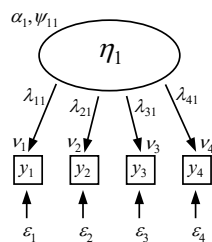# Mean Structures

Means are not required and they are not of particular interest for many structural models, but there are a few types of models, such as latent growth curve models, that entail hypotheses specifically about means. Incorporating means, or "mean structure", in a model has no impact on the model's fit, but scaling constraints are needed to ensure that latent factor means are identified. Means for latent or measured exogenous variables are called "means", and means for latent or measured endogenous variables are called "intercepts". The intercept is always the conditional mean, with its value interpreted when all predictors are equal to 0.
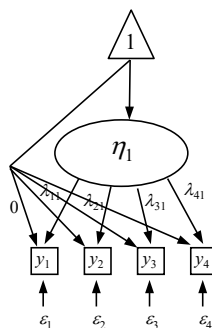
Notation
Inclusion of mean structures in SEM expands the measurement model to include a measurement intercept (Sörbom, 1974). Here is the measurement equation for one of the indicators.

$$y_1 = \nu_1 + \lambda_{11}\eta_1 + \varepsilon_1$$

This equation is the same as the measurement model equation we have seen before, except that there is a measurement intercept, ν "nu", included. Notice that the equation follows a regression equation, which now has an intercept. The intercept represents the predicted value of y_1 when the latent variable η_1 is equal to 0. I like to represent the means with symbols α or ν next to the ellipse or rectangle, as in the following diagram:

Most authors, however, use a different style of diagram [the reticular action model (RAM) notation: McArdle & McDonald, 1984], something like what is shown below:

Here, the triangle represents the mean structure for both the latent variable and the measured variables. The 1 inside the triangle is sometimes given to represent a vector of 1s which is used for computation of means in the matrix equations. Means are not usually estimated by default in most SEM packages, but they will be added if missing data or mean-adjusted robust estimates are requested.

Identification
Degrees of freedom for the model must now take into account the means estimated, so the formula is altered slightly.

$$df = \left[\frac{v(v+1)}{2} + v \right] - p,$$

Here, v ("vee" not "nu" here) is the number of measured variables, and p is number of parameters estimated in the model. For a single latent variable with four indicators, like the figure above, the number of variance-covariance elements in the obtained matrix is [4(4 + 1)]/2 = 10 and the number of observed means is 4, which equals 14 parameters available to be estimated. We count the parameters the same way except we must now add 4 means, either four measurement intercepts or three measurement intercepts and one factor mean, depending on how we have identified the factor mean.

The factor means must be identified, and this is usually done in one of two ways: setting a referent intercept to 0 or setting the factor mean to 0 (not both!). Mplus and lavaan set the intercept for the first indicator to 0 by default.

An alternative identification strategy was identified by Little, Slegers, and Card (2006) uses a complex equality constraint to identify the factor mean. Each loading is constrained to be a function of the other loadings and each intercept is constrained to be a function of the other intercepts. The following formulas illustrate how this would be done for a factor with three indicators. The constraints need to be made as part of the model statements, and not all SEM programs have this capability (in the MODEL CONSTRAINT: section in Mplus or using := statements in lavaan).

$$\lambda_{11} = 3 - \lambda_{21} - \lambda_{31}$$

$$v_1 = 0 - v_2 - v_3$$

Measurement Error
Measurement error affects estimates of means differently from the way it affects variances. Since measurement error is assumed to be random, it has a 0 mean on average. If this is the case, then following the classical test theory formula, the mean of the true score and the observed score will be equal on average

$$\mathrm{E}(X) = \mathrm{E}(T) + \mathrm{E}(e)$$
$$\mathrm{E}(X) = \mathrm{E}(T) + 0$$
$$\mathrm{E}(X) = \mathrm{E}(T)$$

$\mathrm{E}(.)$ indicates the expected value (or average), X is the observed score, T is the true score, and e is the error. We usually assume the average for measurement error is 0, so this term drops out, and we see that the mean of the observed score is the same as the true score. In classical test formula terms, measurement error has no impact on the true score, because its expected value $\mathrm{E}(e)$ is zero.

Interpretation
Means and intercepts in structural modeling can be a little difficult to wrap your head around at first. The only good way I know to explain their interpretation is to consider them in terms of measurement model equations. The expected value for some observed indicator y_j can be expressed in terms of the measurement intercept, the loading, and the factor mean, α_k.

$$E\left(y_j\right) = v_j + \lambda_{jk}\alpha_k$$

The intercept then is an adjusted mean for y_j.

$$v_j = E(y_j) - \lambda_{jk}\alpha_k$$

Notice that if the factor mean is set to 0, then the observed mean of y_j and the measurement intercept are equal. Conversely, factor mean is a function of the observed means and the intercept and if the first intercept is set to 0 and the first loading is set to 1, then the latent mean, α_k, is equal to the observed mean of the first indicator y_1.

$$\alpha_k = \frac{E(y_1) - v_1}{\lambda_{11}}$$

What all of this says is that the "latent means" are simply a function of the observed means of the indicators. When referent identification (setting the intercept to 0 and the loading to 1) is used, the latent mean takes on the value of the first indicator mean.

When effects coding identification is used, the factor means are a weighted average of the indicators, and this can be a desirable interpretation for latent means.

Estimation
To estimate means in the model, the observed matrix to be analyzed must include mean information (i.e., fitting the moment matrix), either by including a vector of means when the variance-covariance matrix is input or by supplying the program with raw data from which the means can be obtained. A special adaptation of maximum likelihood estimation (mean-weighted maximum likelihood) is needed to fit models with mean structures (refer to Bentler & Yuan, 2000, and Hayduk, 1987, for an explanation of this rationale).

Mean Comparisons and Connection to ANOVA
Multigroup or longitudinal equality constraints on measured or factor means can be used to obtain tests that are equivalent to between-subjects (e.g., Green & Thompson, 2012; Thompson & Green, 2006) and/or within-subjects (e.g., Newsom, 2015, Chapter 3; Voelkle, 2007) analysis of variance (ANOVA). The approach allows for greater flexibility for tests and relaxation of the usual ANOVA assumptions. In addition, the same approaches can be used in conjunction with binary and ordinal variables to obtain equivalent tests to chi-square and loglinear models (Newsom, 2017). Latent variables allow for estimation of measurement error, which increases the precision of these tests (Hancock, 2003), and full information maximum likelihood estimation allows for estimation in the presence of missing data, both features that offer important advantages to an SEM approach to mean comparisons.

References and Suggested Readings
Bentler, P. M., & Yuan, K.-H. (2000). On adding a mean structure to a covariance structure model. *Educational and Psychological Measurement,60*, 326-339.
Green, S. B., & Thompson, M. S. (2012). A flexible structural equation modeling approach for analyzing means. In R. H. Hoyle (Ed.), Handbook of structural equation modeling (pp. 393–416). New York: Guilford.
Hancock, G. R. (2003). Fortune cookies, measurement error, and experimental design. Journal of Modern Applied Statistical Methods, 2, 293–305.
Hayduk, L.A. (1987). *Structural equations modeling with lisrel.* Baltimore, MD: Johns Hopkins University Press
Kline, R.B. (2010). *Principles and practice of structural equation modeling (Third Edition).* New York: Guilford Press.
Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling, 13*, 59-72.
McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the reticular action model for moment structures. *British Journal of Mathematical and Statistical Psychology, 37*, 234-251.
Newsom, J.T. (2017). Structural models for binary repeated measures: Linking modern longitudinal structural equation models to conventional categorical data analysis for matched pairs. *Structural Equation Modeling: A Multidisciplinary Journal, 24*, 626-635.
Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology, 27*, 229-239.
Thompson, M. S., & Green, S. B. (2006). Evaluating between-group differences in latent variable means. In G. R. Hancock & R. O. Mueller (Eds.), Structural equation modeling: A second course (pp. 119–169). Greenwich, CT: Information Age.
Voelkle, M. C. (2007). Latent growth curve modeling as an integrative approach to the analysis of change. *Psychology Science, 49*, 375-414.