerful computer facilities are available, the Hotelling  $T^2$  gives comparable power to this approximate test and more precise control of Type I error.

UPPER PERCENTILE OF THE F DISTRIBUTION WITH  
FRACTIONAL DEGREES OF FREEDOM

To carry out the approximate tests discussed above, it is necessary to determine the 100 (1 -  $\alpha$ ) percentile of the F distribution with  $\underline{h}$  and  $\underline{mh}$  degrees of freedom. Since  $\underline{h}$  and  $\underline{mh}$  are determined by multiplying the traditional degrees of freedom by  $\hat{\epsilon}$  or  $\tilde{\epsilon}$ , both  $\underline{h}$  and  $\underline{mh}$  may be fractional. (Actually, the quantity  $\underline{h}$ , which equals the product of  $\hat{\epsilon}$  or  $\tilde{\epsilon}$  times  $k - 1$ , is potentially fractional while  $\underline{m}$ , which equals  $n - 1$  or  $N - g$ , is integral.) When  $\underline{h}$  and  $\underline{mh}$  are small, interpolation between integral values of  $\underline{h}$  is rather untrustworthy. It is then mandatory to determine these values more directly and precisely. Imhof (1962) computed these values for  $\alpha = .05$  and  $.01$  and for  $\underline{m} = 2(1), 8$  and  $\underline{h} = 1(.05), 1.5; 1.5(.10)$  and  $3.0; 3.0(.2), 5.8$ . For the present study his tables were extended to two additional levels of the nominal probability, namely,  $\alpha = .10$  and  $.025$ . Researchers who may have use for these tables may obtain copies from the second author.

REFERENCES

- Anderson, T. W. An introduction to multivariate statistical analysis. New York: John Wiley, 1966.
- Bennett, G. K., Seashore, H. G., & Wesman, A. G. Differential Aptitude Tests, (4th ed.). New York: Psychological Corporation, 1966.
- Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effect of inequality of variance and correlation of errors in the two-way classification. Annals of Mathematical Statistics, 1954, 25, 484-498.
- Collier, R. O., Baker, F. B., Mandeville, G. K., & Hayes, T. F. Estimates of test size for several test procedures based on conventional variance ratios in the repeated measures design. Psychometrika, 1967, 32, 339-353.
- Geisser, S., & Greenhouse, S. Extension of Box's results on the use of the F distribution in multivariate analysis. Annals of Mathematical Statistics, 1958, 29, 885-891.
- Greenhouse, S. W., & Geisser, S. On methods in the analysis of profile data. Psychometrika, 1959, 24, 95-112.
- Hajek, J. Inequalities for the generalized student's distribution and their applications. Selected Translations in Mathematical Statistics and Probabilities, Institute of Mathematical Statistics, 1962, 2, 63-74.
- Hieronymus, A. N., & Lindquist, E. F. Iowa Tests of Basic Skills, Forms 5 & 6. Boston: Houghton Mifflin, 1974.
- Hotelling, H. The generalization of Student's ratio. Annals of Mathematical Statistics, 1931, 2, 360-378.
- Huynh, H. Effect of heterogeneity of covariance on the levels of significance of certain proposed tests for the treatments by subjects and type I designs. Unpublished doctoral dissertation, University of Iowa, 1969.
- Huynh, H., & Feldt, L. S. Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. Journal of the American Statistical Association, 1970, 65, 1582-1589.

- Imhof, J. P. Testing the hypothesis of fixed main-effects in Scheffé's mixed model. Annals of Mathematical Statistics, 1962, 33, 1086-1095.
- Lesser, G. S., Fifer, G., & Clark, D. H. Mental abilities of children from different social-class and cultural groups. Monographs of the Society for Research in Child Development, 1965, 30 (4, Serial No. 102).
- Stoloff, P. H. Correcting for heterogeneity of covariance for repeated measures designs of the analysis of variance. Educational and Psychological Measurement, 1970, 30, 909-924.
- Thurstone, L. L., & Thurstone, T. G. Primary mental abilities, Psychometric Monograph, No. 1. Chicago: University of Chicago Press, 1938.
- Thurstone, L. L., & Thurstone, T. G. Factorial studies of intelligence, Psychometric Monograph, No. 2. Chicago: University of Chicago Press, 1941.
- Wechsler, D. The measurement and appraisal of adult intelligence. Baltimore: Williams and Wilkins, 1958.
- Wechsler, D. Wechsler Intelligence Scale for Children - revised. New York: Psychological Corporation, 1974.
- Winer, B. Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill, 1971.

#### AUTHORS

HUYNH, HUYNH Address: School of Education, Boston College, Chestnut Hill, MA 02167. Title: Assistant professor. Degrees: B.A. University of Saigon, M.A. George Peabody College, Ph.D. University of Iowa. Specialization: Educational measurement and statistical methods.

FELDT, LEONARD S. Address: Lindquist Center, University of Iowa, Iowa City, IA 52242. Title: Professor. Degrees: B.Sc., M.A. Rutgers University, Ph.D. University of Iowa. Specialization: Educational measurement and statistical methods.

*[Manuscript received April, 1975; revised September, 1975.]*