Type of the Paper: Article

Correcting the Correction: A Revised Formula to Estimate Partial Correlations between True Scores

Wetcher-Hendricks, Debra*

* Moravian College, Department of Sociology and Anthropology
   Correspondence: Email: wetcher-hendricksd@moravian.edu Tel.: 610-861-1415

Abstract: Bohrnstedt's (1969) attempt to derive a formula to compute the partial correlation coefficient and simultaneously correct for attenuation sought to simplify the process of performing each task separately. He suggested that his formula, developed from algebraic and psychometric manipulations of the partial correlation coefficient, produces a corrected partial correlation value. However, an algebraic error exists within his derivations. Consequently, the formula proposed by Bohrnstedt does not appropriately represent the value he intended it to estimate. By correcting the erroneous step and continuing the derivation based upon his proposed procedure, the steps outlined in this paper ultimately produce the formula that Bohrnstedt desired.

Keywords: Classical Test Theory, Classical True Score Theory, Correction for Attenuation, Partial Correlation Coefficient

1. Introduction

Psychometricians have long acknowledged the presence and impact of measurement error. The very first principle of Classical Test Theory dictates that X=T+E. This equation suggests that an observed score (X) represents not only a measurement of the entity of interest, or the true score (T), but also some level of error (E) that distorts the researcher's observation. If researchers could determine the portion of an observed score that consists of the true score and of error, they could simply focus their analyses upon true-score values to obtain actual representations of the conditions they study. However, because error scores exist as latent components of observed scores, researchers cannot determine the exact extent to which they affect their data and the statistics computed from that data.

Despite the lack of such detailed information, some general understanding of the consequences that error scores can have upon statistics exists. Psychometricians know, for instance, that the presence of measurement error causes the correlation between observed scores to fall below that of true scores, a phenomenon known as attenuation (Carrol, Ruppert, and Stefanski 1995). But, the inability to obtain an actual true score or error score makes it impossible to determine the exact degree to which the observed score correlation underestimates the true score correlation.

This attenuation effect was first recognized by Spearman in 1904. Spearman, resigned to only estimate the correlation between true scores, proposed a formula to theoretically counteract the effects of attenuation on pairwise correlation coefficients. In his formula,

$$\rho_{T_xT_Y} = \frac{\rho_{XY}}{\rho_{XX'}\rho_{YY'}} \quad . \qquad\qquad (1)$$

$\rho_{T_xT_y}$ represents an estimate of the Pearson correlation coefficient between true scores the numerator contains the observed correlation coefficient, and the values in the denominator represent reliabilities. Applying the formula produces a value that exceeds the observed-score correlation.

Spearman's correction for attenuation was groundbreaking with regards to Classical Test Theory. It provided increased accuracy in estimates of the relationship between variables because it eliminated the effects of measurement error to the extent possible. Consequently, as reported by Hakistian, Schroeder, and Rogers (1988), esteemed statisticians and research analysts (Gulliksen, 1950; Block, 1963; Lord and Novick, 1968; Cochran, 1970; Lord, 1970; Crohnbach, 1971) have given much credibility to Spearman's correction for attenuation.

One limitation of Spearman's formula, though, is its application only to pairwise situations. To address partial, part, or multiple correlations, one must apply the correction to each individual pairwise coefficient before inserting it into the relevant formula. This process becomes extremely tedious and time consuming.

In response, Bohrnstedt (1969) proposed a formula for a partial correlation coefficient[1] containing an implicit correction for attenuation. He had intended for his formula to allow the computation of an estimate for the partial correlation between true scores in a single step and his published explanations of his formula describe it as such. However, an algebraic mistake in Bohrnstedt's derivation makes the resulting formula incorrect.

---

[1] The partial correlation coefficient allows for consideration of the linear relationship between the independent and dependent variables, exclusive of any confounding effects that other factors may have upon these variables.

2. Bohrnstedt's Derivation

Bohrnstedt bases his derivation on two psychometric equivalences. One states that dividing the covariance between two data sets by the product of the data sets' standard deviations produces the pairwise correlation coefficient, as shown in Equation 2. This formula is one of many commonly used to calculate pairwise correlation coefficients.

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \qquad (2)$$

The other equivalence used by Borhnstedt allows him to adapt Equation 2 to partial correlations. It follows the leads of others, including Carter (1949) and Wright, et. al. (1959), by focusing upon residuals between variables (e.g σ_{XZ}). Simply, the covariances and standard deviations that appear in Equation 2 become covariances and standard deviations of residuals between these same variables and the potential confounding variable (Z). Bohrnstedt, therefore, defines the corrected partial correlation as the ratio between the covariance of estimated residuals and the square root of the product of the residuals' variances, or

$$\rho_{XY.Z} = \frac{\sigma_{(XZ)(YZ)}}{\sqrt{\sigma_{XZ}^2 \ \sigma_{YZ}^2}}^2 \qquad (3)$$

To acknowledge the focus upon true scores, not observed scores, Bohrnstedt turns to rather sophisticated psychometric principles. In particular, he utilizes the principle that

---

[2] Although, the exponents and square root sign in the denominator of this fraction may seem senseless, as one "undoes" the other, Bohrnstedt's derivation relies upon expansion of the squared values. Thus, the square root sign maintains equality to the more typical partial correlation equation denominator of $(\sigma_{XZ})(\sigma_{YZ})$.

$\sigma_{XZ} = \sigma_{X-b_{XZ}Z}$ (and the corresponding equivalence for Y and Z) to propose that

$$\rho_{T_X T_Y . T_Z} = \frac{\sigma_{(X-b_{XZ} Z)(Y-b_{YZ} Z)}}{\sqrt{\sigma_{X-b_{XZ} Z}^2 \ \sigma_{Y-b_{YZ} Z}^2}} \ . \quad (4)$$

The insertion of these expressions into the existing subscripts, as shown in Equation 4, indicates the consideration of multiple independent variables.

Expanding the covariance in equation 4 and calculating the squared value of the terms in the denominator produces

$$\rho_{T_X T_Y . T_Z} = \frac{\sigma_{XY} - b_{YZ} \sigma_{XZ} - b_{XZ} \sigma_{YZ} + b_{YZ} b_{XZ} \sigma_Z^2}{\sqrt{(\sigma_X^2 - 2b_{XZ} \sigma_{XZ} + \sigma_Z^2 b_{XZ}^2)(\sigma_Y^2 - 2b_{YZ} \sigma_{YZ} + \sigma_Z^2 b_{YZ}^2)}} \ . \quad (5)$$

Once again following psychometric theory, Bohrnstedt redefines b_{xz} as $\rho_{xz}\sigma_x / \rho_{zz'}\sigma_z$ and makes the comparable change for b_{yz}. This manipulation, along with restatement of each covariance as the product of the relevant correlation coefficient and standard deviations (e.g. $\sigma_{XY} = \sigma_X \sigma_Y \rho_{XY}$), produces

$$\rho_{T_X T_Y . T_Z} = \frac{\sigma_X \sigma_Y \rho_{XY} - \frac{\rho_{YZ}\sigma_Y}{\rho_{ZZ'}\sigma_Z}\sigma_X \sigma_Z \rho_{XZ} - \frac{\rho_{XZ}\sigma_X}{\rho_{ZZ'}\sigma_Z}\sigma_Y \sigma_Z \rho_{YZ} + (\frac{\rho_{YZ}\sigma_Y}{\rho_{ZZ'}\sigma_Z})(\frac{\rho_{XZ}\sigma_X}{\rho_{ZZ'}\sigma_Z})\sigma_Z^2}{\sqrt{(\sigma_X^2 - \frac{2\rho_{XZ}^2\sigma_X^2}{\rho_{ZZ'}} + \frac{\rho_{XZ}^2\sigma_X^2}{\rho_{ZZ'}^2})(\sigma_Y^2 - \frac{2\rho_{YZ}^2\sigma_Y^2}{\rho_{ZZ'}} + \frac{\rho_{YZ}^2\sigma_Y^2}{\rho_{ZZ'}^2})}} \ . \quad (6)$$

A simplified version,

---

[3] The "b" in this expression emerges from the regression equation $y=a+bx$, and represents the expected change in the dependent variable given a one-unit change in the independent variable.

$$\rho_{T_X T_Y.T_Z} = \frac{\sigma_X \sigma_Y \rho_{XY} - \frac{2\rho_{XZ}\rho_{YZ}\sigma_X\sigma_Y}{\rho_{ZZ'}} + \frac{2\rho_{XZ}\rho_{YZ}\sigma_X\sigma_Y}{\rho_{ZZ'}^2}}{\sqrt{(\sigma_X^2 - \frac{2\rho_{XZ}^2\sigma_X^2}{\rho_{ZZ'}} + \frac{\rho_{XZ}^2\sigma_X^2}{\rho_{ZZ'}^2})(\sigma_Y^2 - \frac{2\rho_{YZ}^2\sigma_Y^2}{\rho_{ZZ'}} + \frac{\rho_{YZ}^2\sigma_Y^2}{\rho_{ZZ'}^2})}} \ , \qquad (7)$$

results from adding and multiplying terms in equation 6.

Further algebraic simplifications involve three steps of factoring and combining terms. First the standard deviations are factored within the numerator and the variances are factored within the denominator. Second, the existing terms within the numerator and within the denominator are added. The final step involves factoring reliabilities within the numerator and denominator as well as removing variances from beneath radical signs, making them standard deviations. These steps respectively appear as

$$\rho_{T_X T_Y.T_Z} = \frac{\sigma_X \sigma_Y (\rho_{XY} - \frac{2\rho_{XZ}\rho_{YZ}}{\rho_{ZZ'}} + \frac{2\rho_{XZ}\rho_{YZ}}{\rho_{ZZ'}^2})}{\sqrt{\sigma_X^2 (1 - \frac{2\rho_{XZ}^2}{\rho_{ZZ'}} + \frac{\rho_{XZ}^2}{\rho_{ZZ'}^2}) \ \sigma_Y^2 (1 - \frac{2\rho_{YZ}^2}{\rho_{ZZ'}} + \frac{\rho_{YZ}^2}{\rho_{ZZ'}^2})}} \ , \qquad (8)$$

$$\rho_{T_X T_Y.T_Z} = \frac{\sigma_X \sigma_Y (\frac{\rho_{XY}\rho_{ZZ'}^2 - 2\rho_{XZ}\rho_{YZ}\rho_{ZZ'} + \rho_{XZ}\rho_{YZ}}{\rho_{ZZ'}^2})}{\sqrt{\sigma_X^2 (\frac{\rho_{ZZ'}^2 - 2\rho_{XZ}^2\rho_{ZZ'} + \rho_{XZ}^2}{\rho_{ZZ'}^2}) \ \sigma_Y^2 (\frac{\rho_{ZZ'}^2 - 2\rho_{YZ}^2\rho_{ZZ'} + \rho_{YZ}^2}{\rho_{ZZ'}^2})}} \ , \qquad (9)$$

and

$$\rho_{T_X T_Y.T_Z} = \frac{\frac{\rho_{ZZ'}(\rho_{XY}\rho_{ZZ'} - 2\rho_{XZ}\rho_{YZ}) + \rho_{XZ}\rho_{YZ}}{\rho_{ZZ'}^2}}{\sqrt{\left[\frac{\rho_{ZZ'}(\rho_{ZZ'} - 2\rho_{XZ}^2) + \rho_{XZ}^2}{\rho_{ZZ'}^2}\right]\left[\frac{\rho_{ZZ'}(\rho_{ZZ'} - 2\rho_{YZ}^2) + \rho_{YZ}^2}{\rho_{ZZ'}^2}\right]}} \ . \qquad (10)$$

It is this point in the derivation at which Bohrnstedt errs. His canceling of standard deviations within the numerator and denominator of equation 10 is appropriate. But, he INCORRECTLY cancels reliabilities within the numerator and denominator[4] to produce

$$\rho_{T_X T_Y . T_Z} = \frac{\frac{(\rho_{XY}\rho_{ZZ\prime} - 2\rho_{XZ}\rho_{YZ}) + \rho_{XZ}\rho_{YZ}}{\rho_{ZZ\prime}}}{\sqrt{\left[\frac{\rho_{ZZ\prime} - 2\rho_{XZ}^2 + \rho_{XZ}^2}{\rho_{ZZ\prime}}\right]\left[\frac{\rho_{ZZ\prime} - 2\rho_{YZ}^2 + \rho_{YZ}^2}{\rho_{ZZ\prime}}\right]}} \ . \qquad (11)$$

Without realizing the flaws in equation 11, Bohrnstedt continued with his derivation. He added components within the numerator and the denominator and then cancelled reliabilities from each. The equations

$$\rho_{T_X T_Y . T_Z} = \frac{\frac{\rho_{XY}\rho_{ZZ\prime} - \rho_{XZ}\rho_{YZ}}{\rho_{ZZ\prime}}}{\sqrt{\left[\frac{\rho_{ZZ\prime} - \rho_{XZ}^2}{\rho_{ZZ\prime}}\right]\left[\frac{\rho_{ZZ\prime} - \rho_{YZ}^2}{\rho_{ZZ\prime}}\right]}} \qquad (12)$$

and

$$\rho_{T_X T_Y . T_Z} = \frac{\rho_{XY}\rho_{ZZ\prime} - \rho_{XZ}\rho_{YZ}}{\sqrt{(\rho_{ZZ\prime} - \rho_{XZ}^2)(\rho_{ZZ\prime} - \rho_{YZ}^2)}} \ , \qquad (13)$$

respectively, result.

The algebraic error leading to equation 11 devalues equation 13, proposed as the formula to compute a partial correlation coefficient corrected for attenuation. However, the

---

[4] Bohrnstedt cancelled the reliability $\rho_{ZZ\prime}$ from the numerator and denominator. Had $\rho_{ZZ\prime}$ appeared in all terms within the numerators of the general numerator and the general denominator, then he would have cancelled correctly, However, because only the first of the terms in these numerators contained $\rho_{ZZ\prime}$, such factoring and, consequently, such cancelling should not have occurred.

equations presented before equation 11 remain valid and can serve as a basis for fulfilling

Bohrnstedt's original goal.

3. Correcting the Derivation

Using Bohrnstedt's reasoning and making the changes necessary to correct equation 11

allows for the development of the formula that Bohrnstedt desired. With proper canceling of

reliabilities, equation 11 should have appeared as

$$\rho_{T_X T_Y . T_Z} = \cfrac{\cfrac{\rho_{ZZ\prime}(\rho_{XY}\rho_{ZZ\prime} - 2\rho_{XZ}\rho_{YZ}) + \rho_{XZ}\rho_{YZ}}{\rho_{ZZ\prime}^2}}{\sqrt{\left[\cfrac{\rho_{ZZ\prime}(\rho_{ZZ\prime} - 2\rho_{XZ}^2) + \rho_{XZ}^2}{\rho_{ZZ\prime}^2}\right]\left[\cfrac{\rho_{ZZ\prime}(\rho_{ZZ\prime} - 2\rho_{YZ}^2) + \rho_{YZ}^2}{\rho_{ZZ\prime}^2}\right]}} \; . \tag{14}$$

Squared reliabilites in the denominators of the general denominator can be removed from the

radical sign. Cancelling this $\rho_{zz'}^2$ value with the same value in the denominator of the general

numerator forms the single fraction

$$\rho_{T_X T_Y . T_Z} = \cfrac{\rho_{ZZ\prime}(\rho_{XY}\rho_{ZZ\prime} - 2\rho_{XZ}\rho_{YZ}) + \rho_{XZ}\rho_{YZ}}{\sqrt{[\rho_{ZZ\prime}(\rho_{ZZ\prime} - 2\rho_{XZ}^2) + \rho_{XZ}^2][\rho_{ZZ\prime}(\rho_{ZZ\prime} - 2\rho_{YZ}^2) + \rho_{YZ}^2]}} \; . \tag{15}$$

Distribution produces the final version of the formula:

$$\rho_{T_X T_Y . T_Z} = \cfrac{\rho_{XY}\rho_{ZZ\prime}^2 - 2\rho_{XZ}\rho_{YZ}\rho_{ZZ\prime}) + \rho_{XZ}\rho_{YZ}}{\sqrt{(\rho_{ZZ\prime}^2 - 2\rho_{XZ}^2\rho_{ZZ\prime} + \rho_{XZ}^2)(\rho_{ZZ\prime}^2 - 2\rho_{YZ}^2\rho_{ZZ\prime} + \rho_{YZ}^2)}} \; . \tag{16}$$

Equation 16 does exactly what Borhnstedt had suggested his formula would do. By

inserting correlation and reliability coefficients into the appropriate positions, one can

simultaneously compute a partial correlation coefficient and correct for attenuation.

4. Conclusions

Further development of the formula in Equation 16 can increase its usefulness in research situations. Reasonable goals for future endeavors include deriving similar formulas for part and multiple correlations as well as adjusting these formulas to make them applicable to situations involving more than three variables.

Given the similarity between the partial and part correlation formulas, the procedure needed to derive a comparable formula for part correlations is likely very similar to that followed to obtain equation 16. A formula to compute multiple correlation coefficients corrected for attenuation would not resemble the corrected partial coefficient quite as closely as the corrected part coefficient formula would. Yet, the existing derivation certainly suggests a general framework.

Expanding these formulas to manage more than three variables could follow patterns similar to those used in deriving the formulas for three-variable situations. However, as more variables become involved, the number of relationships between variables grows and, obviously, the complexity of the resulting formulas would increase with the addition of each variable.

Optimally, these developments could lead to formulas that correct for attenuation while computing partial, part, and multiple correlation coefficients for any number of variables. The generalizability of such formulas would make them highly useful for researchers wishing to analyze complex data.

References

1. Carrol R.J., Ruppert, D., and Stafanski, L.A. (1995). *Measurement Error in Nonlinear Models.* New York: Chapman and Hall.

2. Spearman, C. (1904). The proof and measurement of association between two things. American Journal of Psychology, 15, 72–101.

3. Hakstian, A. Ralph, Schroeder, Marsha L., and Rogers, W. Todd (1988). Inferential procedures for correlation coefficients corrected for attenuation. *Psychometrika*, 53, 27-43.

4. Bohrnstedt, George W. (1989). Observations on the measurement of change. In Borgatta, Edgar F. (ed.) *Sociological Methodology*. San Francisco: Jossey-Bass, Inc.

5. Carter, A. H. (1949). The estimation and comparison of residual regressions where there are two or more related sets of observations. *Biometrika,36*(1-2), 26-46.

6. Wright, E. M., Manning, W. H., & Dubois, P. H. (1959). Determinants in multivariate correlation. *The Journal of Experimental Education,27*(3), 195-202.