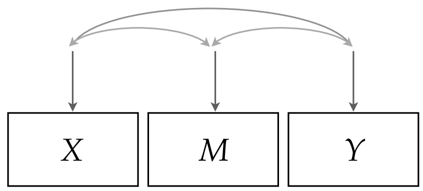
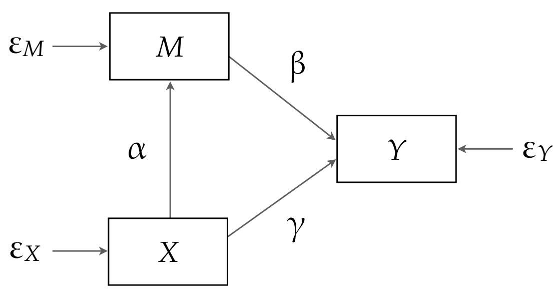
FIML Estimation for a Path Analysis Model

We previously examined full information maximum likelihood (FIML; i.e., raw-data) estimation for a multivariate normal model with a mean vector and covariance matrix, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, as its parameters. For three variables, X, M, and Y, the path diagram for this so-called saturated model is as follows.



A path model imposes structure on the associations, essentially replacing double-headed curved arrows (covariances) with single-headed straight ones (regression slopes). This illustration will use the basic single-mediator model shown in the path diagram below.



The corresponding univariate regression equations are

$$X_i = I_X + \varepsilon_{Xi} \tag{1}$$

$$M_i = I_M + \alpha X_i + \varepsilon_{Mi} \tag{2}$$

$$Y_i = I_Y + \gamma X_i + \beta M_i + \varepsilon_{Yi} \tag{3}$$

and the matrix version of the model is.

$$\begin{pmatrix} X_i \\ M_i \\ Y_i \end{pmatrix} = \begin{pmatrix} I_X \\ I_M \\ I_Y \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ \alpha & 0 & 0 \\ \gamma & \beta & 0 \end{pmatrix} \begin{pmatrix} X_i \\ M_i \\ Y_i \end{pmatrix} + \begin{pmatrix} \varepsilon_{Xi} \\ \varepsilon_{Mi} \\ \varepsilon_{Yi} \end{pmatrix} = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{Y}_i + \boldsymbol{\varepsilon}_i$$
$$\boldsymbol{\varepsilon}_i \sim N_3(0, \boldsymbol{\Psi}) \qquad \boldsymbol{\Psi} = \begin{pmatrix} \sigma^2_{\varepsilon_X} & 0 & 0 \\ 0 & \sigma^2_{\varepsilon_M} & 0 \\ 0 & 0 & \sigma^2_{\varepsilon_Y} \end{pmatrix} \tag{4}$$

A key feature of any classic structural equation model (SEM) is that the regression parameters combine to make predictions about the multivariate normal distribution's parameters. That is, weighted combinations of the mediation model parameters in $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\Psi}$ give a predicted mean vector and covariance matrix, $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. The notation conveys that the mean vector and variance-covariance matrix are functions of the mediation model parameters in $\boldsymbol{\theta}$. Later I refer to individual elements of these matrices, as shown below.

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \begin{pmatrix} \mu_X(\theta) \\ \mu_M(\theta) \\ \mu_Y(\theta) \end{pmatrix} \qquad \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \sigma^2_X(\theta) & & \\ \sigma_{MX}(\theta) & \sigma^2_M(\theta) & \\ \sigma_{YX}(\theta) & \sigma_{YM}(\theta) & \sigma^2_Y(\theta) \end{pmatrix} \tag{5}$$

The matrix expressions below give the model-predicted parameters of the multivariate normal distribution.

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = (\mathbf{I} - \boldsymbol{\beta})^{-1}\boldsymbol{\alpha} \qquad \boldsymbol{\Sigma}(\boldsymbol{\theta}) = (\mathbf{I} - \boldsymbol{\beta})^{-1}\boldsymbol{\Psi}\left((\mathbf{I} - \boldsymbol{\beta})^{-1}\right)^T \tag{6}$$

While these expressions are computationally efficient, applying covariance algebra rules provides greater insight into their, especially for people who are not so familiar with SEMs.

Covariance Algebra Rules

Covariance algebra is a way to derive the variance and covariances predicted by an SEM. David Kenny's out-of-print book *Correlation and Causation* has an excellent chapter on covariance algebra, and a pdf copy of the book can be downloaded from the bottom of the following website: https://davidakenny.net/cm/causalm.htm. The procedure starts by specifying the covariance between two variables as $cov(X,Y)$. Next, you substitute each variable's names with its regression equation, and you simplify the expression by applying different combinations of four rules:

1. The covariance between a variable and a constant equals zero. For example, $cov(Y, \beta) = 0$.
2. If a variable is multiplied by a constant, the constant can be factored out. For example, $cov(X, \beta X) = \beta \cdot cov(X,X)$.
3. The covariance of a variable with itself is the variance of that variable. For example, $cov(X,X) = var(X)$.
4. The covariance between a variable X and a sum equals the sum of covariances involving each component of the sum. For example, $cov(X, \beta X + I_M) = cov(X, \beta X) + cov(X, I_M)$.

Model-Predicted Mean Vector and Variance Covariance Matrix

The model-predicted means in $\boldsymbol{\mu}(\boldsymbol{\theta})$ can be obtained by taking expectations of the terms on the right side of Equations 1 through 3. First, the residual terms drop from the equation because $E(\varepsilon_{Xi}) = E(\varepsilon_{Mi}) = E(\varepsilon_{Yi}) = 0$. Thus, the model-predicted mean of X is just $\mu_X(\theta) = I_X$. Substituting $\mu_X(\theta)$ for X in Equation 2 then gives $\mu_M(\theta)$, and substituting $\mu_X(\theta)$ and $\mu_M(\theta)$ into Equation 3 gives $\mu_Y(\theta)$. The following expression shows how each element in $\boldsymbol{\mu}(\boldsymbol{\theta})$ varies as a function of the mediation model parameters.

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \begin{pmatrix} \mu_X(\theta) \\ \mu_M(\theta) \\ \mu_Y(\theta) \end{pmatrix} = \begin{pmatrix} I_X \\ I_M + \alpha I_X \\ I_Y + \gamma I_X + \beta(I_M + \alpha I_X) \end{pmatrix} \tag{7}$$

Next, applying the previous covariance algebra rules gives the mediation model's predictions about the variances and covariances in $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. To begin, the model-predicted variance of X and the covariance between M and X are as follows.

$$\begin{aligned}\sigma_X^2(\theta) &= cov(X,X) = cov(I_X+\varepsilon_X, I_X+\varepsilon_X) \\ &= cov(I_X,I_X)+cov(I_X,\varepsilon_X)+cov(\varepsilon_X,I_X)+cov(\varepsilon_X,\varepsilon_X) = 0+0+0+var(\varepsilon_X)\end{aligned} \tag{8}$$

$$\begin{aligned}\sigma_{MX}(\theta) &= cov(M,X) = cov(I_M+\alpha X+\varepsilon_M, I_X+\varepsilon_X) \\ &= cov(I_M+\alpha X+\varepsilon_M, I_X)+cov(I_M+\alpha X+\varepsilon_M,\varepsilon_X) \\ &= 0+cov(I_M,\varepsilon_X)+cov(\alpha X,\varepsilon_X)+cov(\varepsilon_M,\varepsilon_X) \\ &= 0+0+\alpha\cdot cov(X,\varepsilon_X)+0 = \alpha\cdot cov(I_X+\varepsilon_X,\varepsilon_X) \\ &= \alpha\cdot\left(cov(I_X,\varepsilon_X)+cov(\varepsilon_X,\varepsilon_X)\right) = \alpha\cdot var(\varepsilon_X)\end{aligned} \tag{9}$$

These equations highlight that any term involving the intercept always produces a zero, so I drop those terms going forward. After dropping I_X, I further simplify the expressions by substituting ε_X for X. The remaining solutions are as follows.

$$\begin{aligned}\sigma_{YX}(\theta) &= cov(Y,X) = cov(\gamma X+\beta M+\varepsilon_Y,\varepsilon_X) \\ &= cov(\gamma X,\varepsilon_X)+cov(\beta M,\varepsilon_X)+cov(\varepsilon_Y,\varepsilon_X) \\ &\qquad\qquad = \gamma\cdot cov(\varepsilon_X,\varepsilon_X)+\beta\cdot cov(M,\varepsilon_X)+0 \\ &= \gamma\cdot var(\varepsilon_X)+\beta\cdot cov(\alpha X+\varepsilon_M,\varepsilon_X) \\ &= \gamma\cdot var(\varepsilon_X)+\beta\cdot\left(cov(\alpha X,\varepsilon_X)+cov(\varepsilon_M,\varepsilon_X)\right) = \gamma\cdot var(\varepsilon_X)+\alpha\beta\cdot cov(X,\varepsilon_X) \\ &= \beta\cdot cov(\alpha X+\varepsilon_M,\varepsilon_X) = \beta\cdot\left(\alpha\cdot cov(X,\varepsilon_X)+cov(\varepsilon_M,\varepsilon_X)\right) = \alpha\beta\cdot cov(X,\varepsilon_X) \\ &= \gamma\cdot var(\varepsilon_X)+\alpha\beta\cdot cov(\varepsilon_X,\varepsilon_X) = \gamma\cdot var(\varepsilon_X)+\alpha\beta\cdot var(\varepsilon_X,\varepsilon_X) \\ &= (\gamma+\alpha\beta)\cdot var(\varepsilon_X)\end{aligned} \tag{10}$$

$$\begin{aligned}\sigma_M^2(\theta) &= cov(M,M) = cov(\alpha X+\varepsilon_M,\alpha X+\varepsilon_M) \\ &= cov(\alpha\varepsilon_X,\alpha\varepsilon_X+\varepsilon_M)+cov(\varepsilon_M,\alpha\varepsilon_X+\varepsilon_M) \\ &= \alpha^2\cdot cov(\varepsilon_X,\varepsilon_X)+\alpha\cdot cov(\varepsilon_X,\varepsilon_M)+\alpha\cdot cov(\varepsilon_M,\varepsilon_X)+cov(\varepsilon_M,\varepsilon_M) \\ &= \alpha^2\cdot var(\varepsilon_X)+0+0+var(\varepsilon_M) \\ &= \alpha^2\cdot var(\varepsilon_X)+var(\varepsilon_M)\end{aligned} \tag{11}$$

$$
\begin{aligned}
\sigma_{YM}(\theta) = cov(Y, M) &= cov(\gamma X + \beta M + \varepsilon_Y, \alpha X + \varepsilon_M) \\
&= cov(\gamma \varepsilon_X, \alpha\varepsilon_X + \varepsilon_M) + cov(\beta M, \alpha\varepsilon_X + \varepsilon_M) + cov(\varepsilon_Y, \alpha\varepsilon_X + \varepsilon_M) \\
&= \ \gamma\alpha \cdot cov(\varepsilon_X, \varepsilon_X) + \gamma \cdot cov(\varepsilon_X, \varepsilon_M) + \alpha\beta \cdot cov(M, \varepsilon_X) + \ \beta \cdot cov(M, \varepsilon_M) + \alpha \\
&\qquad \cdot cov(\varepsilon_Y, \varepsilon_X) + cov(\varepsilon_Y, \varepsilon_M) \\
&= \ \gamma\alpha \cdot var(\varepsilon_X) + 0 + \alpha\beta \cdot cov(\alpha X + \varepsilon_M, \varepsilon_X) + \ \beta \cdot cov(\alpha X + \varepsilon_M, \varepsilon_M) + 0 + 0 \\
&= \ \gamma\alpha \cdot var(\varepsilon_X) + \alpha\beta \cdot \left(\alpha \cdot cov(\varepsilon_X, \varepsilon_X) + cov(\varepsilon_M, \varepsilon_X)\right) + \ \beta \\
&\qquad \cdot \left(\alpha \cdot cov(\varepsilon_X, \varepsilon_M) + cov(\varepsilon_M, \varepsilon_M)\right) \\
&= \ \gamma\alpha \cdot var(\varepsilon_X) + \alpha\beta \cdot (\alpha \cdot var(\varepsilon_X) + 0) + \ \beta \cdot \left(0 + var(\varepsilon_M)\right) \\
&= \ \gamma\alpha \cdot var(\varepsilon_X) + \alpha^2\beta \cdot var(\varepsilon_X) + \ \beta \cdot var(\varepsilon_M) \\
&= \left(\alpha^2\beta + \gamma\alpha\right) \cdot var(\varepsilon_X) + \ \beta \cdot var(\varepsilon_M)
\end{aligned} \tag{12}
$$

$$
\begin{aligned}
\sigma_Y^2(\theta) = cov(Y, Y) &= cov(\gamma X + \beta M + \varepsilon_Y, \gamma X + \beta M + \varepsilon_Y) \\
&= cov(\gamma\varepsilon_X, \gamma\varepsilon_X + \beta M + \varepsilon_Y) + cov(\beta M, \gamma\varepsilon_X + \beta M + \varepsilon_Y) \\
&\qquad + cov(\varepsilon_Y, \gamma\varepsilon_X + \beta M + \varepsilon_Y) \\
&= cov(\gamma\varepsilon_X, \gamma\varepsilon_X) + cov(\gamma\varepsilon_X, \beta M) + cov(\gamma\varepsilon_X, \varepsilon_Y) + cov(\beta M, \gamma\varepsilon_X) \\
&\qquad + cov(\beta M, \beta M) + cov(\beta M, \varepsilon_Y) + cov(\varepsilon_Y, \gamma\varepsilon_X) + cov(\varepsilon_Y, \beta M) \\
&\qquad + cov(\varepsilon_Y, \varepsilon_Y) \\
&= \gamma^2 \cdot var(\varepsilon_X) + \gamma\beta \cdot cov(\varepsilon_X, \alpha X + \varepsilon_M) + 0 + \gamma\beta \cdot cov(\alpha X + \varepsilon_M, \varepsilon_X) + \beta^2 \\
&\qquad \cdot cov(\alpha X + \varepsilon_M, \alpha X + \varepsilon_M) + \beta \cdot cov(\alpha X + \varepsilon_M, \varepsilon_Y) + 0 + \beta \\
&\qquad \cdot cov(\varepsilon_Y, \alpha X + \varepsilon_M) + var(\varepsilon_Y) \\
&= \gamma^2 \cdot var(\varepsilon_X) + \gamma\beta \cdot \left(cov(\varepsilon_X, \alpha\varepsilon_X) + cov(\varepsilon_X, \varepsilon_M)\right) + \gamma\beta \\
&\qquad \cdot \left(cov(\alpha\varepsilon_X, \varepsilon_X) + cov(\varepsilon_M, \varepsilon_X)\right) + \beta^2 \\
&\qquad \cdot \left(cov(\alpha\varepsilon_X, \alpha\varepsilon_X) + cov(\alpha\varepsilon_M, \alpha\varepsilon_X) + cov(\alpha\varepsilon_X, \varepsilon_M)\right. \\
&\qquad \left. + cov(\varepsilon_M, \varepsilon_M)\right) + \beta \cdot \left(cov(\alpha\varepsilon_X, \varepsilon_Y) + cov(\varepsilon_M, \varepsilon_Y)\right) + \beta \\
&\qquad \cdot \left(cov(\varepsilon_Y, \alpha\varepsilon_X) + cov(\varepsilon_Y, \varepsilon_M)\right) + var(\varepsilon_Y) \\
&= \gamma^2 \cdot var(\varepsilon_X) + \gamma\alpha\beta \cdot cov(\varepsilon_X, \varepsilon_X) + \gamma\alpha\beta \cdot cov(\varepsilon_X, \varepsilon_X) + \beta^2 \\
&\qquad \cdot \left(\alpha^2 \cdot cov(\varepsilon_X, \varepsilon_X) + cov(\varepsilon_M, \varepsilon_M)\right) + var(\varepsilon_Y)
\end{aligned} \tag{13}
$$

$$= \gamma^2 \cdot var(\varepsilon_X) + \gamma\alpha\beta \cdot var(\varepsilon_X) + \gamma\alpha\beta \cdot var(\varepsilon_X) + \beta^2\alpha^2 \cdot var(\varepsilon_X) + \beta^2 \cdot var(\varepsilon_M) + var(\varepsilon_Y)$$
$$= \left(\gamma^2 + 2\gamma\alpha\beta + \beta^2\alpha^2\right) \cdot var(\varepsilon_X) + \beta^2 \cdot var(\varepsilon_M) + var(\varepsilon_Y)$$

To summarize, the elements of the model-predicted mean vector and variance–covariance matrix, $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, are computed as follows.

$$\begin{aligned}
\mu_X(\theta) &= I_X \\
\mu_M(\theta) &= I_M + \alpha I_X \\
\mu_Y(\theta) &= I_Y + \gamma I_X + \beta(I_M + \alpha I_X) \\
\sigma^2_X(\theta) &= var(\varepsilon_X) \\
\sigma_{MX}(\theta) &= \alpha \cdot var(\varepsilon_X) \\
\sigma_{YX}(\theta) &= (\gamma + \alpha\beta) \cdot var(\varepsilon_X) \\
\sigma^2_M(\theta) &= \alpha^2 \cdot var(\varepsilon_X) + var(\varepsilon_M) \\
\sigma_{YM}(\theta) &= \left(\alpha^2\beta + \gamma\alpha\right) \cdot var(\varepsilon_X) + \beta \cdot var(\varepsilon_M) \\
\sigma^2_Y(\theta) &= \left(\gamma^2 + 2\gamma\alpha\beta + \beta^2\alpha^2\right) \cdot var(\varepsilon_X) + \beta^2 \cdot var(\varepsilon_M) + var(\varepsilon_Y)
\end{aligned} \tag{14}$$

Structured Model Derivatives

The previous expressions show that each mean, variance, and covariance in $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is a weighted combination of the mediation model parameters. We previously used Newton's algorithm as an optimizer for identifying the values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ that maximize the log-likelihood function (i.e., maximize the data's evidence). Estimating a path model requires an additional matrix of derivatives that summarize the linkages between the mediation model parameters and $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. This matrix Δ contains coefficients (derivatives) that capture the amount by which the model-implied moments in the rows of Δ change as a function of the mediation model parameters in the columns of Δ. To get these derivatives, you differentiate each of the functions in Equation 14 with respect to each of the mediation model parameters. For illustration, the table below displays each element of Δ, and more elegant matrix expressions for these quantities are available in the classic SEM literature (e.g., Bentler & Weeks, 1980).

	I_X	I_M	I_Y	α	γ	β	$\sigma^2_{\varepsilon_X}$	$\sigma^2_{\varepsilon_M}$	$\sigma^2_{\varepsilon_Y}$
$\mu_X(\theta)$	1	0	0	0	0	0	0	0	0
$\mu_M(\theta)$	α	1	0	I_X	0	0	0	0	0
$\mu_Y(\theta)$	$\gamma + \alpha\beta$	β	1	βI_X	I_X	$I_M + \alpha I_X$	0	0	0
$\sigma^2_X(\theta)$	0	0	0	0	0	0	1	0	0
$\sigma_{MX}(\theta)$	0	0	0	$\sigma^2_{\varepsilon_X}$	0	0	α	0	0
$\sigma_{YX}(\theta)$	0	0	0	$\beta\sigma^2_{\varepsilon_X}$	$\sigma^2_{\varepsilon_X}$	$\alpha\sigma^2_{\varepsilon_X}$	$\gamma + \alpha\beta$	0	0
$\sigma^2_M(\theta)$	0	0	0	$2\alpha\sigma^2_{\varepsilon_X}$	0	0	α^2	1	0
$\sigma_{YM}(\theta)$	0	0	0	$(\gamma + 2\alpha\beta)\sigma^2_{\varepsilon_X}$	$\alpha\sigma^2_{\varepsilon_X}$	$\alpha^2\sigma^2_{\varepsilon_X} + \sigma^2_{\varepsilon_M}$	$\alpha\gamma + \alpha^2\beta$	β	0
$\sigma^2_Y(\theta)$	0	0	0	$(\gamma + \alpha\beta)2\beta\sigma^2_{\varepsilon_X}$	$(\gamma + \alpha\beta)2\sigma^2_{\varepsilon_X}$	$2\beta\sigma^2_{\varepsilon_M} + 2\sigma^2_{\varepsilon_X}(\alpha^2\beta + \alpha\gamma)$	$\gamma^2 + 2\alpha\beta\gamma + \alpha^2\beta^2$	β^2	1

Newton's Algorithm

Newton's algorithm for multivariate normal data used the following updating step

$$\tilde{\boldsymbol{\theta}}^{(t+1)} = \tilde{\boldsymbol{\theta}}^{(t)} - \boldsymbol{H}_{\text{O}}^{-1}\left(\tilde{\boldsymbol{\theta}}^{(t)}\right)\nabla LL^{(t)} \tag{15}$$

where $\tilde{\boldsymbol{\theta}}$ is a vector of parameter values, t indexes the iterations, $\boldsymbol{H}_{\text{O}}^{-1}\left(\tilde{\boldsymbol{\theta}}^{(t)}\right)$ is the inverse of the Hessian (the matrix of second derivatives) evaluated at the current parameter estimates at iteration t, and $\nabla LL^{(t)}$ is a vector of first derivatives (the gradient vector) evaluated at the current estimates. In that context, $\tilde{\boldsymbol{\theta}}$ was a q-element vector containing all unique elements in $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, the Hessian was a q by q symmetric matrix, and the gradient vector was also comprised of q elements. A previous handout gave the analytic expressions for the derivatives and gradient vector.

Newton's algorithm for the mediation model introduces the derivative matrix Δ from the previous section as follows

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \left(\Delta^T \boldsymbol{H}_{\text{O}}\left(\boldsymbol{\theta}^{(t)}\right)\Delta\right)^{-1}\left(\Delta^T \nabla LL^{(t)}\right) \tag{16}$$

where $\left(\Delta^T \boldsymbol{H}_{\text{O}}\left(\boldsymbol{\theta}^{(t)}\right)\Delta\right)$ is the Hessian (second derivative matrix) of the mediation model, and $\Delta^T \nabla LL^{(t)}$ its gradient vector. Conceptually, the presence of Δ in the equation translates the derivative information from the metric of $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ to the metric of the mediation model parameters. Note that this expression is asymptotic and assumes that the model is correct (the single-mediation perfectly predicts the means and variance-covariance matrix, so that is true here). Savalei and Rosseel (in press) give more precise expressions for the structured model's Hessian and gradient vector that do not require these assumptions. Finally, adding a step size parameter λ introduces flexibility for adaptively tailoring the jump sizes.

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \lambda\left(\Delta^T \boldsymbol{H}_{\mathrm{O}}\left(\boldsymbol{\theta}^{(t)}\right)\Delta\right)^{-1}\left(\Delta^T \nabla LL^{(t)}\right) \tag{17}$$

The accompanying R program adaptively modifies λ during the initial phases of optimization and sets $\lambda = 1$ at later iterations.

References

Bentler, P. M., & Weeks, D. G. (1980). Linear Structural Equations with Latent-Variables. *Psychometrika, 45*(3), 289–308. doi:Doi 10.1007/Bf02293905

Savalei, V., & Rosseel, Y. (in press). Computational options for standard errors and test statistics with incomplete normal and nonnormal data. *Structural Equation Modeling: A Multidisciplinary Journal*.