DOCUMENT RESUME

ED 049 276 TM 000 448

AUTHOR Sullins, Walter L.
TITLE The Effect of Sequential Dependence on the Sampling Distributions of KR-20, KR-21, and Split-Halves Reliabilities.
PUB DATE Feb 71
NOTE 9p.; Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, February 1971

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Comparative Analysis, Correlation, Item Analysis, Measurement Techniques, *Reliability, Sampling, *Standard Error of Measurement, *Statistical Analysis, *Test Construction, True Scores
IDENTIFIERS *Kuder Richardson Formulas, Monte Carlo Experiments

ABSTRACT

Five-hundred dichotomously scored response patterns were generated with sequentially independent (SI) items and 500 with dependent (SD) items for each of thirty-six combinations of sampling parameters (i.e., three test lengths, three sample sizes, and four item difficulty distributions). KR-20, KR-21, and Split-Half (S-H) reliabilities were computed for these 36,000 response matrices and cumulative distribution functions, each based on 500 trials, compiled. Comparison of the sampling distributions suggests that there is more similarity between the SI and SD cumulative distribution functions for KR-21 than KR-20, and either of these have more similarity than S-H. For KR-20, KR-21, and S-H, the similarity of sampling distributions under SI and SD varies directly with test length: the longer the test, the more similar the distributions. On the other hand, similarity varies inversely with sample size: the larger the sample, the more dissimilar the distributions. For both SI and SD item sets, the expected value of KR-21 is less than or equal to that of KR-20 which is less than or equal to the expected value of S-H. Also, for both item sets, the standard error of S-H is greater than that of KR-21 which is greater than or equal to the standard error of KR-20. The standard errors of the three reliability estimates vary inversely with sample size and test length, for both SI and SD items. In every case, SD increased the expected value of the estimate and decreased its standard error. (Author/AE)

The Effect of Sequential Dependence on the Sampling Distributions of KR20, KR21, and Split-Halves Reliabilities

Walter L. Sullins

Indiana State University
Terre Haute, Indiana

A Paper Presented to the Annual Meeting of the American Educational Research Association, New York

February 1971

INTRODUCTION

The general purpose of this study was to investigate the effects of sequential dependence (SD) on the sampling distributions of three commonly used reliability estimates -- Kuder-Richardson Formula 20 (KR20), Kuder-Richardson Formula 21 (KR21), and odd-even Split-Halves (S-H).

Two items are <u>sequentially independent</u> (SI) if and only if the success or nonsuccess on one does not affect the probability of success on the other. Symbolically, if S_i represents success on item i and S_i' indicates failure, items i and j are sequentially independent if and only if $P(S_i \mid S_j) = P(S_i)$ and $P(S_i \mid S_j') = P(S_i)$.

Two items are <u>sequentially dependent</u> (SD) if and only if the success or nonsuccess on one affects the probability of success on the other. Symbolically, items i and j are sequentially dependent if and only if $P(S_i \mid S_j) \neq P(S_i)$ and $P(S_i \mid S_j') \neq P(S_i)$.

In practice, it is often the case that achievement tests contain two or more items that are SD. The most obvious instance of SD is given by a pair of items, one of which refers to the other and the solution of the one is strictly contingent upon solution of the other. Frequently, however, the sequential dependence between items is more subtle than in this extreme case. Since SD does occur in practice and since the classical reliability estimates are often applied to these SD tests, it is important to investigate the effects of SD on the sampling distributions of the three commonly used estimates.

In the present study, computational forms of KR20 and KR21 (Kuder and Richardson, 1937) were used, and the S-H coefficients were obtained by applying the Spearman-Brown (Spearman, 1910; Brown, 1910) formula to the Pearson correlation between the scores made on the odd and even item sets.

METHODOLOGY

The following sampling parameters were used:

Test Length Tests containing 20, 30 and 50 items.

Sample Size Sample sizes of 20, 32, and 52.

Item Difficulty Distribution Four different distributions of item difficulty were used. Three of these distributions were constant difficulties of .3, .5, and .7; and the fourth was an equally-spaced rectangular distribution of the following form:

Let $i = 1, 2, 3, \ldots, k$ be the item index over the k items.

The difficulty for item i is $\frac{i}{k+1}$.

In this distribution, the items range in difficulty from $\frac{1}{k+1}$ to $\frac{k}{k+1}$.

True Score Distribution For each of the four item difficulty distributions, a symmetric unimodal distribution of true scores was forced in the following way:

Let D_i be the difficulty of item i, and m_{1s}, m_{2s} and m_{3s} be selected so that:

$D_i = m_{1s} (.25) D_i + m_{2s} (.50) D_i + m_{3s} (.25) D_i$.

In this notation, the subscript s is an index over the n subjects, and the following relationships hold among the weights:

$m_{11} = m_{12} = \cdots = m_{1d}$, where $d = n/4$

$m_{2(d+1)} = m_{2(d+2)} = \cdots = m_{2(3d-1)}$

$m_{3(3d)} = m_{3(3d+1)} = m_{3n}$

These conditions on the weights (m_1, m_2, m_3) give a three-point distribution of true scores for each parameter combination. The particular weights used in the study are presented in Table 1.

TABLE 1

WEIGHTS USED TO FORM TRUE SCORE DISTRIBUTION

Difficulty Distribution	m_{1s}	m_{2s}	m_{3s}
Rectangular	.5	1.0	1.5
Constant .3	.5	1.0	1.5
Constant .5	.5	1.0	1.5
Constant .7	.7	1.0	1.3

From the preceding discussion, the probability of subject s correctly answering item i is $m_{1s} D_i$ if $1 \leq s \leq n/4$, $m_{2s} D_i$ if $n/4 < s < 3n/4$, and $m_{3s} D_i$ if $3n/4 \leq s \leq n$. In this discussion the probability of success on an item does not depend on performance on prior items, but only on the relative position of the subject's true score. Thus, the items which comprise the subject's test score are SI.

SD among items was induced by changing the probability of success on a given item according to the subjects performance on the immediately preceding item. Performance on item 1 is unaffected by SD.

In the case of the equally-spaced rectangular distribution of item difficulties, $P(S_i) = \frac{i}{k+1}$. The following dependence probabilities were used:

$$P(S_i \mid S_{i-1}) = \frac{i+1}{k+1}$$

$$P(S_i \mid S'_{i-1}) = \frac{(k+1)i-(i^2-1)}{(k+1)\ (k-i+2)}$$

If D is the constant difficulty for each of the three constant item difficulty distributions, the dependence probabilities are:

$$P(S_i \mid S_{i-1}) = \frac{Dk+1}{k}$$

$$P(S_i \mid S'_{i-1}) = \frac{D(k-Dk-1)}{k-Dk} \ .$$

It may be easily shown that these dependence probabilities satisfy the following necessary condition:

$$n_{r(i)} = n.P(S_i) = n_{r(i-1)} \cdot P(S_i \mid S_{i-1}) +$$
$$n_{w(i-1)} \cdot P(S_i \mid S'_{i-1}) \ ,$$

where $n_{r(i)}$ is the number correctly answering item i, n is sample size, $n_{r(i-1)}$ is the number correctly answering item i-1, and $n_{w(i-1)}$ is the number incorrectly answering item i-1.

The procedures of the study required generation of nine sequences of pseudo-random numbers in the interval (0, 1). The sequences ranged from 200,000 to 1,300,000 in length. The multiplicative congruence method (Lehmer, 1951) was employed, using Pike and Hill's (1965) algorithm, to generate the sequences.

Subject x Item response matrices were generated in an element-wise manner by comparing the probability of success on each item by each subject with a pseudo-random number. The matrices were generated in pairs -- one matrix containing SI items and the other containing SD items. The same random number was used for comparison in each matrix for a given position.

Let XI be the n x k matrix of responses for the SI items and XD be the n x k matrix of responses for the SD items. Also, let XI_{si} and XD_{si} denote elements of these response matrices. Then, using R_{si} for the random number associated with position si, and S_{si} for success on item i by subject s, the matrices were generated in the following way:

If $R_{si} < P(S_{si})$, then $XI_{si} = 1$

If $R_{si} \geq P(S_{si})$, then $XI_{si} = 0$

If $R_{si} < P(S_{si} \mid S_{si-1})$, then $XD_{si} = 1$; otherwise $XD_{si} = 0$

If $R_{si} \geq P(S_{si} \mid S'_{si-1})$, then $XD_{si} = 0$; otherwise $XD_{si} = 1$

RESULTS AND CONCLUSIONS

These Monte Carlo experiments were conducted on an IBM 7094 computer using a Fortran Program written by the researcher (copies of the program are available upon request). Five-hundred dichotomously scored response patterns were generated with SI items and 500 with SD items for each of the thirty-six combinations of sampling parameters (i.e., three test lengths, three sample sizes, and four item difficulty distributions). KR20, KR21, and S-H were computed for each of these 36,000 response matrices and their cumulative distribution functions, each based on 500 trials, compiled.

The Kolmogorov-Smirnov two-sample procedure (Siegel, 1956) was used to compare the sampling distributions of the three estimates under SI and SD. Kolmogorov-Smirnov D values were obtained by computing the absolute value of the maximum vertical distance between the cumulative distribution functions. Considering all parameter combinations, the mean absolute D values for the three estimates were:

KR20	KR21	S-H
.307	.303	.333

Since Kolmogorov-Smirnov D values vary inversely with distribution similarity, it may be concluded that there is more similarity between the SI and SD cumulative distribution functions for KR21 than KR20, and either of these have more similarity than S-H.

Mean absolute D values were also computed for the SI and SD cumulative distribution functions under the various sampling conditions. These means are presented in Table 2.

For KR20, KR21, and S-H, the similarity of their sampling distributions under SI and SD varies directly with test length, i.e., the longer the test, the more similar are the sampling distributions. On the other hand, the similarity of the sampling distributions of the three estimates under SI and SD varies inversely with sample size, i.e., the larger the sample size, the more dissimilar are the sampling distributions. The different item difficulty distributions effect different degrees of similarity between SI and SD distributions. For both KR20 and KR21, the order of similarity of the sampling distributions under SI and SD is: constant .3, equally-spaced rectangular, constant .5, and constant .7. For S-H, the order is: equally-spaced rectangular, constant .3, constant .5, and constant .7.

For both SI and SD item sets, the expected value of KR21 is less than or equal to that of KR20 which is less than or equal to the expected value

TABLE 2

MEAN KOLMOGOROV-SMIRNOV ABSOLUTE D VALUES
UNDER THE VARIOUS MONTE CARLO CONDITIONS

Difficulty Distribution	KR20	KR21	S-H
Rectangular	.238	.229	.274
Constant .5	.385	.350	.378
Constant .3	.196	.204	.285
Constant .7	.408	.427	.397
Test Length	**KR20**	**KR21**	**S-H**
20	.422	.427	.463
30	.309	.296	.333
50	.190	.185	.205
Sample Size	**KR20**	**KR21**	**S-H**
20	.261	.265	.274
32	.297	.303	.329
52	.364	.339	.397

of S-H. Also, for both SI and SD item sets, the standard error of S-H is greater than that of KR21 which is greater than or equal to the standard error of KR20. The standard error of each of the three reliability estimates varies inversely with sample size and test length, for both SI and SD items. In every case, SD increased the expected value of the estimate and decreased its standard error.

REFERENCES

Brown, W. "Some Experimental Results in the Correlation of Mental Abilities," British Journal of Psychology, Volume 3, 1910, pp. 296-322.

Kuder, G. F. and Richardson, N. W. "The Theory of Estimation of Test Reliability," Psychometrika, Volume 2, 1937, pp. 151-160.

Lehmer, D. H. "Mathematical Methods in Large Scale Computing Units," Annals of the Computation Laboratory of Harvard University, Volume 26, 1951, pp. 141-146.

Pike, M. C. and Hill, I. D. "Algorithm 266: Pseudo-Random Numbers," Communications of the Association for Computing Machinery, Volume 8, 1965, pp. 605-606.

Siegel, Sidney Nonparametric Statistics for the Behavioral Sciences, New York: McGraw-Hill, 1956.

Spearman, C. "Correlation Calculated With Faulty Data," British Journal of Psychology, Volume 3, 1910, 271-295.