DOCUMENT RESUME

ED 056 097 TM 000 909

AUTHOR	Livingston, Samuel A.; Stanley, Julian C.
TITLE	What is the True Coefficient of Partial Correlation?
INSTITUTION	Johns Hopkins Univ., Baltimore, Md. Center for the Study of Social Organization of Schools.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
REPORT NO	R-107
PUB DATE	Jul 71
GRANT	OEG-2-7-061610-0207
NOTE	14p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Correlation; Multiple Regression Analysis; *Statistical Analysis; *True Scores

ABSTRACT

Although partial correlation is a correlation of residuals, the correlation of the true-score components of these residuals is not equivalent to the partial correlation of the true scores themselves. The source of this discrepancy is explained and its implications are briefly discussed. (Author)

Center for Social Organization of Schools

ED056097

REPORT NO. 107

JULY, 1971

WHAT IS THE TRUE COEFFICIENT OF PARTIAL CORRELATION?

SAMUEL A. LIVINGSTON

JULIAN C. STANLEY

The Johns Hopkins University

STAFF

John L. Holland, Director

James M. McPartland, Assistant Director

Virginia Bailey
Thelma Baldwin
Zahava D. Blum
Judith P. Clark
James S. Coleman
David DeVries
Keith Edwards
Doris R. Entwisle
Gail Fennessey
James Fennessey
Jean Finnegan
Ellen Greenberger
Rubie Harris
Edward J. Harsch
Robert T. Hogan
John H. Hollifield
Nancy L. Karweit
Judith Kennedy
Steven Kidder
Hao-Mei Kuo
Samuel Livingston
Edward L. McDill
Rebecca J. Muraro
Jeanne O'Connor
Suzanne K. Pieper
Peter H. Rossi
Leslie Schnuelle
Aage B. Sørensen
Julian C. Stanley
Mary C. Viernstein
Diana F. Ward
Murray A. Webster
Phyllis K. Wilson

What Is the True Coefficient of Partial Correlation?

GRANT NO. OEG-2-7-061610-0207

PROJECT NO. 61610-01

Samuel A. Livingston

Julian C. Stanley

REPORT NO. 107

JULY, 1971

Published by the Center for Social Organization of Schools, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of the Office of Education, and no official endorsement by the Office of Education should be inferred.

The Johns Hopkins University

Baltimore, Maryland

INTRODUCTORY STATEMENT

The Center for Social Organization of Schools has two primary objectives: to develop a scientific knowledge of how schools affect their students, and to use this knowledge to develop better school practices and organization.

The Center works through five programs to achieve its objectives. The Academic Games program has developed simulation games for use in the classroom, and is studying the processes through which games teach and evaluating the effects of games on student learning. The Social Accounts program is examining how a student's education affects his actual occupational attainment, and how education results in different vocational outcomes for blacks and whites. The Talents and Competencies program is studying the effects of educational experience on a wide range of human talents, competencies, and personal dispositions in order to formulate-- and research--important educational goals other than traditional academic achievement. The School Organization program is currently concerned with the effects of student participation in social and educational decision-making, the structure of competition and coo[illegible]ratic reward systems, ability-g[illegible]ools, and effects of school quality. The Careers and Curricula program bases its work upon a theory of career development. It has developed a self-administered vocational guidance device to promote vocational development and to foster satisfying curricular decisions for high school, college, and adult populations.

This report, like others occasionally published by the Center, deals with a subject common to all programs -- that of scientific measurement.

Acknowledgment

We thank Frederic M. Lord and Carl E. Bereiter for their help in clarifying the problems presented in this article.

Abstract

Although partial correlation is a correlation of residuals, the correlation of the true-score components of these residuals is not equivalent to the partial correlation of the true scores themselves. The source of this discrepancy is explained and its implications are briefly discussed.

The correction of the first-order partial correlation coefficient for attenuation due to errors of measurement reveals a seeming paradox in test-score theory. Lord (1958, pp. 440-441) and Bereiter (1963, p. 8) have touched on the problem in connection with measures of change, as has Stanley (1971, pp. 389-390) in connection with the reliability of errors of estimate. However, none of these authors directly confronts the discrepancy discussed below.

Partial correlation is defined as a correlation of residuals,

$$\rho_{xy.w} = \rho_{(x - \hat{x}|w)(y - \hat{y}|w)} ,$$

where $\hat{x}|w$ and $\hat{y}|w$ are the least-squares linear regression estimates of $\underline{x}$ and $\underline{y}$, respectively, from $\underline{w}$. One might then expect the true partial correlation coefficient to be $\rho_{(T_x - T_{\hat{x}|w})(T_y - T_{\hat{y}|w})}$, where $T_{\hat{x}|w}$ and $T_{\hat{y}|w}$ are the true-score components of the estimates. The seeming paradox is that when this expression is expanded, it yields an expression which is not equivalent to the partial correlation of the true scores themselves. The partial correlation of the true scores is the correlation of a different set of residuals:

$$\rho_{T_xT_y.T_w} = \rho_{(T_x - \hat{T}_x|T_w)(T_y - \hat{T}_y|T_w)} ,$$

where $\hat{T}_x|T_w$ and $\hat{T}_y|T_w$ are the least-squares linear-regression estimates of T_x and T_y, respectively, from T_w.

The source of the discrepancy--that the two correlations are not equivalent--lies in the fact that the true-score component of the estimate of $\underline{x}$ from $\underline{w}$ is generally not equal to the estimate of T_x from T_w. We can, without loss of generality, let w, x, and y represent deviations from their respective means. Then

$T_{\hat{x}|w} \not\equiv \hat{T}_x | T_w$; that is,[1]

$T_{(\beta_{xw} w + \alpha_{xw})} \not\equiv \beta_{T_x T_w} T_w$.

Similarly, $T_{\hat{y}|w} \not\equiv \hat{T}_y | T_w$.

In fact, the two expressions $T_{\hat{x}|w}$ and $\hat{T}_x | T_w$ will be equivalent only when $\rho_{ww'} = 1$, that is, when the predictor is perfectly reliable.

To demonstrate this fact, we must first show that the regression coefficient β_{xw} is constant over forms (i.e., repeated measurements). It is, if we assume that the individual true scores T_x and T_w and the variance error of measurement $\sigma^2_{e_w}$ are constant from form to form, and that $\sigma_{e_w e_x}$, $\sigma_{T_w e_x}$, $\sigma_{e_w T_x}$, and $\sigma_{T_w e_w}$ are all zero.[2]

---

[1] The symbol $\not\equiv$ is used here to indicate the absence of an identity; the two expressions it separates are not equal for all values of the variables involved.

[2] Note that $\beta_{xw} = \sigma_{wx}/\sigma^2_w = \sigma_{T_w T_x}/(\sigma^2_{T_w} + \sigma^2_{e_w})$.

The true score of $\hat{x}|w$ is defined as its expectation over forms:

$$T_{\hat{x}|w} = E_f(\hat{x}_f|w) = E_f(\beta_{xw}w_f) \quad .$$

Since β_{xw} and α_{xw} are constant over forms,

$$T_{\hat{x}|w} = \beta_{xw}E_f(w_f) = \beta_{xw}T_w \quad .$$

Therefore, the regression coefficient for estimating the true-score component of $\hat{x}|w$ from the true-score component of w is the same as the regression coefficient for estimating $\underline{x}$ from $\underline{w}$. Figure 1 illustrates this fact.

What about the regression coefficient for estimating the true-score component of $\underline{x}$ itself (rather than of $\hat{x}|w$) from the true-score component of w? It is

$$\beta_{T_xT_w} = \frac{\sigma_{T_wT_x}}{\sigma^2_{T_w}} = \frac{\sigma_{wx}}{\rho_{ww'}\sigma^2_w} = \frac{\beta_{xw}}{\rho_{ww'}} \quad ,$$

which is β_{xw} corrected for attenuation. Since $0 \leq \rho_{ww'} \leq 1$, $\beta_{T_xT_w} \geq \beta_{xw}$. Therefore, for a given set of scores, $\hat{T}_x|T_w$ will have greater variance than $T_{\hat{x}|w}$. For any individual observed score w (other than the mean), $T_{\hat{x}|w}$ will lie closer to the mean than will

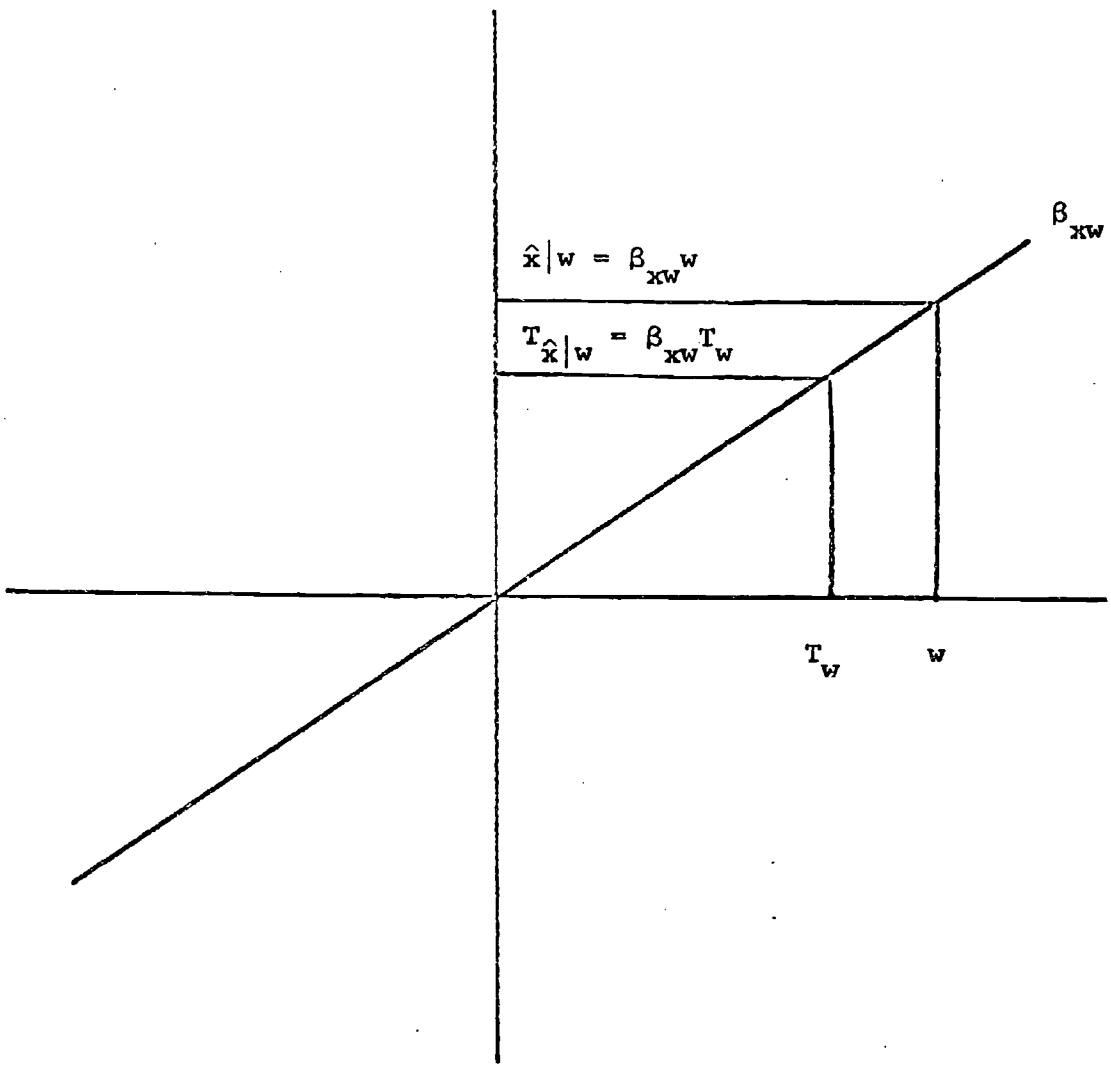

Figure 1

The true-score component of a linear-regression estimate.

$\hat{T}_x | T_w$. Figure 2 illustrates this fact.

When the first of the two correlations, $\rho_{(T_x - T_{\hat{x}|w})(T_y - T_{\hat{y}|w})}$, is expressed in terms of the zero-order correlations, reliability coefficients, and standard deviations of $\underline{w}$, $\underline{x}$, and $\underline{y}$, the standard deviations cancel out, leaving

$$(1) \quad \frac{\rho_{xy} - \rho_{wx}\rho_{wy}}{\sqrt{[\rho_{xx'} - \rho_{wx}^2(2 - \rho_{ww'})][\rho_{yy'} - \rho_{wy}^2(2 - \rho_{ww'})]}} .$$

When the second of the two correlation coefficients, $\rho_{(T_x - \hat{T}_x|T_w)(T_y - \hat{T}_y|T_w)}$, is expanded in a similar manner, the resulting expression is

$$(2) \quad \frac{\rho_{ww'}\rho_{xy} - \rho_{wx}\rho_{wy}}{\sqrt{(\rho_{ww'}\rho_{xx'} - \rho_{wx}^2)(\rho_{ww'}\rho_{yy'} - \rho_{wy}^2)}} .$$

These two expressions are generally not equivalent. They become equivalent when $\rho_{ww'} = 1$, that is, when the predictor variable is perfectly reliable.

Formula (2) is equivalent to the expression that results from correcting the zero-order correlations for attenuation before entering them into the partial correlation formula. That is,

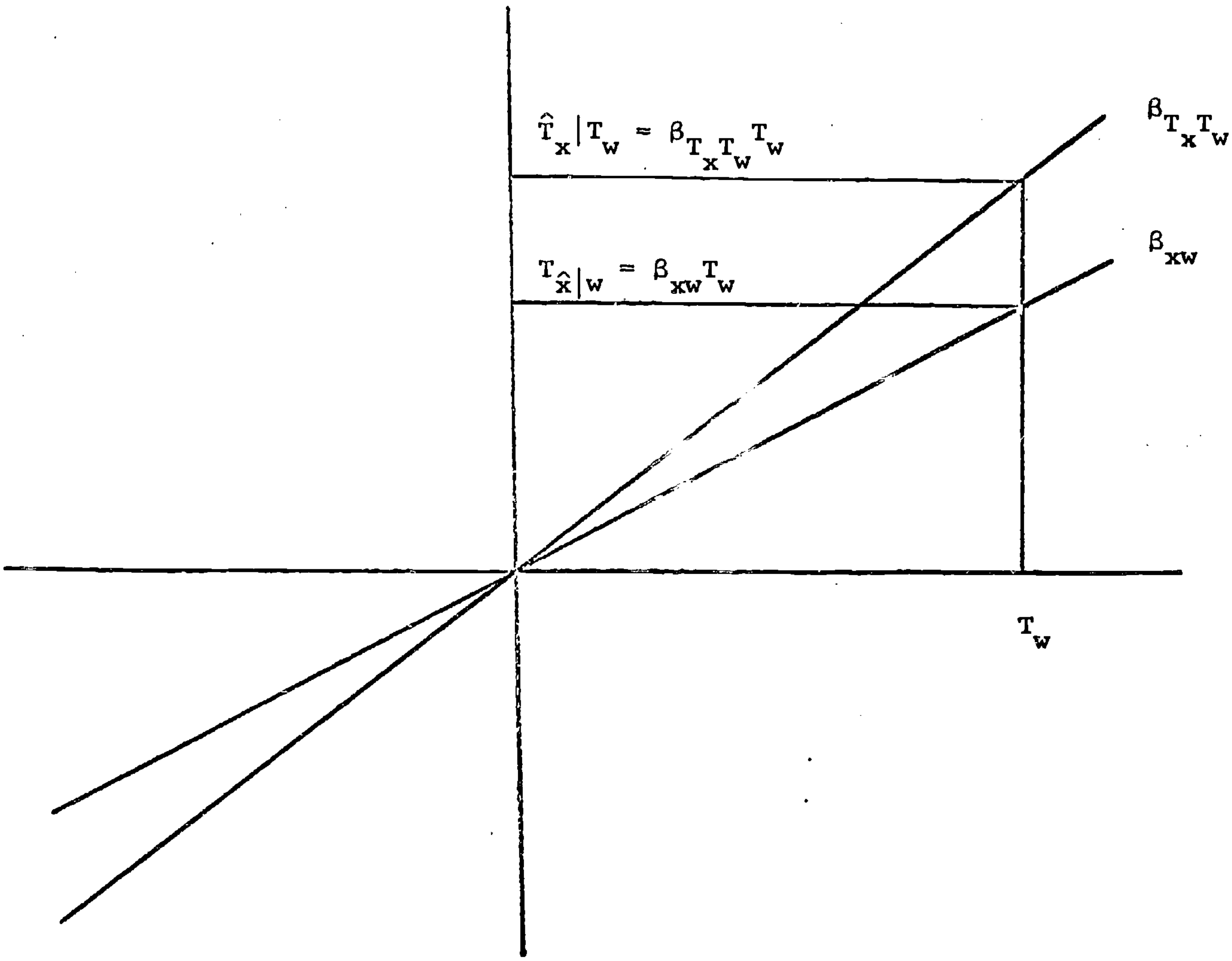

Figure 2.

The true-score estimate of $\underline{x}$ lies further from the mean than the true-score component of the raw-score estimate of $\underline{x}$.

(3) $$\rho_{T_xT_y\cdot T_w} = \rho_{(T_x - \hat{T}_x|T_w)(T_y - \hat{T}_y|T_w)}$$

$$= \frac{\rho_{T_xT_y} - \rho_{T_wT_x}\rho_{T_wT_y}}{\sqrt{(1 - \rho^2_{T_wT_x})(1 - \rho^2_{T_wT_y})}}$$

$$= \frac{\dfrac{\rho_{xy}}{\sqrt{\rho_{xx'}}\sqrt{\rho_{yy'}}} - \left(\dfrac{\rho_{wx}}{\sqrt{\rho_{ww'}}\sqrt{\rho_{xx'}}}\right)\left(\dfrac{\rho_{wy}}{\sqrt{\rho_{ww'}}\sqrt{\rho_{yy'}}}\right)}{\sqrt{\left(1 - \dfrac{\rho^2_{wx}}{\rho_{ww'}\rho_{xx'}}\right)\left(1 - \dfrac{\rho^2_{wy}}{\rho_{ww'}\rho_{yy'}}\right)}} .$$

Thus the discrepancy. Although the partial correlation coefficient is a correlation of residuals that result when $\underline{x}$ and $\underline{y}$ are estimated from $\underline{w}$, the partial correlation based on the true scores is not equivalent to the correlation of the true-score components of the residuals that result when $\underline{x}$ and $\underline{y}$ are estimated from $\underline{w}$.

The question posed in the title of this paper can then be answered as follows: If one wishes to know what the value of a partial correlation would have been if all three variables had been measured without error, he should use $\rho_{T_xT_y\cdot T_w}$, which he can compute by correcting each of the three zero-order correlations for attenuation before entering them into the partial correlation formula. On the other hand, if he has computed the residuals $(x - \hat{x}|w)$ and $(y - \hat{y}|w)$ from observed scores and wishes to know how highly these residuals themselves would correlate if corrected for attenuation, he must use $\rho_{(T_x - T_{\hat{x}|w})(T_y - T_{\hat{y}|w})}$, which he can compute from the <u>uncorrected</u> zero-order correlations by formula (1).

References

Bereiter, C. E. Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), Problems in measuring change. Madison: University of Wisconsin Press, 1963. Pp. 3-20

Lord, F. M. Further problems in the measurement of growth. Educational and Psychological Measurement, 1958, 16, 437-454.

Stanley, J. C. Reliability. Ch. 13 in R. L. Thorndike (Ed.), Educational measurement. (2nd ed.) Washington: American Council on Education, 1971.