DOCUMENT RESUME

ED 078 003 TM 002 815

AUTHOR Petersen, Anne C.; Bock, R. Darrell
TITLE A Generalized Correction for Attenuation.
PUB DATE 73
NOTE 12p.; Paper presented at Annual Meeting of American Educational Research Association (New Orleans, Louisiana, February 25-March 1, 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Analysis of Covariance; Correlation; *Matrices; Measurement Techniques; Research Methodology; *Statistical Analysis; Tables (Data)

ABSTRACT

Use of the usual bivariate correction for attenuation with more than two variables presents two statistical problems. This pairwise method may produce a covariance matrix which is not at least positive semi-definite, and the bivariate procedure does not consider the possible influences of correlated errors among the variables. The method described here, a generalized correction for attentuation, is a multivariate method which accommodates both the negative variance component problem and the collinearity problem. (Author/KM)

A GENERALIZED CORRECTION FOR ATTENUATION

Anne C. Petersen and R. Darrell Bock
The University of Chicago

Paper presented at the annual meeting of the
Americal Educational Research Association in
New Orleans, February 25-March 1

Correcting data for attenuating measurement error is now widely recognized as an essential procedure whenever the researcher is interested in relationships among the true scores of the variables. Use of the usual bivariate correction for attenuation with more than two variables, however, presents two statistical problems. This pairwise method may produce a covariance matrix which is not at least positive semi-definite. Furthermore, the bivariate procedure does not consider the possible influences of correlated errors among the variables. The method described here, a generalized correction for attenuation, is a multivariate method which accommodates both the negative variance component problem and the collinearity problem.

In the p-variate case, we shall consider two components of the theoretical observed covariance Σ_Y : Σ_T, the true covariance matrix and Σ_E, the error covariance matrix or

$$\Sigma_Y = \Sigma_T + \Sigma_E$$

Theoretically, we can find Σ_T, the true covariance matrix by simply rearranging the equation. Since Σ_Y is exactly equal to $\Sigma_T + \Sigma_E$ then $\Sigma_Y \geq \Sigma_E$ and $\Sigma_T = \Sigma_Y - \Sigma_E \geq 0$. With estimates S of these theoretical Σ's, we cannot be certain that $S_T = S_Y - S_E \geq 0$. It is entirely possible that $S_Y < S_E$, giving a non-positive semi-definite matrix for S_T.

There is an obvious algebraic solution to this dilemma. We can simultaneously diagonalize S_Y and S_E using the spectral decomposition (Rao, 1965). Since both of these are real symmetric matrices, and S_E is positive definite, there exists a non-singular matrix T such that

$$T' S_Y T = L \text{ and } T' S_E T = I$$

where L is a diagonal matrix and I is the identity matrix. The diagonal elements of L, $l_1 \geq \ldots \geq l_p$, of L are the roots of $|S_Y - l\, S_E| = 0$. The ith column vector T_i of T satisfies $S_Y T_i = l_i S_E T_i$. If $l_i \geq 1$ for all i then,

$$\begin{aligned} S_T &= S_Y - S_E \\ &= B' L B - B'B \quad \text{where } B = T^{-1} \\ &= B' (L-I) B \\ &= B' L^* B. \end{aligned}$$

If we restrict the l_i^* so that

$$l_i^* = \max(l_i - 1, 0) \text{ for all } i$$

we can insure that S_T is at least positive semi-definite.

While the algebraic solution just described solves the negative variance component problem, it would be more reassuring to know that the resulting estimates possess desirable properties such as maximum likelihood properties. We will now show that these estimates are maximum likelihood estimates.

S is a sample covariance matrix with n degrees of freedom and is an unbiased estimate of Σ. S has a Wishart distribution with density function

$$f(S) = C(S) \left|\Sigma\right|^{-n/2} \exp\left(-\frac{n}{2} \operatorname{tr} S \Sigma^{-1}\right)$$

where $C(S) = C|S|^{\frac{1}{2}(n-p-1)}$ is a constant.

The joint log likelihood equation is

$$\log L(\Sigma_Y, \Sigma_E) = C_S - \frac{n_Y}{2} \log|\Sigma_Y| - \frac{n_Y}{2} \operatorname{tr} S_Y \Sigma_Y^{-1} - \frac{n_E}{2} \log|\Sigma_E| - \frac{n_E}{2} \operatorname{tr} S_E \Sigma_Y^{-1}.$$

Recall that $S_Y = B' L B$ and $S_E = B'B$.

Substituting these expressions into the log likelihood equation and rearranging terms in the trace:

$$= C_S - \frac{nY}{2} \log|\Sigma_Y| - \frac{nY}{2} \mathrm{tr}(L\, B\Sigma_Y^{-1} B')$$
$$- \frac{nE}{2} \log|\Sigma_E| - \frac{n_E}{2} \mathrm{tr}(B\Sigma_E^{-1} B').$$

Let us define new parameters P and Q such that

$$Q = B'^{-1} \Sigma_Y B^{-1} \quad \text{and} \quad P = B'^{-1} \Sigma_E B^{-1}.$$

The determinants after this reparameterization then are

$$\det Q = (\det B)^{-2} \det \Sigma_Y$$

and $$\det P = (\det B)^{-2} \det \Sigma_E.$$

The logs of the determinants of Σ_Y and Σ_E are

$$\log|\Sigma_Y| = 2 \log|B| + \log|Q|$$
$$= C + \log|Q|$$

and $$\log|\Sigma_E| = C + \log|P|$$

The log likelihood equation is now

$$\ell = C_S' - \frac{nY}{2} \log|Q| - \frac{n_Y}{2} \mathrm{tr}\, L\, Q^{-1} - \frac{nE}{2} \log|P| - \frac{n_E}{2} \mathrm{tr}\, P^{-1}.$$

We can show that the determinant is maximum when the off-diagonal elements are zero. Since we want the maximum solution we can focus on the diagonal elements of the determinants and hence the diagonal elements of all terms in the equation. So now

$$\ell \leq C'_S + \frac{nY}{2} \log|Q_D^{-1}| + \frac{nE}{2} \log|P_D^{-1}| - \frac{nY}{2} \mathrm{tr}\, L\, Q_D^{-1} - \frac{nE}{2} P_D$$

where the subscript D denotes that we are sidering only the diagonal elements of the matrix.

Then $$\ell_{max} = \sum_{i=1}^{p} \left(\frac{n_Y}{2} \log q^{ii} + \frac{n_E}{2} \log p^{ii} - \frac{n_Y}{2} l_i q^{ii} - \frac{n_E}{2} p^{ii}\right)$$

where p^{ii} is the inverse of the ith diagonal element of P.

Therefore we can maximize each l_i separately and we have reduced the problem to a univariate one. In differentiating to solve the equation we must consider the restriction on the parameters in order to achieve a positive semi-definite matrix for S_T. This constraint, in terms of the new parameters, is that $q^{ii} \geq p^{ii}$ or $q_{ii} \leq p_{ii}$. If we consider the unrestricted problem, the maximum likelihood estimates are

$$\hat{p}_{ii} = 1 \quad \text{and} \quad \hat{q}_{ii} = 1_{ii}.$$

Since we cannot insure that the constraint above will not be violated by these unrestricted estimates, we must consider the restricted estimates.

One technique for incorporating restrictions into maximum likelihood estimation is the application of Kuhn-Tucker conditions, assuming that the constraint is convex and differentiable. This assumption holds with the constraint in this problem so the conditions are

$$\pi_i \, (q^{ii} - p^{ii}) = 0 \text{ for } \pi_i \neq 0 \text{ and}$$

$$\frac{\partial \ell_i}{\partial u} + \frac{\partial g_i}{\partial u} = 0 \text{ where u is } q^{ii} \text{ or } p^{ii} \text{ and } g_i = \pi_i (q^{ii} - p^{ii})$$

for the equation

$$\ell_i = \frac{n_Y}{2} \log q^{ii} + \frac{n_E}{2} \log p^{ii} - \frac{n_Y}{2} 1_i \, q^{ii} - \frac{n_E}{2} p^{ii}.$$

The two equations to be solved for the restricted estimates are

$$\frac{n_E}{2} p_{ii} - \frac{n_E}{2} - \pi_i = 0 \qquad (1)$$

$$\frac{n_Y}{2} q_{ii} - \frac{n_Y}{2} 1_i + \pi_i = 0 \quad (2)$$

The first Kuhn-Tucker condition implies that $q^{ii} = p^{ii}$ indicating that the restricted solution is on the boundary. Solving (1) and (2) above:

$$\hat{p}_{ii} = \frac{n_E + n_Y \, 1_i}{n_E + n_Y} = \hat{q}_{ii}\,.$$

Recall that

$$S_T = B' \hat{Q} B - B' \hat{P} B$$

$$= B' (\hat{Q} - \hat{P}) B.$$

We have just shown that

$$\hat{q}_{ii} - \hat{p}_{ii} = \begin{cases} l_i - 1 & \text{if } l_i \geq 1 \text{ (unrestricted estimate)} \\ 0 & \text{if } l_i < 1 \end{cases}$$

So $$S_T = B' L^* B \quad \text{where } l_i^* = \max(l_i - 1, 0).$$

Therefore the algebraic estimate is the maximum likelihood estimate and we can estimate the true covariance matrix in terms of the eigenvectors and eigenvalues of the observed covariance matrix in the metric of the error covariance matrix.

Applications

Two examples will demonstrate the usefulness of the generalized correction for attenuation. The first example uses data from the Louisville Twin Study (Bock and Vandenberg, 1968). The second example uses data from the Fels Research Institute Longitudinal Study (Petersen, 1973).

Example 1.

The variables for this example are the subtests of the Differential Aptitude Test: Spatial, Numerical, Abstract, Verbal, Mechanical, Clerical, Spelling, and Sentences. The subjects are monozygotic and dizygotic twin pairs. The purpose of the study was to estimate the heritable variation of the mental test scores. The components of the total variation are

Σ_H: covariance matrix of the heritable components

Σ_{EN}: covariance matrix of the environmental components

Σ_{ER}: covariance matrix of the measurement error

The sample quantities used are the mean-product matrices M_{WM} and M_{WD}, for within-monozygotic and within-dizygotic pairs, respectively, and are calculated from the between-twin differences for pairs of variables. The expected mean-product matrices are

$$E(M_{WD}) = \Sigma_H + \Sigma_{EN} + \Sigma_{ER}$$

$$E(M_{WM}) = \Sigma_{EN} + \Sigma_{ER} \; .$$

The naive estimate then for Σ_H is $M_{WD} - M_{WM}$. Tables 1 and 2 give these mean-product matrices and Table 3 shows their difference. The mechanical subtest has a negative estimated variance due to heredity. In this example the heritable covariance is analogous the true covariance and the mean-product matrices are analogous to the observed and error matrices. The true correlation matrix, analogous to the heritable correlation matrix, is given in Table 4. The eigenvalues for this matrix, in Table 5, indicate that there are only five dimensions in the data, with only three of these eigenvalues larger than 1. The eigenvectors corresponding to these eigenvalues are given in Table 6. The first eigenvector appears to be a general factor with clerical ability barely included, contrasted with mechanical ability. The second eigenvector is primarily spatial, mechanical, and clerical ability and the third is abstract reasoning contrasted with clerical ability. All three of these factors were consistent with other results in the heritability study thus validating, to some extent, the generalized correction for attenuation method.

Example 2.

The variables for this example are ratings of physical characteristics for androgenicity. The generalized correction for attenuation was used on all sets of data; only the data for males and females at age 18 are included here. The error covariance matrix is the variance due to differences between raters. Table 7 has the observed correlation matrices and Table 8 displays the true correlation matrices after correcting for the attenuating measurement error. The error has decreased all the correlations, some substantially. There were three eigenvalues greater than one for females and two eigenvalues for males. The corresponding eigenvectors were extremely interpretable in terms of the underlying dimension of interest. When the error matrices were simply subtracted from the observed matrices, there were some negative elements on the diagonal in all cases. The diagonal elements in this resulting matrix are identical to the variances which would be obtained with pairwise application of the bivariate correction for attenuation, indicating that this method would have produced some negative variance estimates.

The generalized correction for attenuation described in this paper produces estimates for the true relationships among variables. This multivariate procedure considers the correlated errors and successfully deals with the problem of negative variance estimates.

REFERENCES

Bock, R. Darrell and Vandenberg, Steven G. Components of heritable variation in mental test scores. In Steven G. Vandenberg (Ed.), Progress in Human Behavior Genetics, Baltimore: The Johns Hopkins Press, 1968.

Petersen, Anne C. The relationship of androgenicity in males and females to spatial ability and fluent production. Unpublished PhD Dissertation, University of Chicago, 1973.

Rao, C.R. Linear Statistical Inference and Its Applications. New York: John Wiley and Sons, 1965.

TABLE 4

TRUE CORRELATION MATRIX

	Spat	Numer	Abst	Verb	Mech	Cler	Spell	Sent
Spatial	1.00							
Numerical	-0.15	1.00						
Abstract	0.44	0.21	1.00					
Verbal	0.21	0.69	0.49	1.00				
Mechanical	0.61	-0.71	-0.01	-0.53	1.00			
Clerical	0.42	-0.08	-0.21	0.18	0.49	1.00		
Spelling	0.24	0.49	0.79	0.87	-0.31	0.14	1.00	
Sentences	0.68	0.14	0.26	0.68	0.02	0.41	0.46	1.00

TABLE 5

EIGENVALUES OF TRUE CORRELATION MATRIX

1	3.43
2	2.48
3	1.16
4	0.57
5	0.37
6	0.00
7	0.00
8	0.00

TABLE 6

EIGENVECTORS OF TRUE CORRELATION MATRIX

	1	2	3
Spatial	-0.21	-0.53	-0.15
Numerical	-0.34	0.35	0.25
Abstract	-0.37	-0.07	-0.66
Verbal	-0.52	0.08	0.19
Mechanical	0.21	-0.56	-0.17
Clerical	-0.08	-0.42	0.57
Spelling	-0.50	0.02	-0.17
Sentences	-0.37	-0.31	0.26

TABLE 1

DIZYGOTIC BOYS COVARIANC

	Spat	Numer	Abst	Verb
Spatial	291.32			
Numerical	-3.49	24.94		
Abstract	45.95	2.17	55.94	
Verbal	13.14	10.85	10.53	27.82
Mechanical	49.48	-7.98	3.96	-6.69
Clerical	63.75	2.60	2.13	9.38
Spelling	64.19	25.13	41.98	61.97
Sentences	100.94	11.52	19.21	30.75

TABLE 2

MONOZYGOTIC BOYS COVAR

	Spat	Numer	Abst	Verb
Spatial	71.94			
Numerical	4.31	18.84		
Abstract	19.07	0.45	42.72	
Verbal	1.57	3.01	3.32	14.96
Mechanical	17.77	10.29	8.10	7.47
Clerical	13.93	6.10	10.25	3.02
Spelling	10.60	7.56	-10.84	9.48
Sentences	8.92	13.92	10.84	5.24

TABLE 3

CT = CY - CE

	Spat	Numer	Abst	Verb
Spatial	219.37			
Numerical	-7.80	6.10		
Abstract	26.89	1.71	13.22	
Verbal	11.57	7.85	6.71	12.85
Mechanical	31.70	-18.26	-4.14	-14.16
Clerical	49.82	-3.50	-8.12	6.31
Spelling	53.59	17.57	52.82	52.49
Sentences	92.01	-2.41	8.87	25.51

TABLE 7

CORRELATIONS AMONG PHYSICAL VARIABLES AT AGE 18[a]

	Musc	Shol	Wst	Hips	Butt	Thi	LegS	Calf	OvRt	G/BS	PubH
Muscles/Fat		.73	.58	.69	.63	.49	.42	.55	.78	.41	-.05
Shoulder Width	.82		.42	.54	.54	.32	.27	.39	.59	.37	.00
Waist Indentan	.69	.68		.78	.66	.43	.29	.47	.75	.18	-.18
Hip Width	.50	.52	.57		.87	.58	.50	.65	.86	.41	-.11
Buttocks Shape	.73	.66	.68	.60		.64	.56	.61	.81	.44	.00
Thigh Shape	.56	.54	.64	.56	.66		.80	.66	.64	.45	-.07
Leg Space	.49	.40	.42	.28	.67	.57		.55	.56	.49	-.13
Calf Shape	.63	.59	.61	.41	.62	.56	.42		.62	.34	-.07
Overall Rating	.80	.71	.73	.58	.73	.67	.49	.63		.45	-.10
Gen or Br Size	.45	.50	.39	.37	.47	.43	.36	.46	.46		-.03
Pubic Hair	.19	.20	.09	.26	.12	.11	-.11	.18	.34	.17	

[a] Girls above diagonal, boys below.
(n=99) (n=106)

TABLE 8

CORRECTED CORRELATIONS AMONG PHYSICAL VARIABLES AT AGE 18[a]

	Musc	Shol	Wst	Hips	Butt	Thi	LegS	Calf	OvRt	G/BS	PubH
Muscles/Fat		.77	.76	.92	.94	.77	.64	.88	.92	.69	-.22
Shoulder Width	.99		.63	.81	.79	.23	.19	.49	.74	.38	-.01
Waist Indentan	.70	.65		.92	.90	.51	.20	.70	.88	.35	-.21
Hip Width	.71	69	.76		.97	.58	.40	.81	.97	.62	-.21
Buttocks Shape	.90	.87	.75	.75		.67	.46	.79	.96	.54	-.05
Thigh Shape	.71	.65	.94	.88	.78		.88	.72	.70	.59	-.20
Leg Space	.56	.50	.45	.37	.80	.42		.52	.57	.70	-.22
Calf Shape	.78	.79	.74	.65	.81	.77	.41		.77	.74	-.32
Overall Rating	.91	.90	.76	.89	.84	.85	.39	.76		.67	-.25
Gen or Br Size	.73	.67	.51	.62	.77	.68	.54	.75	.68		-.35
Pubic Hair	.21	.26	-.02	.46	.12	.20	-.26	.11	.49	.08	