DOCUMENT RESUME

ED 095 223		TM 003 910

AUTHOR Stocking, Martha; And Others

TITLE A Program for Estimating the Relative Efficiency of Tests at Various Ability Levels, for Equating True Scores, and for Predicting Bivariate Distributions of Observed Scores.

INSTITUTION Educational Testing Service, Princeton, N.J.
SPONS AGENCY National Science Foundation, Washington, D.C.
REPORT NO ETS-RM-73-24
PUB DATE Oct 73
NOTE 20p.

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Computer Programs; *Equated Scores; *Statistical Analysis; Test Reliability; *Test Results; Test Validity; *True Scores

ABSTRACT

For two tests measuring the same trait, the program, BIV20, equates the scores using the two True score distributions estimated by the univariate method 20 program (see Wingersky, Lees, Lennon, and Lord, 1969) and, with these equated true scores and their distributions, estimates the bivariate distribution scores and the relative efficiency of the two tests at various ability levels. The method is described and formulas given in the appendix. If desired, this estimated distribution is compared to an actual bivariate observed-score distribution provided by the user, and a chi-square between the estimated and actual distributions is computed. The program is written in Fortran IV for the IBM 360/65. Its only restriction is that the maximum number of items for each test is 50. (Author/RC)

RM-73-24

RESEARCH MEMORANDUM

A PROGRAM FOR ESTIMATING THE RELATIVE EFFICIENCY OF TESTS AT VARIOUS ABILITY LEVELS, FOR EQUATING TRUE SCORES, AND FOR PREDICTING BIVARIATE DISTRIBUTIONS OF OBSERVED SCORES

Martha Stocking, Marilyn S. Wingersky,
Diana M. Lees, Virginia Lennon, and Frederic M. Lord

This Memorandum is for interoffice use. It is not to be cited as a published report without the specific permission of the authors.

Educational Testing Service
Princeton, New Jersey
October 1973

BEST COPY AVAILABLE

A Program for Estimating the Relative Efficiency of Tests at Various Ability Levels, for Equating True Scores, and for Predicting Bivariate Distributions of Observed Scores*

For two tests measuring the same trait, the program, BIV20, equates the true scores using the two true-score distributions estimated by the univariate method 20 program (see Wingersky, Lees, Lennon, & Lord, 1969) and, with these equated true scores and their distributions, estimates the bivariate distribution of observed scores and the relative efficiency of the two tests at various ability levels. The method is described and formulas given in Appendix A. If desired, this estimated distribution is compared to an actual bivariate observed-score distribution provided by the user, and a chi-square between the estimated and actual distributions is computed.

This program is written in Fortran IV for the IBM 360/65. Its only restriction is that the maximum number of items for each test is 50.

Uses

Some of the uses of the program are the following (see Lord, 1965, for details):

1. To check on the mathematical model used to estimate the true-score distributions (Lord, 1967);

*This updates the Research Memorandum 69-2 titled "A program for estimating a bivariate distribution of test scores from the marginals" by Martha Stocking, Diana M. Lees, Virginia Lennon and Frederic M. Lord supported in part by contract Nonr-2752(00) between the Office of Naval Research and Educational Testing Service. Additions and revisions to the program were supported by grant GB-32781X from National Science Foundation.

2. To estimate how a group of examinees that has taken only a short form of a test would have performed on a long form. (This is useful in setting up test norms.)
3. To estimate how a group of examinees that has been selected on a certain test would perform on a parallel test.
4. To try to equate two rather different tests measuring the same trait when both tests cannot be given to the same group of examinees.
5. To investigate whether or not two different tests measure the same psychological trait. (If the program does not provide a good estimate of the actual bivariate scatterplot, it may be possible to conclude that the assumption--that the two tests measure the same psychological trait has been violated.)
6. To compute the relative efficiency of the two tests at various ability levels.

An illustration equating true scores and estimating relative efficiencies for several pairs of tests is given in Lord, 1973b.

Assumptions

The results are based on the following assumptions (see Lord, 1967, for a detailed and rigorous statement):

1. The conditional distribution of observed scores for fixed true score is a (certain approximation to a) compound binomial distribution.

4. The true-score distribution in the group tested is "smooth."
5. The true scores on the two tests studied are perfectly, although possibly curvilinearly, related.

Input

The output from a univariate method-20 program is required for both tests being investigated. For detailed description of the input formats, reader is referred to Appendix C.

Output

The user receives as printed output the following:

1. The estimated bivariate observed-score distribution and the regression of each estimated test on the other (row and column means of the estimated bivariate observed-score distribution).
2. The estimated marginal observed-score distributions; the mean, variance, and Kuder-Richardson K_{21} for each of these distributions; also for the estimated bivariate distribution, the estimated correlation between observed scores and the correlation ratios for Test X given Test Y and Test Y given Test X.
3. The estimated probability distribution of true scores for each test.
4. An equating of the true scores of both tests by the equipercentile method.
5. The relative efficiencies of the two tests at various ability levels.

If the actual observed-score bivariate distribution is provided, the user receives the additional output:

6. The chi-square between the estimated and actual bivariate observed-score distributions, and its probability level; the grouping of the two bivariate distributions for the chi-square, the estimated and actual group frequencies, and the group contributions to chi-square.
7. A graph showing for each cell in the bivariate distribution a "." if the estimated frequency for the cell was greater than 1, also the sign of the contribution to the X^2 of the group containing the cell if the contribution was greater in absolute value than 2.
8. The regression of each actual test on the other (row and column means of actual bivariate observed-score distribution). These can be compared with item 1 above.
9. The actual marginal observed-score distributions; the mean, variance, and Kuder-Richardson R_{21} for each of these distributions, the actual correlation between the observed scores, and the correlation ratio for Test X given Test Y and for Test Y given Test X for the observed bivariate distribution. These may be compared with item 2 above.

The program punches the logarithms of the estimated relative efficiencies, the true-score percentiles at which they are computed and the estimated and actual marginal distributions for both tests along with identifying information. The equated true scores may also be punched.

Availability

A copy of the program may be obtained upon written request from the authors, Educational Testing Service, Princeton, New Jersey 08540. The user must provide a tape on which the program will be loaded in 80 character card images and must specify whether the tape should be blocked, in EBCDIC or BCD, 7-track or 9-track, and the tape density and parity. The tape will be unlabeled.

Disclaimer

Although the program has worked satisfactorily on the data we have tried, no claim is made that the program is free of error.

Appendix A

The program BIV20 computes the estimated bivariate distribution of observed scores between two tests, Test X and Test Y, using the parameters ~~of~~ the corresponding univariate true-score distributions. If we let $\phi(x,y)$ be the population bivariate distribution of observed scores for the two tests, ζ be the true score on Test X, $g(\zeta)$ be the distribution of true scores for Test X, $h(x,y|\zeta)$ be the joint conditional distribution of observed scores x and y on Test X and Test Y, given ζ, and a_ζ and b_ζ be the lower and upper limits respectively, of the true-score distribution for Test X, then

$$(1) \qquad \phi(x,y) = \int_{a_\zeta}^{b_\zeta} g(\zeta)h(x,y|\zeta)\,d\zeta \quad .$$

If we let η be a true score on Test Y and $g'(\eta)$ be the distribution of true scores for Test Y, then under assumption 3, η is a function of ζ. When $g(\zeta)$ and $g'(\eta)$ ~~are given, this functional relation, $\psi(\zeta)$, say,~~ can be determined numerically, for any fixed ζ, from

$$(2) \qquad \int_{a_\zeta}^{\zeta} g(\zeta)\,d\zeta = \int_{a_\eta}^{\psi(\zeta)} g'(\eta)\,d\eta \quad ,$$

where a_η is the lower limit of the true-score distribution for Test Y, by inverse interpolation methods. Under the assumption that the errors of measurement are independently distributed when ζ is fixed, equation (1) becomes

$$(3) \qquad \phi(x,y) = \int_{a_\zeta}^{b_\zeta} g(\zeta)h(x|\zeta)h'(y|\eta)\,d\zeta \quad ,$$

where $h(x|\zeta)$ is the conditional distribution of observed score x on Test X given a fixed true score ζ, and $h'(y|\eta)$ is the conditional distribution of observed score y on Test Y for a fixed true score η where $\eta = \psi(\zeta)$. If we let $\hat{\ }$ denote estimated quantities, then equation (2) becomes

$$(2)' \qquad \int_{a_\zeta}^{\zeta} \hat{g}(\zeta)\, d\zeta = \int_{a_\eta}^{\hat{\psi}(\zeta)} \hat{g}'(\eta)\, d\eta$$

and equation (3) becomes

$$(3)' \qquad \hat{\phi}(x,y) = \int_{a_\zeta}^{b_\zeta} \hat{g}(\zeta) h(x|\zeta) h'(y|\hat{\psi}(\zeta))\, d\zeta \; .$$

The estimated bivariate distribution of observed scores, $\hat{\phi}(x,y)$, is obtained by first determining $\hat{\psi}(\zeta)$ from equation (2)', then performing the integration indicated in equation (3)' by quadrature.

The reader is referred to equation 21 (Lord, 1967) for the mathematical form of $\hat{g}(\zeta)$ and $\hat{g}'(\eta)$, and to equation 7 (Lord & Lees, 1967) for the mathematical form of $h(x|\zeta)$ and $h'(y|\eta)$.

The estimated relative efficiency of Test Y versus Test X is given by the formula

$$R.E._{yx} = \frac{n_y^2}{n_x^2} \frac{(n_x - 2k_x)\, \zeta(1 - \zeta)}{(n_y - 2k_y)\, \eta(1 - \eta)} \left(\frac{\hat{g}(\zeta)}{\hat{g}'(\eta)} \right)^2$$

where k_x and k_y are defined in Lord, 1965, eq. 9, n_x is the number of items on Test X and n_y is the number of items on Test Y. For the derivation of this formula see Lord (1973a). The relative efficiencies are computed for equally spaced ζ 's between a_ζ and b_ζ. A graph is produced of $\hat{G}(\zeta)$ vs $\log_{10}$ R.E.$_{yx}$ where

$$\hat{G}(\zeta) = \int_{a_\zeta}^{\zeta} \hat{g}(\zeta)\, d\zeta$$

for the equally spaced ζ 's.

Appendix B

Accuracy of Computations

As a general rule the user obtains estimated marginal observed-score distributions which agree to at least three decimal places with the estimated observed-score distributions obtained from a univariate method-20 program. There are two major sources of numerical error.

1. The numerical solution of equation 2, Appendix A. This solution is accomplished by an iterative inverse interpolation procedure using Bessel's interpolation formula (Scarborough, 1955, p. 77), and, as with any interpolation process, this generates computational errors.

The accuracy attempted by this iterative procedure is controlled by an input option available to the user.

2. Equation 3 of Appendix A is integrated with Simpson's rule where the user specifies the number of intervals. Accuracy can be increased by using more intervals. However, from past experience approximately 50 invervals have been sufficient.

For true-score distributions where $g(\zeta = a_\zeta)$ and/or $g(\zeta = b_\zeta)$ are not close to 0, the error in integrating the corresponding tail of $f(x,y)$ can be quite large. There are two ways to possibly reduce this error. The user may specify that the Simpson's rule interval containing a_ζ and the interval containing b_ζ be subdivided into much smaller intervals for a more accurate computation of the area under the curve in these intervals. Also it may be possible to increase the accuracy by reversing the abscissa and ordinate, putting Test Y on the abscissa and Test X on the ordinate. This will only help if, for Test Y, $g(\eta = a_\eta)$ and $g(\eta = b_\eta)$ are approximately 0.

The maximum number of intervals plus subintervals in the tails is 280.

Appendix C

Input to Bivariate Program for Method-20 (BIV20)

In the following description of the input required by BIV20, the test on the abscissa of the bivariate scatterplot is referred to as "Test X." The test on the ordinate of the bivariate scatterplot is referred to as "Test Y." Zeta (ζ) denotes the true score for Test X, and eta (η) denotes the true score for Test Y. All input information marked with an asterisk (*) comes from the univariate method-20 output. All other information must be supplied by the user. All input information followed by a (+) can be continued on more than one card. For detailed descriptions of the mathematical functions for the input variables, the user is referred to Lord and Lees (1967) and Lord (1967).

Card Number	Description	
*1.	Title card for Test X	
	col. 1 - 60	title of test, to be used as heading information on output.
	col. 63 - 65	number of items in test, must be less than or equal to 50. The format is (I3).
	col. 66 - 72	number of examinees taking test. The format is (F7.0).
*2.	Title card for Test Y with the same information and format as card number 1.	
*3.	Parameter card for Test X. These parameters are necessary to compute the estimated true-score distribution for Test X. The format is (5E15.8).	
	col. 1 - 15	a_ζ
	col. 16 - 30	b_ζ

Card Number	Description		
*3. (cont'd)	col. 31 - 45	d_ζ	parameters for the "smoothing function" $\gamma(\zeta)$ described in Lord, 1969.
	col. 46 - 60	Δ_ζ	
	col. 61 - 75	k_x	

(If the simplified method-20 program has been used, $a_\zeta = 0$, $b_\zeta = 1$, $d_\zeta = \Delta_\zeta = 0$.)

*4. Parameter card for Test Y with the same information and format as card number 3.

5. Options Card

col. 1 - 5 — punch the number of quadrature intervals to be used in computing $\hat{\Phi}(x,y)$. This number must be less than or equal to 280, and must be even. Generally this number ranges between 50 and 100.

col. 6 - 10 — punch 0 if the ζ 's and η 's are to be equated by the program. This is the usual option.

1 if equated ζ 's and η 's are supplied by the user. The user must then provide cards number 13 (b)+.

-1 if the ζ 's and η 's are assumed to be identical.

col. 11 - 15 — punch 0 if the grouping for the chi-square is to be computed by the program. This is the usual option.

1 if the grouping for the chi-square is to be computed by the program, and written on scratch tape 4.

2 if the grouping for the chi-square is to be read from tape unit 4.

Card Number	Description			
5. (cont'd)		col. 11 - 15 (cont'd)	3	if the grouping for the chi-square is to be read from cards. The user must provide cards number 19, number 20+, number 21+.
		col. 16 - 20	punch 0	if equated true scores are not to be punched on cards. This is the usual option.
			1	if equated true scores are to be punched on cards. The format will be (5E15.8).
		col. 21 - 25		punch the number of subintervals to be used in integrating the interval containing a_ζ and the interval containing b_ζ. Two times this number plus the contents of cols. 1-5 must be less than 280.

*6. Parameters required for the conditional distributions of observed scores for a given true score for Test X. If the general univariate method-20 program has been used, these are part of its output. If the simplified version has been used, these values may be obtained from Lord and Lees (1967), Figures 10 and 11. The format is (2E15.8).

col. 1 - 15 α

col. 16 - 30 β

*7. The same information and format as card number 6 for Test Y.

*8. The constants associated with the smoothing functions $\gamma(\zeta)$ and $\gamma(\eta)$ required to estimate the true-score distributions for Test X and Test Y. These constants are the reciprocals of

Card Number	Description
*8. (cont'd)	$\int_{a_\zeta}^{b_\zeta} \zeta^{d_\zeta}(1-\zeta)^{\Delta_\zeta}\, d\zeta$ for Test X, and $\int_{a_\eta}^{b_\eta} \eta^{d_\eta}(1-\eta)^{\Delta_\eta}\, d\eta$ for Test Y. If the general univariate method-20 program has been used, these are part of the output. If the simplified version has been used, both these constants are equal to 1. The format is (2E15.8).

col. 1 - 15 $\gamma(\zeta)$ constant.

col. 16 - 30 $\gamma(\eta)$ constant.

Card Number	Description
*9.	The coarse grouping of frequencies of observed scores used in estimating the true-score distribution of Test X. The format is (40I2).

col. 1 - 2 the number of coarse groups. This number must be less than or equal to 26.

col. 3 - 4
col. 5 - 6
col. 7 - 8
col. 9 - 10
⋮
for each group, the number of frequencies in that group. The listing of groups starts with the group containing the frequency for the lowest observed score, and ends with the group containing the frequency of the highest observed score.

Card Number	Description
*10.	The same information and format as card number 9 for Test Y.
*11.	The parameters λ_u of the true-score distribution for Test X. The format is (5E15.8).
*12.	The same information and format as card number 11 for Test Y.

Card Number	Description
13. (a)	If equated true scores are to be computed by the program (col. 10 of card number 5 equals 0), the epsilon for the iterative inverse interpolation process is used to equate true scores. If the results of two successive iterations differ by less than epsilon, the iterative procedure is halted. This epsilon is usually .00001. The format is (F15.8).
(b)+	If equated true scores are to be supplied by the user (col. 10 of card number 5 equals 1), the true scores for Test X followed (starting on a new card) by the true scores for Test Y. The format must be (14F5.5). There must be the same number of ζ 's and η 's, and this number must be one greater than the number of quadrature intervals given in col. 1-5 of card number 5.
	Note that 13 (a) and 13 (b)+ are mutually exclusive.
14.	Criteria for the grouping of the estimated and actual bivariate observed-score distributions for the chi-square. (Note the information required in col. 47-48 before proceeding.) The format is (2F8.3,16I2).

col. 1 - 8 — punch the minimum group size required. Usually if the number of examinees is greater than 3000, the minimum is 30; if it is less than or equal to 3000, the minimum is 20.

col. 9 - 16 — punch the minimum cell frequency for the estimated bivariate observed-score distribution. All cells with a frequency less than or equal to this value will be set to zero before the grouping for chi-square is done. This minimum is usually .005.

col. 17 - 46 — punch, in (15I2) format, the following numbers: 00,01,02,05, 08,09,10,13,16,17,18,20,25,28, 32. These are distance criteria used to determine whether a prospective cell should be included in a group.

Card Number	Description		
14. (cont'd)	col. 47 - 48	punch 0	if the actual bivariate observed-score distribution will be read, and the chi-square computed. The user must provide cards number 15, number 16+, number 17, number 18+.
		1	if the actual bivariate observed-score distribution will not be read. Col. 1-46 may be left blank. The program will halt after the computation of the estimated bivariate observed-score distribution.
15.	If the actual bivariate observed-score distribution is to be read by the program (col. 47-48 of card number 14 is zero) punch this card and cards number 16+, number 17, and number 18+.		
	col. 1 - 30		the integer format of the actual bivariate observed-score distribution which must be read in by rows, with each row being the frequency distribution of observed scores on Test Y (in ascending order by observed score) for a given observed score on Test X. For example, (20I4) punched in these columns would mean that the actual bivariate scatterplot will be punched with 20 four-digit integers per card.
	col. 31 - 35		the number of elements per card. The format is (I5).
	col. 36 - 40		the observed score on Test X for the first row to be read in (usually 0). Rows must be read in ascending order of observed scores on Test X. This means that initial rows of all zeroes need not be punched, but will be assumed zero by the program. The format is (I5).

Card Number	Description
15. (cont'd)	col. 41 - 45 punch 0 if the input matrix has not been transposed. 1 if the input matrix has been transposed. Note: This option is used only when col. 36-40 is zero. The format is (I5).
16.+	Actual bivariate scatterplot, punched by rows in the format given by card number 15. There are two ways in which the scatterplot may be punched, depending upon col. 36-40 of card number 15. If col. 36-40 of card number 15 is not equal to zero, consecutive zero cells may be replaced by a negative number whose magnitude equals the number of zero cells to be omitted. Each new row does not start on a new card. If col. 36 - 40 of card number 15 equals zero, all consecutive zero cells must be punched, and each new row must start a new card.
17.	Punch in col. 1 - 30 the integer format of the actual marginal frequency distributions which will be read.
18.+	Actual marginal frequency distribution of Test X, followed (on a new card) by that of Test Y, both punched in the format given by card number 17. Each marginal frequency distribution is punched in ascending order of observed score.
19.	If the grouping for the chi-square is to be read from cards (col. 15 of card number 5 is equal to number 3), punch the integer format in which the grouping will be read in col. 1-60. The user must also punch cards number 20+ and number 21+. (The information in col. 1-46 of card number 14 will be ignored.)
20.+	Matrix of group numbers, one for each cell in the bivariate scatterplot, punched by rows in the format given by card number 19, with each new row starting a new card. The ij -th element of this matrix is the number of the group to which the ij -th cell of the bivariate scatterplot belongs.

Card Number	Description
21.+	Information required for the combining of undersized groups, punched in the integer format given by card number 19.

punch first the total number of groups

second the number of groups to be combined

third old group number, new group number

old group number, new group number

old group number, new group number

⋮

(for each group to be moved, starting with the lowest old group number.)

References

Lord, F. M. A strong true-score theory, with applications. Psychometrika, 1965, 30, 239-270.

Lord, F. M. Estimating true-score distributions in psychological testing (An empirical Bayes estimation problem). Psychometrika, 1969, 34, 259-299.

Lord, F. M. The relative efficiency of two tests as a function of ability level. Research Bulletin 73-41. Princeton, N.J.: Educational Testing Service, 1973. (a)

Lord, F. M. True-score equating, relative efficiency and the Anchor Test Study data. Research Bulletin 73-00. Princeton, N.J.: Educational Testing Service, 1973. (b)

Lord, F. M., & Lees, D. M. Estimating true-score distributions for mental tests (Methods 12, 14, 15). Research Memorandum 67-1 and ONR Technical Report, Contract Nonr-2752(00). Princeton, N.J.: Educational Testing Service, 1967.

Scarborough, J. B. Numerical mathematical analysis. (3rd ed.). Baltimore, Md.: The Johns Hopkins Press, 1955.

Wingersky, M., Lees, D. M., Lennon, V., & Lord, F. M. A computer program for estimating true score distributions and graduating observed-score distributions. Educational and Psychological Measurement, 1969, 29, 689-692.