Correcting for Restriction of Range in Both *X* and *Y* When the Unrestricted Variances are Unknown

Ralph A. Alexander, Paul J. Hanges, and George M. Alliger
The University of Akron

Correction of correlation coefficients that have arisen from range restricted populations is commonly suggested and practiced in research on testing and measurement. Until recently, that research has operated under two important limitations. First, the majority of the research has dealt with range restriction on one variable only, and second, the correction formulas have assumed that the variance of the variable(s) in the unrestricted population was known. This article presents a method for estimating such corrections from the data in the restricted sample and applies the method to a recently developed approximation for restriction on both *X* and *Y*. The procedure is evaluated and found to produce sufficiently accurate results to be useful in many practical range restriction settings.

The theoretical and applied research on correlation coefficients that have been affected by range restriction has grown to a sizable body of literature since the problem was first summarized by Pearson (1903). Until recently, most of this research has suffered from two limitations. First, the correction of coefficients for range restriction has been limited almost entirely to formulas that assume restriction on only a single variable. Second, such corrections have been applied under the assumption that the unrestricted population variance of one or more of the variables of interest is known or can be reasonably estimated from some source other than the sample data itself.

The correction of sample *r*s for range restriction is by now a widely accepted practice in the psychological testing and measurement literature (American Psychological Association, Division 14, 1980), and such corrected correlations have generally been found to be less biased than the uncorrected values, even under a variety of assumption violations (Linn, 1983). The most frequently used correction formulas have been researched since the early 1920s (Kelley, 1923; Otis, 1922), but were brought into more common use by Thorndike (1947). These correction formulas (usually referred to in the psychological literature as the Thorndike Case 1, Case 2, and Case 3 formulas) are only applicable for correcting the *X*-*Y* correlation when restriction has occurred on either *X* or on *Y* or on some third variable.

To date, very little consideration has been given to the problem of correlations that have arisen from distributions restricted on both *X* and *Y*. Although Pearson (1908) provided a partial solution to this problem, it is only recently that further advances have been made. Weiler (1959) and Regier and Hamdan (1971) developed the mathematics for calculating the range-restricted correlation from properties of the

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 9, No. 3, September 1985, pp. 317–323

0146-6216/85/030317-07$1.55

unrestricted distribution and the extent of truncation on X and Y. However, an exact expression for correcting the range-restricted correlation has not yet been found (Regier & Hamdan, 1971). Wells and Fruchter (1970) presented an approximate correction procedure for this case, and the procedure has been found to perform reasonably well (Alexander, Carson, Alliger, & Barrett, 1984). Alexander, Carson, Alliger, and Carr (1984) derived an improved approximation which they reported yields corrected values within ±1% of the true population values under a wide variety of cuts on both X and Y. That approximation formula is:

$$r = \frac{r'^2 - 1}{2r'} U_x U_y \pm \left[\frac{(1 - r'^2)^2}{4r'^2} U_x^2 U_y^2 + 1 \right]^{1/2} , \quad (1)$$

where r' is the range-restricted correlation,
U_x and U_y are the ratios of restricted to unrestricted standard deviations in X and Y, respectively (e.g., $U_x = S'_x/S_x$), and
the "±" sign takes the sign of r'.

As is the case in the Thorndike corrections for unidimensional range restriction, Equation 1 requires knowledge of the unrestricted standard deviations (e.g., S_x) of both variables. Virtually all of the prior literature on restriction of range has estimated this value from population norms or from assumed values of the truncation point on the variable (see, e.g., Alexander, Carson, Alliger, & Barrett, in press; Sackett & Wade, 1983; Schmidt, Hunter, & Urry, 1976). Recently, Alexander, Alliger, and Hanges (1984) noted that, at times, such information is not available, and they reported a procedure for estimating the unrestricted standard deviation directly from the restricted sample data when restriction has occurred on one of the two variables of interest (X or Y). They reported that the procedure is quite accurate in recapturing the underlying correlation and that the sampling variance of such corrected rs is only slightly greater than that of rs corrected under the assumption of a known population variance.

The purpose of this article is to extend the Alexander, Alliger, and Hanges (1984) method to the estimation of unrestricted standard deviations in both X and Y when both variables are truncated. These values are then applied to Equation 1, and the accuracy and sampling variance of the corrected rs is examined.

Method

Cohen (1959) showed that for a sample drawn from a bivariate normal X-Y distribution truncated, for example, on X, there is a unique sample statistic that will provide information about the restriction. This is referred to as Cohen's Ratio (CR):

$$CR_x = S'^2_x/(\bar{X} - X_c)^2 , \quad (2)$$

where S'^2_x is the variance,
$\bar{X}$ is the mean, and
X_c is the lowest observed value,

all observed on X in the restricted sample. (Throughout this discussion it is assumed that truncation has occurred from below. If truncation occurs from above, the highest rather than the lowest observed value would be used.) Alexander, Alliger, and Hanges (1984) tabled two functions of the truncated normal distribution of X and used these tabled values to correct for unidimensional truncation. The two functions are the z-score representing the truncation point and the standard deviation of X in standardized form in the restricted distribution.

The extension of this method to bidimensional truncation (restriction on both X and Y) is straightforward. Cohen's Ratio on Y becomes:

$$CR_y = S_y'^2/(\bar{Y} - Y_c)^2 \quad . \tag{3}$$

For any given pair of cutting scores z_x on X and z_y on Y, there will be a unique pair of CRs and standardized standard deviations. Since each of the restricted standardized standard deviations represents the proportional reduction in the standard deviation due to range restriction, each is equal to the ratio of the restricted to unrestricted standard deviation—that is, U_x and U_y.

In order to use this method, then, it is necessary to first construct a table of these values. In testing the method for this study, a computer program was written that constructed such a table for all possible combinations of cuts on X and Y ranging from -2σ to $+2\sigma$ in increments of $.10\sigma$. For each pair of truncation points, the values of CR_x, CR_y, U_x, and U_y were calculated. Since that table is far too lengthy to reproduce here, selected values are shown in Table 1.

The range restriction correction method proceeds as follows. From the empirical sample at hand, calculate r_{xy}, and CR_x, CR_y (from Equations 2 and 3). Find the pair of tabled CR values ($CR_{(TAB)}$) that is closest to the pair of CR values calculated from the sample ($CR_{(OBS)}$). This is done by finding the pair of tabled values that minimizes the sum of the squared differences (i.e., MIN[$(CR_{x(OBS)} - CR_{x(TAB)})^2 + (CR_{y(OBS)} - CR_{y(TAB)})^2$]). (A computer program is available from the first author that constructs the described table, calculates the sample values needed, finds the minimum function and displays the associated values.)

Table 1
Selected Values of Cohen's Ratio and Corresponding Restricted Standard Deviation of z Score for Truncation Point

Cohen's Ratio on X	Cohen's Ratio on Y	U = S'/S		Truncation Points	
$S_x'^2/(\bar{X}-X_c)^2$	$S_y'^2/(\bar{Y}-Y_c)^2$	U_x	U_y	z_{xc}	z_{yc}
.204	.204	.936	.936	-2.0	-2.0
.180	.375	.920	.793	-2.0	-1.0
.140	.569	.903	.603	-2.0	0.0
.104	.722	.890	.446	-2.0	1.0
.375	.180	.793	.920	-1.0	-2.0
.348	.348	.796	.796	-1.0	-1.0
.287	.552	.813	.608	-1.0	0.0
.215	.715	.841	.448	-1.0	1.0
.569	.140	.603	.903	0.0	-2.0
.552	.287	.608	.813	0.0	-1.0
.498	.498	.633	.633	0.0	0.0
.412	.685	.687	.462	0.0	1.0
.722	.104	.446	.890	1.0	-2.0
.715	.215	.448	.841	1.0	-1.0
.685	.412	.462	.687	1.0	0.0
.622	.622	.502	.502	1.0	1.0

Once the closest pair of tabled CR values is found, the U_x and U_y values associated with that pair are used in Equation 1.

The performance of this method was tested with computer simulations. Using the *IMSL Library* (1979), bivariate normal samples were generated for $\rho = .2, .4, .5, .6$, and $.8$; n-sizes of 30, 60, and 100; and truncation points on X and Y of $z = -2, -1, 0$, and $+1$. Each simulation was run with 2,000 random samples. For each random sample r_{xy}, CR_x, and CR_y were calculated, the closest pair of tabled values of CR was found and Equation 1 used to compute the corrected r.

The algorithms needed to calculate CR_x, CR_y, U_x, and U_y under bidimensional truncation are reported elsewhere (Alexander, Carson, Alliger, & Barrett, 1984, in press) and in the interest of brevity are not repeated here. Those algorithms require knowledge of the underlying unrestricted population ρ in order to compute functions of a bidimensionally truncated distribution. The simulations reported here all used an assumed value of $\rho = .50$. These simulations were repeated for several other values of ρ and the results were very similar. This is due to the fact that the CR and U values vary only slightly as a function of ρ except for extreme ρ (approaching .05 or .95).

Results and Discussion

Table 2 contains the mean observed and corrected correlations as a function of ρ and truncation points for sample sizes of 60. Under moderate truncation on both variables, the procedure overcorrects slightly (maximum mean overcorrection is .02); however, this tendency decreases as the population ρ increases. If only one variable is severely truncated while the other is moderately restricted, the procedure is again relatively accurate (maximum mean deviation is $-.04$). The procedure tends to show maximum overcorrection when the population ρ is close to .50. However, in general, as the truncation on one variable becomes more severe and as the population ρ approaches .80, the procedure undercorrects. In the case of extreme restriction on both variables, the procedure consistently undercorrects (maximum mean undercorrection is $-.06$); the severity of this undercorrection diminishes, however, as the population ρ increases.

Thus, the procedure outlined in this paper accurately recaptures, on the average, the true population ρ for sample sizes of 60. The results were comparable for the simulations with sizes of 30 and 100, though the maximum mean deviation from the true ρ did differ (maximum mean deviation for $n = 30$ was $-.11$, maximum mean deviation for $n = 100$ was $-.04$). The method overcorrects for slight truncation and low ρ and overcorrects for severe restriction on both variables and low ρ.

Table 3 presents the standard deviations for the uncorrected correlations, the corrected correlations when the population variance is known, and the corrected correlations when the population variance is estimated with the present procedure for varying truncation points, population ρ, and sample sizes. Although the current procedure does increase the standard deviations above those for the observed uncorrected correlations, these values are only slightly greater for the proposed method than when the population variance is assumed to be known. In fact, the maximum inflation of the standard deviation is only .07 for $n = 30$, .05 for $n = 60$, and .04 for $n = 100$. When $\rho = .80$, the mean standard deviations for the correction formulas are almost equal to those for the uncorrected correlations. Thus, the present procedure not only adequately corrects correlation coefficients in the more common case when the population variance is unknown, but it does so with a minimal increase in variability from currently used procedures when knowledge of the population variance is necessary.

Conclusions

This paper extended the Alexander, Alliger, and Hanges (1984) univariate restriction of range correction procedure to the bidimensional case. The results of the simulations indicate that correction for

Table 2
Computer Simulation Results: Comparison of Mean Observed and Corrected Correlations for Differing ρ and Truncation Points

Truncation Point on X	Truncation Point on Y			
	-2.0	-1.0	0.0	1.0
For ρ = .2				
-2.0	.18 (.22)	.15 (.22)	.12 (.22)	.09 (.21)
-1.0		.13 (.22)	.10 (.22)	.08 (.20)
0.0			.08 (.21)	.07 (.19)
1.0				.05 (.14)
For ρ = .4				
-2.0	.36 (.42)	.32 (.43)	.25 (.43)	.19 (.40)
-1.0		.29 (.43)	.23 (.43)	.18 (.39)
0.0			.19 (.42)	.15 (.38)
1.0				.12 (.35)
For ρ = .5				
-2.0	.46 (.52)	.41 (.53)	.32 (.52)	.25 (.49)
-1.0		.38 (.53)	.31 (.52)	.24 (.49)
0.0			.26 (.52)	.22 (.49)
1.0				.18 (.46)
For ρ = .6				
-2.0	.56 (.62)	.50 (.62)	.41 (.61)	.32 (.58)
-1.0		.47 (.63)	.40 (.61)	.31 (.58)
0.0			.35 (.61)	.29 (.58)
1.0				.24 (.56)
For ρ = .8				
-2.0	.77 (.81)	.72 (.81)	.62 (.79)	.51 (.76)
-1.0		.70 (.81)	.62 (.79)	.51 (.76)
0.0			.59 (.79)	.50 (.76)
1.0				.46 (.76)

Note. Corrected values are in parentheses. Data based on an n-size of 60 with 2,000 iterations.

correlations when the population unrestricted variances are unknown may be useful. These results will, of course, strictly hold only for bivariate normal distributions under conditions of strict truncation. Further research is needed to determine the robustness of the procedure under violations of these assumptions. The accuracy of this new procedure appears to be comparable to the currently used procedures that assume knowledge of the unrestricted population variances.

References

Alexander, R. A., Alliger, G. M., & Hanges, P. J. (1984). Correcting for range restriction when the population variance is unknown. *Applied Psychological Measurement, 8,* 431–437.

Alexander, R. A., Carson, K. P., Alliger, G. M., & Barrett, G. V. (1984). Correction for restriction of range when both *X* and *Y* are truncated. *Applied Psychological Measurement, 8,* 231–241.

Alexander, R. A., Carson, K. P., Alliger, G. M., & Barrett, G. V. (in press). Further considerations of the power to detect non-zero validity coefficients under range restriction. *Journal of Applied Psychology.*

Table 3
Standard Deviations of Restricted Uncorrected Correlations and Correlations Corrected Using Known Variance and Estimated Variance (for ρ = .50)

Truncation Point on X*	Truncation Point on Y											
	-2.0			-1.0			0.0			1.0		
	n_1=30	n_2=60	n_3=100	n_1=30	n_2=60	n_3=100	n_1=30	n_2=60	n_3=100	n_1=30	n_2=60	n_3=100
-2.0												
a	149	105	081	157	111	087	162	122	092	176	126	098
b	144	098	075	163	111	084	202	145	105	286	192	141
c	164	113	087	187	128	099	230	166	128	298	211	162
-1.0												
a				164	116	091	172	128	095	181	130	101
b				183	122	094	229	159	115	300	206	146
c				208	142	109	257	180	137	311	224	171
0.0												
a							185	131	102	193	136	104
b							286	189	139	363	247	184
c							316	214	160	374	260	199
1.0												
a										196	139	111
b										423	299	238
c										426	310	248

Note. Based on 2,000 replications. Decimals are omitted. Tabled values for ρ = .2, .4, .6, .8 are available on request.
*a = standard deviation of range restricted, uncorrected r. b = standard deviation of corrected r assuming unrestricted population variance (σ_X) is known. c = standard deviation of corrected r assuming unrestricted population variance (σ_X) is not known.

Alexander, R. A., Carson, K. P., Alliger, G. M., & Carr, L. (1984). *An improved approximation for correcting the bivariate normal correlation when truncation has occurred on both variables*. Manuscript submitted for publication.

American Psychological Association, Division of Industrial Organizational Psychology. (1980). *Principles for the validation and use of personnel selection procedures* (2nd ed.). Dayton OH: Author.

Cohen, A. C. (1959). Simplified estimators for the normal distribution when samples are singly censored or truncated. *Technometrics, 1,* 217–237.

IMSL Library. (1979). Houston TX: International Mathematical and Statistical Libraries.

Kelley, T. L. (1923). *Statistical method*. New York: Macmillan.

Linn, R. L. (1983). Pearson selection formulas: Implications for studies of predictive bias and estimates of educational effects in selected samples. *Journal of Educational Measurement, 20,* 1–15.

Pearson, K. (1903). Mathematical contributions to the theory of evolution-XI. On the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society, London, Series A, 200,* 1–66.

Pearson, K. (1908). On the influence of double selection on the variation and correlation of two characters. *Biometrika, 6,* 111–112.

Otis, A. A. (1922). A method of inferring the change in a coefficient of correlation resulting from a change in the heterogeneity of the group. *Journal of Educational Psychology, 13,* 293–294.

Regier, M. H., & Hamdan, M. A. (1971). Correlation in a bivariate normal distribution with truncation in both variables. *Australian Journal of Statistics, 13,* 77–82.

Sackett, P. R., & Wade, B. E. (1983). On the feasibility of criterion-related validity: The effects of range restriction assumption on needed sample size. *Journal of Applied Psychology, 68,* 374–381.

Schmidt, F. L., Hunter, J. E., & Urry, V. W. (1976). Statistical power in criterion-related validation studies. *Journal of Applied Psychology, 61,* 473–485.

Thorndike, R. L. (1947). *Research problems and techniques* (Report No. 3). Washington DC: AAF Aviation Psychology Program Research Reports, U.S. Government Printing Office.

Weiler, H. (1959). Means and standard deviations of a truncated normal bivariate distribution. *Australian Journal of Statistics, 1,* 73–81.

Wells, D. G., & Fruchter, B. (1970). Correcting the correlation coefficient for explicit restriction in both variables. *Educational and Psychological Measurement, 30,* 925–934.

Author's Address

Send requests for reprints or further information to Ralph A. Alexander, Department of Psychology, The University of Akron, Akron OH 44325, U.S.A.