DOCUMENT RESUME

ED 154 003 TM 006 999

AUTHOR Huynh, Huynh
TITLE Estimation of the KR20 Reliability Coefficient When Data Are Incomplete.
PUB DATE Apr 77
NOTE 21p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Analysis of Variance; *Comparative Statistics; *Mathematical Models; Statistical Analysis; *Test Reliability
IDENTIFIERS *Estimation; *Kuder Richardson Formula 20; Monte Carlo Methods

ABSTRACT

Three techniques for estimating Kuder Richardson reliability (KR20) coefficients for incomplete data are contrasted. The methods are: (1) Henderson's Method 1 (analysis of variance, or ANOVA); (2) Henderson's Method 3 (FITCO); and (3) Koch's method of symmetric sums (SYSUM). A Monte Carlo simulation was used to assess the precision of the three reliability estimation procedures. The ANOVA method was judged to be the most useful of the three methods. Similar results were obtained by random sampling with real data. (CTM)

ESTIMATION OF THE KR20 RELIABILITY COEFFICIENT WHEN DATA ARE INCOMPLETE

Huynh Huynh
University of South Carolina

Abstract

Attention is focussed on the estimation of the KR20 reliability coefficient when data are incomplete. In terms of point estimate, it is found that the analysis of variance procedure (Henderson's Method 1) would be most suitable among the three techniques under study. Confidence coefficients are provided for the KR20 when data are incomplete.

April 1977

A paper presented at the annual meeting of the American Educational Research Association, New York, April 4-8, 1977.

ESTIMATION OF THE K20 RELIABILITY COEFFICIENT WHEN DATA ARE INCOMPLETE

1. Introduction

The calculation of a reliability of a test such as the Kuder-Richardson coefficient twenty (KR20) or its more genral counterpart alpha [Cronbach, 1951] frequently assumes that data are complete. In other words each examinee is expected to respond to every item. Yet data completion is rarely realized in many situations. For known or untold reasons, many students choose to skip a difficult item rather than to contemplate a wild or educated guess when penalties are imposed for theincorrect responses. If the proportion of missing responses is fairly small, one may estimate these responses and then apply an appropriate formula to the complete set of data. When testing is construed as the realization of a two-way factorial design [Lord, 1955; Kristof, 1963, 1970; Feldt, 1965], various techniques described in such texts as Cochran and Cox [1966] or Winer [1971] may be called upon. The extent to which the estimated responses would bias the computed reliability, however, does not seem to have been fully explored. Hence it appears desirable to base reliability computations directly on the available data and to disregard totally the missing responses.

The present study aims at the exploration of several estimation procedures for test reliability when data are incomplete. Attention will be focussed on the KR20 index (e.g. with test items scored as 0 or 1). Conceivably the results would be expected to hold for the more general index alpha since the 0 - 1 scoring presents the most serious violation on the assumptions of the model used in the study. It will be assumed on the part of the reader familiarity with the content of Chapter 7 of Lord and Novick (1968), the linear model [Searl, 1971] and the work of Feldt [1965] on modeling testing as a random effect, two-way factorial design.

2. Description of three Estimation Procedures for the KR20 Coefficient

It will now be assumed that the response of subject i ($i = 1,\ldots, a$) on item j ($j = 1,\ldots, b$) may be represented as $y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$, where μ is a constant, $\alpha_i \sim L(0,\sigma_\alpha^2)$, $\beta_j \sim L(0,\sigma_\beta^2)$ and $e_{ij} \sim L(0,\sigma_e^2)$ and all random variables are independent. The notation $L(0, \sigma^2)$ represents any distribution with mean 0 and variance σ^2. It is not required that this distribution is normal. Given the various parameters as defined, the population reliability index is $\rho_{20} = b\sigma_\alpha^2 / (b\sigma_\alpha^2 + \sigma_e^2)$.

The estimation of ρ_{20} on the basis of sample test data may be somewhat facilitated by considering the estimation of σ_α^2 and σ_e^2 separately. Thus a reasonable estimate r_{20} of ρ_{20} may be arrived at by replacing σ_α^2 and σ_e^2 respectively by two suitably chosen unbiased estimates s_α^2 and s_e^2. This in no way guarantees that r_{20} would average out at ρ_{20}, however unbiasedness should be expected to hold at least asymptotically under mild regularity conditions. Under normality, it is known [Kristof, 1970] that an unbiased estimate for the population reliability may be given by a linear function of the sample reliability. Granting that the pattern of missing data is fixed from sample to sample, there are at least three ways to estimate both σ_α^2 and σ_e^2. All methods yield unbiased statistics for σ_α^2 and σ_e^2, and reduce to the traditional variance component estimation when data are complete.

The first two procedures, namely the analysis of variance (ANOVA) and fitting constants (FITCO) techniques are due to Henderson [1953]. They are also referred to as Henderson's Method 1 and Henderson's Method 3. Details about these methods along with their rationale may be found in Searl [1971, Chapter 10]. Basically, ANOVA attempts to extend the formulation of the sum of squares in a balanced design to the case of unbalanced data. Thus three sources of variation are identified, namely the total variation SS_{tot}, the row (subject) variation SS_A and the (item) column variation SS_B [See Table 1 for the computational procedures.] The unbiased estimates s_α^2 and s_e^2 will be found among the linear combinations of these variations. Henderson's Method 3 (FITCO), on the other hand, focuses on the equation $y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$ as a linear model, considering (in this study) the row variation SS_A as the sum of squares explained by the α_i's after being adjusted for the effects of the β_j's and μ [See Table 2]. Unbiased estimates for σ_α^2 and σ_e^2 may be found as previously described.

The third procedure, due to Koch [1968], is called the symmetric sums (SYSUM) method. It is based on the symmetric sums of squared differences of the form $(y_{ij} - y_{i'j'})^2$. Three symmetric sums [refered to as h_A, h_B and h_{AB} in Table 1] are there linearly combined to form the unbiased estimates s_α^2 and s_e^2.

Tables 1 and 2

It may be noted that both procedures ANOVA and SYSUM are computationally simple and can be implemented easily by hand. The FITCO method is more complex, involving the inversion of an usually large matrix (of order of b - 1). Of course this should not present any problem where computer facilities are available.

To provide the readers with some feeling about the three described methods, a numerital example is presented in Table 3.

Table 3

---

3. Logic of the Simulations

When normality is assumed for the distributions described early in Section 2, and when the pattern of missing data is fixed in advance, the sampling variances of s_α^2 and s_e^2 may be computed [Searl, 1971]. Bush and Anderson [1963] studied several cases of the two-way design with planned unbalancedness. The general trend which may be deduced from their results is that when σ_e^2 is much larger than σ_α^2, the ANOVA estimates have smaller variance than the FITCO estimates, but otherwise the FITCO estimates are less variable. Thus in estimating test reliability ρ_{20} with fixed (hence, known) pattern of missing data, it seems that the ANOVA procedure would provide a more stable estimate r_{20} than the FITCO does when ρ_{20} is fairly low. Conversely, estimates based on the FITCO technique would show less fluctuation than those deduced from the ANOVA when ρ_{20} is high.

In mental testing, it seems likely that the pattern of missing responses is function of the ability level of the examinees. When guessing is not a major factor, it would be reasonable to assume for example that examinees with lower ability will leave more items untouched than examinees with higher ability. For these situations the variance calculations displayed in Searl [1971, Chapter 11] and the findings of Bush and Anderson as stated would not automatically hold.

It may be noted that the model $y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$ implies that each response y_{ij} takes at least four separate values. The lower bound is occasioned when all random variables α_i, β_j, and e_{ij} have positive probabilities at only two identical points. The estimation procedures previously described for ρ_{20} are therefore applicable as long as there are at least four response categories for each test item. This requirement is satisfied for data collected from scales of the Likert type. The 0 - 1 scoring procedure, on the other hand, clearly violates the functional form $y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$. Some data reported in Feldt [1965], however, indicate that this linear model may still serve as an aproximation for reliability studies when the set of 0 - 1 responses is complete.

Even under normality any theoretical computation of the sampling variance of each of the three estimates r_{20} would involve awful efforts when the pattern of missing data is let to vary from sample to sample. With 0 - 1 data such a computation may not be possible unless of course one is prepared to rely on an exhaustive enumeration Given the computer time constraint on the part of the author, it seems appropriate to resort to Monte Carlo simulations to compare the performance of the ANOVA, FITCO and SYSUM procedures. The next section will present the design of the first simulation based on artificial data.

4. Design of Simulation Study 1

For reason of mathematical tractability, the beta-binomial model was chosen to generate the test responses in the first simulation. This model assumes that (i) each examinee with true ability θ responds independently to the items with θ being the probability of obtaining a correct answer; and

(ii) the true ability θ for the population of examinees follows a beta density of the form

$$p(\theta) = \frac{1}{B(u,v)} \theta^{u-1} (1-\theta)^{v-1}$$

where $u > o$ and $v > o$.

With the beta-binomial framework so stated, the population reliability for a test consisting of b items is known to be $\rho_{20} = b/(b + u + v)$.

It seems reasonable to postulate that the proportion of missing response is a nonincreasing function of the ability of the examinee. (One would assume of course that guessing is not a serious factor.) A missing response function $m(\theta)$ will have to be specified. Though the set of nonincreasing functions is infinite, a linear form was chosen for $m(\theta)$ for reason of simplicity. In other words, the proportion of missing responses is

$$m(\theta) = c\theta + d$$

where c and d are two suitably chosen constants. Over the total population of examinees, the expected value for $m(\theta)$ may be noted to be

$$E\, m(\theta) = cu/\ (u + v) + d.$$

If p denotes the overall proportion of missing responses, then

$$cu/\ (u + v) + d = p.$$

To get a second equation so that c and d may be solved, it was assumed that examinees with highest ability ($\theta = 1$) would not miss any item. This implied that $c + d = 0$, and hence

$$m(\theta) = p\ (u + v)\ (1 - \theta)/v.$$

With b being the number of items, an examinee with true ability θ would be expected to skip $bm(\theta)$ items. Of course this number had to be rounded to the nearest integer in the study. It may be noted that though only one thousand samples were generated in this study, the overall observed proportion of missing responses is almost identical to the posted value p.

As the first step in the simulation process, the percentile points of the true ability distribution were computed via the IBM subroutine BDTR [1971]. The IBM subrouting RANDU [1971] was then used to (i) generate a random sample of a true ability values, (ii) generate at each true ability θ the b 0 - 1 responses, and (iii) generate at each true ability θ the pattern of the $bm(\theta)$ missing responses. These three steps generated an incomplete matrix of responses. The three estimation procedures ANOVA, FITCO and SYSUM were then applied to obtain the sample r_{20}'s. The whole process was repeated one thousand times to estimate the means $\bar{r}_{20}$'s and the corresponding mean square errors (MSE) which were used to compare the performance of the estimation procedures under a variety of situations.

5. Results of Simulation Study 1

Table 4 reports the means and mean square errors of the three sample estimates r_{20}'s for eighteen situations with a = 40 and b = 20. The population reliability was chosen to be ρ_{20} = .5 (low), .7 (moderate) and .9 (high). For each reliability level, three true ability distributions were selected to represent several levels of skewness γ. Finally, except for one case, the proportion of missing data was chosen as 20% (moderate) and 40% (high).

The following trends may be inferred from the data on Table 4:

(a) Under all situations under consideration, both the ANOVA and FITCO provide estimates which behave almost identically in terms of bias and mean square error.

(b) For symmetric (true ability or test score) distributions, the three methods are about equally efficient. Though the SYSUM estimate tends to show more bias or more variability than the other two estimates, the discrepancies are not substantial.

(c) For skewed distributions, the SYSUM estimate tends to display more pronounced bias or variability than the other two estimates.

(d) Other things being equal, the means of both the ANOVA and FITCO estimates relate positively to the skewness level of the true ability or test score distributions. The reverse trend holds for the SYSUM estimate.

(e) As expected, the three estimates perform better when the proportion of missing responses is smaller.

Table 4

These findings indicate clearly that for all practical purposes, the ANOVA estimate is the most suitable in dealing with incomplete data, at least with those conforming to the beta-binomial model. This estimate is fairly simple to compute and tends to show less bias and less variability than the FITCO (a computational nightmare) and SYSUM estimates in most situations.

6. Simulation Study 2

While the beta-binomial model was chosen mainly for its mathematical tractability, most real life test data rarely conform to any simple distribution. It would seem desirable to replicate some of the conclusions of Section 5 using some data of this type.

A second simulation was conducted via random sampling (without replacement) from three data sets. (i) Data 1 consist of the responses of 582 examinees to a 138 item test. The test score distribution has mean 76.75, standard deviation 27.09, skewness .259 and reliability .964. Using the Spearman-Brown formula for projection, it was found that a test consisting of k = 20 items of Data 1 would have a KR20 index of .796. (ii) Data 2 is a subset of Data 1, consisting of the responses of 428 examinees. The test score distribution has mean 63.67, standard deviation 17.77, skewness .096 and reliability .909. The projected reliability of a 20-item test is ρ_{20} = .590. (iii) Finally Data 3 is a subset of Data 2 with 314 examinees. The test score distribution has mean 55.23, standard deviation 12.17, skewness -.271 and reliability .805. The projected reliability for a 20-item test is ρ_{20} = .375.

For the present simulation, the missing response function was chosen to be a decreasing linear function of the test score obtained from the 138-item test for each examinee. As in Section 4, it was assumed that examinees with a score of 138 would not miss any item.

The results of the second simulation study are reported in Table 5. They indicate clearly that the ANOVA and FITCO estimates behave almost identical in terms of bias and mean square error in all circumstances under investigation. The SYSUM estimate tends to show more bias and/or more variability than the other two estimates. These observations reinforce the major conclusion of the first simulation study that the ANOVA procedure seems to be most suitable in dealing with incomplete data.

Table 5
----

7. Confidence Interval for the KR20 Reliability Based on Incomplete Data

Under normality, confidence intervals for the KR20 reliability may be established when data are complete [Feldt, 1965]. An extension will be made to the case of incomplete data.

Referring to the notations of Table 1 and 2 it may be noted that under normality, the various sums of squares are distributed as chi-squares up to a multiplicative constant. More specifically, it is well known that (i) the adjusted sum of squares due to A, $SS_A = R(\mu, \alpha, \beta) - T_B$ is distributed as $[h\,\sigma_\alpha^2 + (a - 1)\sigma_e^2]\,\chi^2(a - 1)$; (ii) the error sum of squares $SS_E = T_0 - R(\mu, \alpha, \beta)$ is distributed as $f\sigma_e^2\chi^2(f)$; and (iii) the two chi-squares are independent. It follows that the ratio

$$\lambda = f\,SS_A/SS_E$$

is distributed as

$$[h\,\frac{\sigma_\alpha^2}{\sigma_e^2} + a - 1]\;F(a - 1, f)$$

where F represents an F variable with listed degrees of freedom. Algerbraic manipulations yield

$$\lambda = h\,\frac{s_\alpha^2}{s_e^2} + a - 1$$

$$= h\,\frac{r_{20}}{b(1 - r_{20})} + a - 1$$

where s_α^2, s_e^2 and hence r_{20} are computed via the FITCO procedure. Let $100(1 - \alpha)\%$

be given confidence level and let

$$L_1 = F_{\frac{\alpha}{2}}(a-1,f) = F_{1-\frac{\alpha}{2}}(f, a-1)$$

and

$$L_2 = F_{1-\frac{\alpha}{2}}(a-1,f).$$

Then a $100(1-\alpha)\%$ confidence interval for the ratio $\sigma_\alpha^2/\sigma_e^2$ is given by the two limits

$$\ell_1 = (\lambda/L_2 - a + 1)/h$$

and

$$\ell_2 = (\lambda/L_1 - a + 1)/h .$$

Since $\rho_{20} = b\sigma_\alpha^2/(b\sigma_\alpha^2 + \sigma_e^2)$ is monotonically increasing with $\sigma_\alpha^2/\sigma_e^2$, the corresponding $100(1-\alpha)\%$ confidence interval of ρ_{20} is given by (ρ_1, ρ_2) where

$$\rho_1 = b\ell_1/(b\ell_1 + 1)$$

and

$$\rho_2 = b\ell_2/(b\ell_2 + 1)$$

<u>Numerical example</u>. Granting that the normal theory can be applied to the data of Table 3, an 80% confidence interval was found to be (.029 - .629) for $\sigma_\alpha^2/\sigma_e^2$. The corresponding 80% confidence interval for ρ_{20} was deduced as (.146 - .791).

To conclude this section, the following remarks may be made.

(a) The data compiled by Feldt [1965] and those of Table 4 with $\gamma = 0$ indicate that the normal model holds up fairly well for symmetric distributions based on 0 - 1 responses. One should hope at least that the procedures previously described would provide an approximate confidence interval for ρ_{20} when the test score distribution is not badly skewed. The Jacknife procedure as implemented by Pandley and Hubert [1975] may be adapted to the case of incomplete data if normality appears to be a bothering assumption.

(b) The data simulated in this study clearly show that the estimates based on ANOVA and FITCQ are almost identical in terms of mean and mean square error. This indicates that in case of need, one may go ahead and use the ANOVA r_{20} to establish a confidence interval for ρ_{20}. This would greatly simplify the computations.

8. Conclusion

Attention was focused in the study on the estimation of the KR20 reliability coefficient when data are incomplete. In terms of point estimation, it was found that the analysis of variance procedure would be most suitable among the three techniques under study. Other techniques are of course available. But by and large they do not seem to be easily implemented for test data. One of them is the minimum normed quadratic unbiased estimation (MINQUE) proposed by Rao [1973; also other references listed in this text]. Under normality, MINQUE provides locally minimum variance unbiased quadratic estimates for both σ_α^2 and σ_e^2. Unfortunately MINQUE at the present time requires inversion of large matrices and therefore may not be computationally justified in most testing situations.

Finally confidence intervals based on normality are provided. The mass of data on the effect of nonnormality on the F test and the forces of the Central Limit Theorem should convince the readers that the procedure so described would provide at least an approximation for a number of testing situations.

Table 1

Computing Formulae for the ANOVA, and SYSUM Variance Estimates*

Model: $y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$,

$i = 1,\ldots,a$; $j = 1,\ldots,b$; with $n_{ij} = 0$ for empty cells and $n_{ij} = 1$ for non-empty cells; $N = \sum_i \sum_j n_{ij}$.

Analysis of variance estimators

Calculate

$$T_o = \sum_i \sum_j y_{ij}^2 \qquad T_\mu = y_{..}^2/N$$

$$T_A = \sum_i y_{i.}^2/n_{i.} \qquad T_B = \sum_j y_{.j}^2/n_{.j}$$

and

$$k_1' = \sum_i n_{i.}^2/N \qquad k_2' = \sum_j n_{.j}^2/N$$

$$\lambda_1 = (N - k_1')/(N - b) \quad \text{and} \quad \lambda_2 = (N - k_2')/(N - a).$$

Then

$$s_e^2 = \frac{\lambda_2(T_o - T_A) + \lambda_1(T_o - T_B) - (T_o - T_\mu)}{N - k_1' - k_2' + 1}$$

and

$$s_\alpha^2 = [T_o - T_B - (N - b)s_e^2]/(N - b).$$

Symetric sums estimators

Calculate $k_1 = Nk_1'$ and $k_2 = Nk_2'$

$$h_A = \sum_j (n_{.j} \sum_i y_{ij}^2 - y_{.j}^2)/(k_2 - N)$$

$$h_B = \sum_i (n_{i.} \sum_j y_{ij}^2 - y_{i.}^2)/(k_1 - N)$$

Table 1
(contd.)

and $$h_{AB} = [\sum_i \sum_j (N - n_{i.} - n_{.j} + n_{ij}) y_{ij}^2$$

$$-(y_{..}^2 - \sum_i y_{i..}^2 - \sum_j y_{.j}^2 + T_0) / (N^2 - k_1 - k_2 + N)$$

Then $$s_\alpha^2 = h_{AB} - h_B \quad \text{and} \quad s_e^2 = h_A + h_B - h_{AB}$$

---

* Adapted from Searle [1971, Chapter 11]. It should be noted that the two expressions for h_A and h_B on page 488 of Searle must be interchanged. The dot (.) notation refers to summation.

Table 2

Computing Formulae for the FITCO Variance Estimates*

Computation for $R(\mu, \alpha, \beta)$

Compute for j and $j' = 1,\ldots,b-1$

$$c_{jj} = n_{.j} - \sum_{i=1}^{a} n_{ij}^2/n_{i.}$$

$$c_{jj'} = - \sum_{i=1}^{a} n_{ij}n_{ij'}/n_{i.}\ ,\ j \neq j'$$

$$r_j = y_{.j} - \sum_{i=1}^{a} n_{ij}\bar{y}_{i.}$$

where $\bar{y}_{i.}$ is the mean of the ith row.

Let $C = \{c_{jj'}\}$ and $r = \{r_j\}$

Then $R(\mu, \alpha, \beta) = T_A + r'C^{-1}r.$

Fitting constants method estimators

Denote $h = N - b$
and $f = N - a - b + 1.$

Then

$$s_e^2 = [T_0 - R(\mu, \alpha, \beta)]/f$$

and

$$s_\alpha^2 = [R(\mu, \alpha, \beta) - T_B - (a-1)s_e^2]/h\ .$$

---

* Adapted from Searl [1971, Chapter 11].

Table 3

A Numerical Example of
the ANOVA, FITCO, and SYSUM Estimates

Examinee	1	2	3	4	5	6	7	8	9	10	11	12
Item												
1	1	1	X	0	1	1	0	0	1	X	1	X
2	1	0	1	X	X	1	X	X	1	0	1	0
3	1	1	1	X	0	1	1	1	0	0	1	X
4	1	1	1	1	1	1	X	1	1	0	1	0
5	0	1	0	0	0	0	0	1	X	X	1	0
6	1	1	[illegible]	1	x	1	0	1	0	0	1	1

Coding: 1 = correct response
0 = incorrect response
X = missing response

Reliability estimate
ANOVA = .615
FITCO = .547
SYSUM = .663

Table 4

Means (and Mean Square Errors) of the ANOVA, FITCO and SYSUM Reliability Estimates Based on Data Simulated from the Beta-Binomial Model

Population parameters		p	Means (MSE) of r_{20} based on ANOVA	FITCO	SYSUM
u = 6.000	γ = .386	20%	.533(.191)	.534(.191)	.508(.210)
v = 14.000	ρ_{20} = .500	40%	.549(.207)	.550(.208)	.489(.264)
u = 10.000	γ = .000	20%	.504(.148)	.504(.149)	.515(.153)
v = 10.000	ρ_{20} = .500	40%	.491(.196)	.493(.196)	.520(.210)
u = 14.000	γ = -.386	20%	.485(.150)	.485(.151)	.564(.130)
v = 6.000	ρ_{20} = .500	40%	.472(.180)	.471(.182)	.643(.128)
u = 2.571	γ = .519	20%	.710(.097)	.710(.097)	.687(.112)
v = 6.000	ρ_{20} = .700	40%	.708(.116)	.708(.118)	.637(.180)
u = 4.286	γ = .000	20%	.692(.082)	.692(.082)	.698(.084)
v = 4.286	ρ_{20} = .700	40%	.685(.107)	.685(.108)	.710(.112)
u = 6.000	γ = -.519	20%	.671(.090)	.670(.091)	.736(.071)
v = 2.571	ρ_{20} = .700	40%	.655(.111)	.653(.111)	.790(.062)
u = .667	γ = .141	20%	.898(.030)	.898(.030)	.888(.035)
v = 1.556	ρ_{20} = .900	40%	.898(.034)	.899(.034)	.875(.056)
u = 1.111	γ = .000	20%	.895(.027)	.895(.027)	.897(.085)
v = 1.111	ρ_{20} = .900	40%	.883(.032)	.883(.032)	.900(.031)
u = 1.556	γ = -.141	20%	.882(.032)	.881(.032)	.910(.024)
v = .667	ρ_{20} = .900	30%*	.862(.036)	.860(.036)	.913(.024)

* The simulation failed at p = 40% due to some negative values for $\hat{m}(\theta)$.

Table 5

Mean (and Mean Square Errors) of the ANOVA, FITCO and SYSUM Reliability Estimates Based on Random Samples from Three Sets of Real Data

Population Parameters			p	Means(MSE) of r_{20} based on		
No.	γ	ρ_{20}		ANOVA	FITCO	SYSUM
1.	.259	.796	20%	.795(.056)	.795(.057)	.815(.053)
			40%	.795(.062)	.793(.061)	.843(.048)
2.	.096	.590	20%	.562(.123)	.563(.123)	.553(.127)
			40%	.544(.159)	.547(.154)	.528(.171)
3.	-.271	.375	20%	.331(.176)	.329(.173)	.318(.176)
			40%	.321(.204)	.322(.202)	.295(.205)

References

Bush, N. and Anderson, R. L. A comparison of three different procedures for estimating variance components. Technometrics, 1963, 5, 421-440.

Cochran, W. G. and Cox, G. M. Experimental designs (Second edition). New York: John Wiley & Sons, 1966.

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

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

Henderson, C. R. Estimation of variance and covariance components. Biometrics, 1953, 9, 226-252.

IBM Application Program, System /360. Scientific subroutine package (360-CM-03X) Version III, Programmer's manual. White Plains, New York: IBM Corporation Technical Publication Department, 1971.

Koch, G. G. Some further remarks on "A general approach to the estimation of variance components." Technometrics, 1968, 10, 551-558.

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

Kristof, W. On the sampling theory of reliability estimation. Journal of Mathematical Psychology, 1970, 7, 371-377.

Lord, F. M. Sampling fluctuations resulting from the sampling of test items. Psychometrika, 1955, 20, 1-22.

References (contd.)

Lord, F. M. and Novick, M. R. Statistical theories of mental test scores. Reading, Mass.: Eddison-Wesley, 1968.

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

Rao, C. R. Linear statistical inference and its applications (Second edition). New York: John Wiley & Sons, 1973.

Searl, S. R. Linear models. New York: John Wiley & Sons, 1971.

Winer, B. J. Statistical principles in experimental designs (Second edition). New York: McGraw-Hill Book Co., 1971.