A Note on Confidence Intervals for Two-Group Latent Mean Effect Size Measures

Jaehwa Choi  
The George Washington University

Weihua Fan  
University of Houston

Gregory R. Hancock  
University of Maryland, College Park

This note suggests delta method implementations for deriving confidence intervals for a latent mean effect size measure for the case of 2 independent populations. A hypothetical kindergarten reading example using these implementations is provided, as is supporting LISREL syntax.

As the importance of population effect size estimates has become increasingly recognized in recent years, so too has the need for confidence intervals around those effect size estimates (see, e.g., Cumming & Finch, 2001; Olejnik & Algina, 2000; Roberts & Henson, 2002; Smithson, 2001; Steiger & Fouladi, 1997; B. Thompson, 2002). Steiger (1990), for example, presented a general procedure for constructing confidence intervals around measures of effect size in univariate analysis of variance designs (see also Fleishman, 1980; Hedges & Olkin, 1985; Steiger, 2004). Interestingly, although there is much discussion of effect size measures for experimental designs involving observed variables...
(with many different types of effect size measures being proposed over the past 2 decades), only fairly recently were some of these principles formally extended to experimental design at a latent variable level.

Specifically, given the simplicity of Cohen’s (1988) experimental design effect size measures and their pervasiveness throughout the social and behavioral sciences, Hancock (2001) proposed parallel effect size measures for between-groups hypothesis testing of means on a single latent construct within the structured means modeling (SMM) framework (Sörbom, 1974). Ultimately, the focus of that work was on such effect size measures’ relation to power analysis and sample size determination rather than the estimation of the effect sizes themselves. In this note, we return to the issue of effect size estimation and propose strategies for constructing confidence intervals around the proposed effect size measure for latent means models. Interval estimates for the case of two populations are addressed and a hypothetical kindergarten reading example, investigating the difference between a traditional phonics-based curriculum and a whole language curriculum, is presented.

**BRIEF REVIEW OF SMM AND LATENT MEAN EFFECT SIZE MEASURES WITH TWO POPULATIONS**

For a single-factor model with \( p \) observed indicators \( (Y) \) of construct \( \eta \), the \( Y \) values may be expressed in a \( p \times 1 \) vector \( y \) as follows: \( y = \tau + \Lambda \eta + \varepsilon \). The \( p \times 1 \) vector \( \tau \) contains \( p \) intercept values, the \( p \times 1 \) vector \( \Lambda \) contains variable loadings on a single construct, and \( \varepsilon \) is a \( p \times 1 \) vector of normal errors. Thus, the first moment vector \( \mu = E[y] = \tau + \Lambda E[\eta] = \tau + \Lambda \kappa \), where \( \kappa \) is the construct mean on \( \eta \). The second moment matrix, assuming constructs and the errors to be independent, is \( E[(y - \mu)(y - \mu)'] = \Sigma = \Lambda \Phi \Lambda' + \Theta \). The scalar \( \Phi \) is the construct variance and \( \Theta \) is the \( p \times p \) covariance matrix of the errors in \( \varepsilon \). Given first and second moment data from multiple \((J)\) samples, and appropriate cross-population parameter constraints reflecting assumptions of intercept \((\tau)\) and loading \((\Lambda)\) invariance (or partial conditions thereof; Byrne, Shavelson, & Muthén, 1989), model-based inferences using SMM can be made about latent means for the purpose of testing the null hypothesis \( H_0: \kappa_1 = \ldots = \kappa_J \). The unfamiliar reader is referred to didactic treatments of this topic (e.g., Bollen, 1989; Hancock, 2004; M. Thompson & Green, 2006).

For the case of \( J = 2 \) populations, the SMM technique offers a test of construct mean equivalence: \( H_0: \kappa_1 = \kappa_2 \). For the between-subjects case, Hancock (2001) recommended a standardized effect size measure of

\[
d = \frac{|\kappa_1 - \kappa_2|}{\sqrt{\Phi}},
\]
where $\phi$ is the variance of the construct $\eta$ if one assumes homogeneity across both populations. The $d$ value may be estimated from samples of size $n_1$ and $n_2$, with or without the variance homogeneity assumption, as

$$d = \frac{|\hat{\kappa}_1 - \hat{\kappa}_2|}{\sqrt{\left(\frac{n_1}{n_1 + n_2}\right)\hat{\phi}_1 + \left(\frac{n_2}{n_1 + n_2}\right)\hat{\phi}_2}},$$

(2)

where $\hat{\kappa}_1$ and $\hat{\kappa}_2$ are sample means on the construct $\eta$ and $\hat{\phi}_1$ and $\hat{\phi}_2$ are sample construct variances. Because the limits of identification in the means portion of a two-population model require that one population serve as a reference with its latent mean constrained, customarily to zero (e.g., $\kappa_1 = 0$), Equation (2) reduces to

$$d = \frac{|\hat{\kappa}_2|}{\sqrt{\left(\frac{n_1}{n_1 + n_2}\right)\hat{\phi}_1 + \left(\frac{n_2}{n_1 + n_2}\right)\hat{\phi}_2}},$$

(3)

This $d$ estimate may be interpreted just as in univariate analyses; for example, a value of $\hat{d} = .25$ would indicate that the two population means on the latent construct $\eta$ are estimated to be one fourth of an error-free (i.e., latent) standard deviation apart along the latent $\eta$ continuum.

CONFIDENCE INTERVALS FOR LATENT MEAN EFFECT SIZE WITH TWO POPULATIONS

The construction of the confidence intervals of interest is quite feasibly facilitated using common structural equation modeling (SEM) software. Constructing such intervals relies on the delta method (see, e.g., Bishop, Feinberg, & Holland, 1988; Raykov & Marcoulides, 2004; Wasserman, 2004) as has been utilized in other SEM contexts (e.g., maximal reliability; Raykov, 2004). Here the results of the delta method will be implemented in two ways, the first analytical (i.e., by hand) and the second numerical (i.e., using software). The delta method uses (typically) a first-order Taylor series expansion to approximate the moments of a differentiable function $g$ of random variables $\theta$ when direct evaluation of the expectation is not feasible (Cramér, 1946). The delta method assumes that $g$ is smooth (in our case, differentiable in a neighborhood of the maximum likelihood estimates) and polynomially bounded. Technical details of these assumptions are beyond the scope of this note, and readers should consult Cramér for more
information regarding smoothness and polynomial bounding. Generically, the variance approximation is

$$\sigma^2_{g(\hat{\theta})} \approx [\nabla g(\theta)^T] \Sigma_{\theta} [\nabla g(\theta)].$$  \hspace{1cm} (4)$$

where $\nabla g(\theta)$ is the gradient vector of the function $g$ and $\Sigma_{\theta}$ is the asymptotic covariance matrix of the elements in $\theta$. This variance gets estimated, then, as

$$\hat{\sigma}^2_{g(\hat{\theta})} \approx [\nabla g(\hat{\theta})^T] \hat{\Sigma}_{\hat{\theta}} [\nabla g(\hat{\theta})].$$  \hspace{1cm} (5)$$

This approach is applied as follows in the context of latent mean effect size with two populations.

**Delta Method, Analytical Implementation**

Assuming $\kappa_1 = 0$ for identification, let us define the function $g$ as

$$g(\kappa_2, \phi_1, \phi_2) = \frac{\kappa_2}{\sqrt{\left(\frac{n_1}{n_1 + n_2}\right) \phi_1 + \left(\frac{n_2}{n_1 + n_2}\right) \phi_2}}.$$  \hspace{1cm} (6)$$

Note that this is equivalent to the previous $d$ but without the absolute value. This definition of $\hat{d} = g(\kappa_2, \phi_1, \phi_2)$ is used for the remainder of this note, allowing the incorporation of the direction of the effect as well as alleviating the mathematical inconvenience of the absolute value. That said, because the target function $g$ is not a linear composite of the parameters $\kappa_2, \phi_1,$ and $\phi_2$, deriving an analytical expression for the variance of the target function is not trivial. As mentioned previously, in such cases one often approximates the variance of $g$ using the delta method’s first-order Taylor series expansion. That is, we can analytically express the (approximated) variance of $g$ as a function of the variance and covariance of $\kappa_2, \phi_1,$ and $\phi_2$, which can be obtained from conventional SEM program output. The approximate variance for $g(\kappa_2, \phi_1, \phi_2)$ via the delta method is given by the expression

$$\sigma^2_{g(\kappa_2, \phi_1, \phi_2)} \approx \begin{bmatrix} \frac{\partial g}{\partial \kappa_2} & \frac{\partial g}{\partial \phi_1} & \frac{\partial g}{\partial \phi_2} \end{bmatrix} \begin{bmatrix} \sigma^2_{\kappa_2} & \sigma_{\phi_1, \kappa_2} & \sigma_{\phi_2, \kappa_2} \\ \sigma_{\phi_1, \kappa_2} & \sigma^2_{\phi_1} & \sigma_{\phi_2, \phi_1} \\ \sigma_{\phi_2, \kappa_2} & \sigma_{\phi_2, \phi_1} & \sigma^2_{\phi_2} \end{bmatrix} \begin{bmatrix} \frac{\partial g}{\partial \kappa_2} \\ \frac{\partial g}{\partial \phi_1} \\ \frac{\partial g}{\partial \phi_2} \end{bmatrix}.$$  \hspace{1cm} (7)$$
where $\sigma^2_{\hat{k}_2}, \sigma^2_{\hat{\phi}_1},$ and $\sigma^2_{\hat{\phi}_2}$ are the asymptotic parameter estimate variances (squared standard errors) and $\sigma^2_{\hat{\phi}_1, \hat{k}_2}, \sigma^2_{\hat{\phi}_2, \hat{k}_2},$ and $\sigma^2_{\hat{\phi}_2, \hat{\phi}_1}$ are the asymptotic covariances as contained in the matrix $\mathbf{\Sigma}_{\hat{k}_2, \hat{\phi}_1, \hat{\phi}_2}$. The estimates of these variances and covariances may be obtained from the asymptotic parameter estimate covariance matrix typically available as optional output in standard SEM software packages.

The gradient vector of the function $g$ in the current case, evaluated at the parameter estimates $(\hat{k}_2, \hat{\phi}_1, \hat{\phi}_2)$, is

$$\nabla G(\hat{k}_2, \hat{\phi}_1, \hat{\phi}_2) = \begin{bmatrix} \frac{1}{2} \hat{k}_2 \left( \frac{n_1}{n_1 + n_2} \right) \left[ \left( \frac{n_1}{n_1 + n_2} \right) \hat{\phi}_1 + \left( \frac{n_2}{n_1 + n_2} \right) \hat{\phi}_2 \right]^{-3/2} \\ -\frac{1}{2} \hat{k}_2 \left( \frac{n_1}{n_1 + n_2} \right) \left[ \left( \frac{n_1}{n_1 + n_2} \right) \hat{\phi}_1 + \left( \frac{n_2}{n_1 + n_2} \right) \hat{\phi}_2 \right]^{-1/2} \end{bmatrix}.$$ (8)

Substituting this expression into Equation (7) in turn defines an analytical expression for the estimated standard error of $\hat{d}, \hat{\sigma}_d$, where

$$\hat{\sigma}_d = \sqrt{\hat{\sigma}^2_g(\hat{k}_2, \hat{\phi}_1, \hat{\phi}_2)}.$$ (9)

The associated $(1 - \alpha) \times 100\%$ confidence interval for $\hat{d}$, with lower (L) and upper (U) boundaries, is therefore

$$[\hat{d}_L, \hat{d}_U] = [\hat{d} - z_{(1-\alpha/2)}\hat{\sigma}_d, \hat{d} + z_{(1-\alpha/2)}\hat{\sigma}_d],$$ (10)

with the interval based on the assumption of normality for the estimated effect size sampling distribution. A hypothetical kindergarten reading example of this analytical implementation of the delta method for standard error and confidence interval estimation is presented in the following section.

**Delta Method, Numerical Implementation (Using LISREL)**

The estimation of standard errors, and in turn the construction of confidence intervals, for latent mean effect size for the two-population case is facilitated by the popular SEM software LISREL (Jöreskog & Sörbom, 2006), as has been done in other contexts (e.g., maximal reliability; Raykov, 2004). By specifying the desired effect size as an additional (auxiliary) parameter within the program, the software augments the parameter vector, computes an estimate for the effect size (i.e., the additional parameter), and automatically determines the standard
CONFIDENCE INTERVALS FOR LATENT MEAN EFFECT SIZES

error estimate using the inverse of the augmented information matrix (which is also estimated by a delta method but using numerical rather than analytical derivatives). This process occurs without affecting model estimation or fit in any way because this auxiliary parameter does not have any implications for the mean or covariance structures of the existing model. The variability associated with this effect size parameter estimate is presented in the requested output, from which a symmetric confidence interval may be computed manually as per Equation (10) using the additional parameter’s estimated standard error. A hypothetical kindergarten reading example of this numerical implementation of the delta method, including sample LISREL syntax, is presented in the following section.

EXAMPLE USING BOTH ANALYTICAL AND NUMERICAL IMPLEMENTATIONS OF THE DELTA METHOD

Hancock (2004) presented summary statistics (i.e., correlations, standard deviations, and means) for a hypothetical quasi-experimental example to illustrate the SMM approach to latent mean differences. Two intact samples of 500 kindergartners were assumed, the first assigned to receive a traditional phonics-based curriculum for 2 years and the second to receive a whole-language curriculum for 2 years. At the end of first grade, three reading assessments were made, focusing on vocabulary (Voc), comprehension (Comp), and language (Lang). These variables were all believed to be observable manifestations of an underlying construct of reading proficiency. Summary statistics for these two samples’ fabricated data are reproduced in Table 1 (note that no raw data existed for Hancock’s example, only the summary statistics).

A structured means model was fit to the summary statistics in Table 1, with all loadings and intercepts constrained equal across populations, with the first indicator (Voc) selected as the scale referent in both populations (i.e., its loading fixed to 1) and with the phonics-based curriculum population treated as the reference population (with its latent mean fixed to 0). Maximum likelihood estimation was used within LISREL 8.8, achieving satisfactory data-model fit (e.g., $\chi^2 = 4.039$, $df = 4$, RMSEA = .005, CFI = 1.000). Key parameter estimates from the standard portion of the output (reported here to three decimal places) included $\hat{\kappa}_2 = -1.871$ (for which $z = -6.804$, $p < .05$), $\hat{\phi}_1 = 16.453$, and $\hat{\phi}_2 = 16.373$, which yield a standardized effect size estimate of $\hat{d} = -.462$; further, by substitution into Equation (8), these lead to an estimated gradient vector of

$$\nabla g(\hat{\kappa}_2, \hat{\phi}_1, \hat{\phi}_2) = \begin{bmatrix} .24683 \\ .00704 \\ .00704 \end{bmatrix}. \quad (11)$$
TABLE 1
Contrived Summary Statistics for Hypothetical Two-Group Example

<table>
<thead>
<tr>
<th></th>
<th>Phonics ($n_1 = 500$)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Voc</td>
<td>Comp</td>
<td>Lang</td>
<td>SD</td>
</tr>
<tr>
<td>Voc</td>
<td>1.000</td>
<td></td>
<td></td>
<td>4.654</td>
</tr>
<tr>
<td>Comp</td>
<td></td>
<td>.770</td>
<td>1.000</td>
<td>3.943</td>
</tr>
<tr>
<td>Lang</td>
<td></td>
<td>.650</td>
<td>.659</td>
<td>3.771</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Whole-Language ($n_2 = 500$)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Voc</td>
<td>Comp</td>
<td>Lang</td>
<td>SD</td>
</tr>
<tr>
<td>Voc</td>
<td>1.000</td>
<td></td>
<td></td>
<td>4.687</td>
</tr>
<tr>
<td>Comp</td>
<td></td>
<td>.747</td>
<td>1.000</td>
<td>3.901</td>
</tr>
<tr>
<td>Lang</td>
<td></td>
<td>.641</td>
<td>.631</td>
<td>4.083</td>
</tr>
</tbody>
</table>

Additional relevant values are the estimated asymptotic parameter estimate variances and covariances associated with $\hat{\xi}_2$, $\hat{\phi}_1$, and $\hat{\phi}_2$, which may be expressed in

$$
\hat{\Sigma}_{\hat{\xi}_2,\hat{\phi}_1,\hat{\phi}_2} = \begin{bmatrix}
.07564 & -.01869 & -.01830 \\
-.01869 & 1.72324 & .32590 \\
-.01830 & .32590 & 1.73118
\end{bmatrix}.
$$

From these values, it follows from the matrix multiplication in Equation (7) that $\hat{d} = .068$. Following Equation (10), this would in turn yield an estimated 90% confidence interval for $d$ of $[\hat{d}_L, \hat{d}_U] = [-.574, -.349]$.

The aforementioned analytical implementation would be expected to be viable regardless of the SEM software package employed. By comparison, to facilitate direct estimation we may use the auxiliary parameter feature in LISREL, the syntax for which is shown in the Appendix. In addition to all the same parameter estimates presented earlier, the resulting direct estimate of $d$ is $-.462$, matching the previously mentioned hand calculation. For $\hat{d}$, the resulting estimate is $0.068$, identical to three decimal places to the value of $0.068$ obtained earlier using the analytical implementation. Following Equation (10), then, this direct strategy yields the same 90% confidence interval for $d$ (to three decimal places) of $[-.574, -.349]$, as would generally be expected.

CONCLUSIONS

Following work on effect size measures and their point estimates for latent means models (Hancock, 2001), this note has offered a delta method approach
for deriving interval estimates for the latent effect sizes with two independent groups. The necessary standard error can be estimated either analytically or numerically, where that standard error estimate in turn facilitates the construction of a symmetric \((1 - \alpha) \times 100\%\) confidence interval for \(d\) (assuming normality of the effect size estimate’s sampling distribution). If one has access to software allowing the construction of additional parameters, then the numerical implementation seems preferred if only for simplicity; for users of SEM software packages without this feature the analytical implementation should be completely satisfactory as well. As the analytical implementation determines parameter estimate standard errors using the same numerical approximation method as employed by LISREL’s auxiliary parameter function, merely involving a matrix without the augmentation of the additional parameter, one would expect these methods to be in agreement within practical decimal accuracy.

Of course, the accuracy of the interval estimate strategies depends on the reasonableness of underlying assumptions, which include those related to proper fit function specification vis-à-vis the distribution of the data, adequacy of sample size, and moderateness of data-model misfit. Other foundational assumptions involve the models themselves, in particular their homogeneity across populations. First, although SMM does not require homogeneous construct variances, for the effect size measure to have useful interpretation those variances should be reasonably homogeneous; otherwise, a measure of latent mean separation in a metric that does not hold even approximately across populations becomes tenuous in meaning. Second, invariance of loadings and intercepts across populations are typically assumed, at least to start, although many researchers will relax these assumptions if modification indices so suggest, yielding a context of partial invariance within which to assess latent means. Should partial invariance exist, the only requirement for accuracy of the estimated standard error and confidence interval methods presented here would be model propriety. That is, as long as any population noninvariance is appropriately reflected in the absence of corresponding cross-group constraints (while still preserving model identification), then the model is correct and hence the point and interval estimates should be reasonable. Once the population invariance status is not mirrored by the model, however, the reasonableness of parameters’ point and interval estimates becomes questionable, whether or not the parameters in question are auxiliary.

Finally, as with other delta method implementations in SEM, the methods proposed in this note rest on the distributional assumptions inherent to the maximum likelihood estimator. Specifically, although parameter estimates still tend to be unbiased if data are not multivariate normal, the standard errors of parameter estimates tend to be compromised. In terms of this note, this implies that, although \(\hat{d}\) should be unbiased, its standard error estimate \(\hat{\sigma}_d\) could be inaccurate, leading in turn to an inaccurate interval estimate for \(d\).
In such cases where the assumption of multivariate normality is not reasonable, the researcher might therefore consider utilizing the more robust standard error estimates proposed by Satorra and Bentler (1994), and available in most SEM software packages, in the derivation of more accurate interval estimates for the latent effect size.

To close, in light of the simple approaches offered herein, our hope is that applied researchers employing latent means models will not just summarize their results using effect size estimates but will also couch those estimates within appropriate intervals. In fact, both point and interval estimates for latent mean effect sizes could easily be built into existing SEM software and would assist researchers in interpreting and communicating the results of their multiple sample latent means models. Extensions to the more general case of \( J \geq 2 \) populations, where effect size work exists (Hancock, 2001) but for which the necessarily asymmetric confidence intervals appear to become considerably more challenging, remain to be addressed.

REFERENCES


### APPENDIX

**Sample LISREL Code for the Numerical Implementation Using the Additional Parameter Feature**

The syntax shown here is for LISREL 8.8, which, unlike previous versions, allows embedded parentheses within the `COMPUTE` command line.

Latent means model: Group One

```
DA NG=2 NI=3 NO=500
KM SY
1.000
   .770 1.000
   .650 .659 1.000
SD
4.654 3.943 3.771
ME
70.944 69.606 66.962
```
MO NX=3 NK=1 LX=FU,FI TD=DI,FR PH=SY,FR TX=FR KA=FI
FR LX(2,1) LX(3,1)
VA 1 LX(1,1)
OU ND=5 ALL

Latent means model: Group Two
DA NO=500
KM SY
  1.000
  .747 1.000
  .641 .631 1.000
SD
4.687 3.901 4.083
ME
69.166 68.046 65.400
MO LX=IN TD=DI,FR PH=SY,FR TX=IN KA=FR AP=1
CO PA(1)=KA(2,1)/((500*PH(1,1,1)+500*PH(2,1,1))/1000)**.5
!PA(1) is the effect size estimate of d
OU ND=5 ALL