Chi-Square Statistics with Multiple Imputation

November 27, 2008

1 Likelihood Ratio Test

In this section we describe how Mplus computes the likelihood ratio test (LRT) with multiple imputations. The LRT is computed only for the ML estimator for single level SEM models using the method described in Meng and Rubin (1992), see also Enders (2008). For these models the LRT is computed for the estimated model against the unrestricted mean and variance/covariance model, i.e., the usual test of fit. For all other estimators and models the fit statistic is not computed but the distribution of the test statistic over the different imputed data sets is reported.

The LRT test statistics is computed as follows. Suppose that there are M imputed data sets. Let T_m be the test of fit statistic for the $m-$th imputed data set. Let the parameter estimates of the H_0 and H_1 models, using the $m-$th imputed data set, be Q_{0m} and Q_{1m}. Let the number of parameters for the H_0 and H_1 models respectively be p_0 and p_1. Define the average quantities as

$$\bar{T} = \frac{1}{M} \sum_{m=1}^{M} T_m$$

$$\bar{Q_0} = \frac{1}{M} \sum_{m=1}^{M} Q_{0m}$$

$$\bar{Q_1} = \frac{1}{M} \sum_{m=1}^{M} Q_{1m}$$

Now compute the LRT test statistic for the H_0 model against the H_1 model where the parameter estimates are fixed to $\bar{Q_0}$ and $\bar{Q_1}$ respectively, using the $m-$th imputed data set and denote this test statistic value by T'_m. This statistic is averaged over all imputed data sets

$$\bar{T}' = \frac{1}{M} \sum_{m=1}^{M} T'_m.$$

Then final test statistics is

$$\frac{\bar{T}'}{1 + r_3}$$

which has approximately a chi-square distribution with the same degrees of freedom as the usual test of fit statistics, i.e., $p_1 - p_0$. The correction factor r_3 is computed as follows

$$r_3 = \frac{M+1}{(M-1)(p_1 - p_0)}(\bar{T} - \bar{T}').$$

The above approximation may be quite poor if the amount of missing data is relatively large or the number of imputations M is low.

2 Wald Test

For imputed data the Wald test is computed for all estimators and models. The computation is based on estimating the joint distribution of the parameter estimates. Suppose that we need to test $F(Q) = 0$ where F is a multivariate function and Q are the model parameters. Let the parameter estimates for the $m-$th imputed data set be Q_m and their estimated asymptotic distribution be V_m. The combined estimates are computed as

$$\bar{Q} = \frac{1}{M} \sum_{m=1}^{M} Q_m$$

and their asymptotic distribution is computed as follows

$$V = \frac{1}{M} \sum_{m=1}^{M} V_m + \frac{M+1}{M(M-1)} \sum_{m=1}^{M} (Q_m - \bar{Q})(Q_m - \bar{Q})^T.$$

The correct Wald test can now be computed the usual way

$$W = F(\bar{Q})(F'(\bar{Q})V(F'(\bar{Q}))^T)^{-1}F(\bar{Q})^T$$

where F' is the first derivative of F with respect to the parameters Q. Under the null hypothesis, $F(Q) = 0$, the distribution of W is a chi-square with d degrees of freedom, where d is the dimension of the restrictions F. In the above formula the delta method is used to obtain the asymptotic covariance of $F(Q)$ from the asymptotic covariance of Q. This method however does not utilize the actual chi-square values across the different imputations as in the previous section. Instead, it computes the Wald test simply by using the estimated asymptotic variance of the parameters, just as the Wald test is computed for complete data analysis.

3 References

Enders, C.K. (2008, in progress). Applied missing data analysis. New York: Guilford.

Meng, X.L. & Rubin, D.B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. Biometrika, 79, 103-111.