Scientific Software International

Multilevel Structural equations modeling

Table of Contents

1   Multilevel Structural Equations Models

Social science research often entails the analysis of data with a hierarchical structure. A frequently cited example of multilevel data is a dataset containing measurements on children nested within schools, with schools nested within education departments.

The need for statistical models that take account of the sampling scheme is well recognized and it has been shown that the analysis of survey data under the assumption of a simple random sampling scheme may give rise to misleading results.

Iterative numerical procedures for the estimation of variance and covariance components for unbalanced designs were developed in the 1980s and were implemented in software packages such as MLWIN, SAS PROC MIXED and

HLM. At the same time, interest in latent variables, that is, variables that cannot be directly observed or can only imperfectly be observed, led to the theory providing for the definition, fitting and testing of general models for linear structural relations for data from simple random samples.

A more general model for multilevel structural relations, accommodating latent variables and the possibility of missing data at any level of the hierarchy and providing the combination of developments in these two fields, was a logical next step. In papers by Goldstein and MacDonald (1988), MacDonald and Goldstein (1989) and McDonald (1993), such a model was proposed. Muthén (1990, 1991) proposed a partial maximum likelihood solution as simplification in the case of an unbalanced design. An overview of the latter can be found in Hox (1993).

General two-level structural equations modeling is available in LISREL. Full information maximum likelihood estimation is used, and a test for goodness of fit is given. An example, illustrating the implementation of the results for unbalanced designs with missing data at both levels of the hierarchy, is also given.

2 A General Two-level Structural Equations Model

Consider a data set consisting of 3 measurements, *math 1*, *math 2*, and *math 3*, made on each of 1000 children who are nested within $N = 100$ schools. This data set can be schematically represented for school i as follows
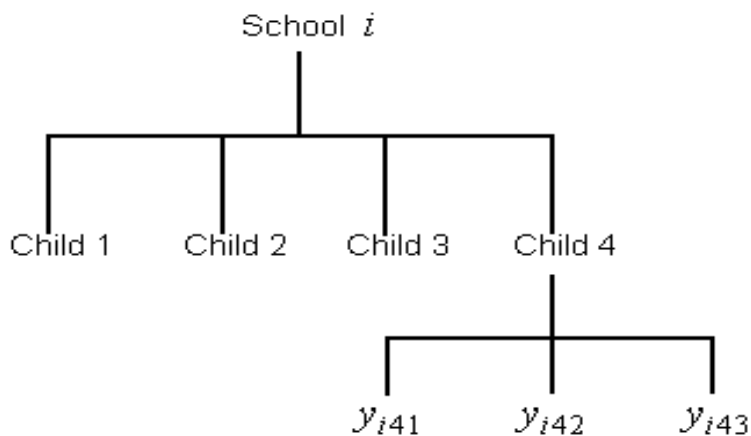
For the i-th level-2 unit (school), we can write

$$\mathbf{y}_i = \begin{pmatrix} \mathbf{y}_{i1} \\ \mathbf{y}_{i2} \\ \mathbf{y}_{i3} \\ \mathbf{y}_{i4} \end{pmatrix},$$

where for child 4 within school i

$$\mathbf{y}_{i4}^{'} = \begin{pmatrix} y_{i41} & y_{i42} & y_{i43} \\ math\,1 & math\,2 & math\,3 \end{pmatrix}.$$

A model which allows for between- and within-schools variation in math scores is the following simple variance component model

$$\mathbf{y}_{i1} = \mathbf{v}_i + \mathbf{u}_{i1}$$
$$\mathbf{y}_{i2} = \mathbf{v}_i + \mathbf{u}_{i2}$$
$$\mathbf{y}_{i3} = \mathbf{v}_i + \mathbf{u}_{i3}$$
$$\mathbf{y}_{i4} = \mathbf{v}_i + \mathbf{u}_{i4},$$

or $\mathbf{y}_{ij} = \mathbf{v}_i + \mathbf{u}_{ij}$, $i = 1, 2, \ldots, N$, where it is assumed that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N$ are i.i.d. $N(0, \mathbf{\Sigma}_B)$ and that $\mathbf{u}_{i1}, \ldots, \mathbf{u}_{iN}$ are i.i.d $N(0, \mathbf{\Sigma}_W)$. It is additionally assumed that

$$Cov(\mathbf{v}_i, \mathbf{u}_{ij}) = 0, i = 1, \ldots, N; j = 1, 2, \ldots, n_i.$$

From the distributional assumptions it follows that

$$Cov(\mathbf{y}_i, \mathbf{y}_i^{'}) = \begin{pmatrix} \mathbf{\Sigma}_B + \mathbf{\Sigma}_W & \mathbf{\Sigma}_B & \mathbf{\Sigma}_B & \mathbf{\Sigma}_B \\ \mathbf{\Sigma}_B & \mathbf{\Sigma}_B + \mathbf{\Sigma}_W & \mathbf{\Sigma}_B & \mathbf{\Sigma}_B \\ \mathbf{\Sigma}_B & \mathbf{\Sigma}_B & \mathbf{\Sigma}_B + \mathbf{\Sigma}_W & \mathbf{\Sigma}_B \\ \mathbf{\Sigma}_B & \mathbf{\Sigma}_B & \mathbf{\Sigma}_B & \mathbf{\Sigma}_B + \mathbf{\Sigma}_W \end{pmatrix}$$

or $Cov(\mathbf{y}_i, \mathbf{y}_i^{'}) = \mathbf{1}\mathbf{1}^{'} \otimes \mathbf{\Sigma}_B + \mathbf{I} \otimes \mathbf{\Sigma}_W$.

It also follows that

$$E(\mathbf{y}_i) = \mathbf{0}.$$

In practice, the latter assumption $E(\mathbf{y}_i) = \mathbf{0}$ is seldom realistic, since measurements such as math scores do not have zero means. One approach to this problem is to use grand mean centering. Alternatively, one can add a fixed component to the model $\mathbf{y}_{ij} = \mathbf{v}_i + \mathbf{u}_{ij}$, so that

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{v}_i + \mathbf{u}_{ij}, \tag{1}$$

where $\mathbf{X}_{ij}$ denotes a design matrix and $\boldsymbol{\beta}$ a vector of regression coefficients.

Suppose that for the example above, the only measurements available for child 1 are *math 1* and *math 3* and for child 2 *math 2* and *math 3*.

Let $\mathbf{S}_{i1}$ and $\mathbf{S}_{i2}$ be selection matrices defined as follows

$$\mathbf{S}_{i1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{ therefore } \mathbf{S}_{i1}\mathbf{v}_i = \begin{bmatrix} v_{i1} \\ v_{i3} \end{bmatrix}$$

and

$$\mathbf{S}_{i2} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{ therefore } \mathbf{S}_{i2}\mathbf{v}_i = \begin{bmatrix} v_{i2} \\ v_{i3} \end{bmatrix}.$$

In general, if p measurements were made, $\mathbf{S}_{ij}$ (see, for example, du Toit, 1995) consists of a subset of the rows of the $p \times p$ identity matrix $\mathbf{I}_p$, where the rows of $\mathbf{S}_{ij}$ correspond to the response measurements available for the (i, j)-th unit.

The above model can be generalized to accommodate incomplete data by the inclusion of these selection matrices. Hence

$$\mathbf{y}_{ij} = \mathbf{X}_{(y)ij}\boldsymbol{\beta} + \mathbf{S}_{ij}\mathbf{v}_i + \mathbf{S}_{ij}\mathbf{u}_{ij} \tag{2}$$

where $\mathbf{X}_{(y)}$ is a design matrix of the appropriate dimensions.

If we further suppose that we have a $q \times 1$ vector of variables $\mathbf{x}_i$ characterizing the level-2 units (schools), then we can write the observed data for the i-th level-2 unit as

$$\mathbf{y}_i' = [\mathbf{y}_{i1}', \mathbf{y}_{i2}', ..., \mathbf{y}_{in_i}', \mathbf{x}_i'],$$

where

$$\mathbf{y}_{ij}' = [y_{ij1}, y_{ij2}, ..., y_{ijp}]$$

and

$$\mathbf{x}_i' = [x_{i1}, x_{i2}, ..., x_{iq}]. \tag{3}$$

We assume that $\mathbf{y}_{ij}$ and $\mathbf{x}_i$ can be written as

$$\mathbf{y}_{ij} = \mathbf{X}_{(y)ij}\boldsymbol{\beta}_y + \mathbf{S}_{ij}\mathbf{v}_i + \mathbf{S}_{ij}\mathbf{u}_{ij}, \ j = 1, 2, ... n_i \tag{4}$$

$$\mathbf{x}_i = \mathbf{X}_{(x)i}\boldsymbol{\beta}_x + \mathbf{R}_i\mathbf{w}_i, \ i = 1, 2, ... N \tag{5}$$

where $\mathbf{X}_{(y)}$ and $\mathbf{X}_{(x)}$ are design matrices for fixed effects, and $\mathbf{S}_{ij}$ and $\mathbf{R}_i$ are selection matrices for random effects of order $p_{ij} \times p$ and $q_i \times q$ respectively. Note that (4) defines two types of random effects, where $\mathbf{v}_i$ is common to level-3 units and $\mathbf{u}_{ij}$ is common to level-1 units nested within a specific level-2 unit.

Additional distributional assumptions are

$$Cov(\mathbf{w}_i) = \boldsymbol{\Sigma}_{xx}, \quad i = 1, 2, ..., N$$
$$Cov(\mathbf{y}_{ij}, \mathbf{w}_i) = \boldsymbol{\Sigma}_{xy}, \quad i = 1, 2, ..., N; \ j = 1, 2, ..., n_i \tag{6}$$
$$Cov(\mathbf{u}_{ij}, \mathbf{w}_i) = \mathbf{0}.$$

From (4) and (5), it follows that

$$\mathbf{y}_i = \begin{bmatrix} \mathbf{X}_{(y)i}\boldsymbol{\beta}_y + \mathbf{S}_i\mathbf{v}_i + \sum_{j=1}^{n_i} \mathbf{Z}_{ij}\mathbf{u}_{ij} \\ \mathbf{X}_{(x)i}\boldsymbol{\beta}_x + \mathbf{R}_i\mathbf{r}_i \end{bmatrix} \tag{7}$$

where

$$\mathbf{X}_{(y)i} = \begin{bmatrix} \mathbf{X}_{(y)i1} \\ \vdots \\ \mathbf{X}_{(y)in_i} \end{bmatrix}, \mathbf{S}_i = \begin{bmatrix} \mathbf{S}_{i1} \\ \vdots \\ \mathbf{S}_{in_i} \end{bmatrix}, \mathbf{R}_i = \begin{bmatrix} \mathbf{R}_{i1} \\ \vdots \\ \mathbf{R}_{in_i} \end{bmatrix},$$

and $\quad \mathbf{Z}_{ij} = \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{S}_{ij} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}.$

From the distributional assumptions given above, it follows that

$$\mathbf{y}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i),$$

where

$$\boldsymbol{\mu}_i = \begin{bmatrix} \mathbf{X}_{(y)i} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{(x)i} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_y \\ \boldsymbol{\beta}_x \end{bmatrix} = \mathbf{X}_i \boldsymbol{\beta}, \tag{8}$$

and

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} \mathbf{V}_i & \mathbf{S}_i \boldsymbol{\Sigma}_{yx} \mathbf{R}_i' \\ \mathbf{R}_i \boldsymbol{\Sigma}_{xy} \mathbf{S}_i' & \mathbf{R}_i \boldsymbol{\Sigma}_{xx} \mathbf{R}_i' \end{bmatrix} \tag{9}$$

where

$$\mathbf{V}_i = Cov \begin{pmatrix} \mathbf{y}_{i1} \\ \vdots \\ \mathbf{y}_{in_i} \end{pmatrix} = \mathbf{S}_i \boldsymbol{\Sigma}_B \mathbf{S}_i' + \sum_{j=1}^{n_i} \mathbf{Z}_{ij} \boldsymbol{\Sigma}_W \mathbf{Z}_{ij}'.$$

Remark

6

If $\mathbf{R}_i = \mathbf{I}_q$ and $\mathbf{S}_{ij} = \mathbf{I}_p$, corresponding to the case of no missing y or x variables, then $\mathbf{S}_i\mathbf{\Sigma}_{yx}\mathbf{R}_i' = \mathbf{1}\otimes\mathbf{\Sigma}_{yx}$ where $\mathbf{1}'$ is a $n_i \times 1$ row vector $(1,1,\ldots,1)$.

Furthermore, for $\mathbf{S}_{ij} = \mathbf{I}_p$, $j = 1, 2, \ldots, n_i$

$$\mathbf{V}_i = \mathbf{I}_{n_i} \otimes \mathbf{\Sigma}_W + \mathbf{1}\mathbf{1}' \otimes \mathbf{\Sigma}_B$$

(see, for example, MacDonald and Goldstein, 1989). The unknown parameters in (8) and (9) are $\boldsymbol{\beta}$, $vecs\mathbf{\Sigma}_B$, $vecs\mathbf{\Sigma}_W$, $vecs\mathbf{\Sigma}_{xy}$ and $vecs\mathbf{\Sigma}_{xx}$.

Structural models for the type of data described above may be defined by restricting the elements of $\boldsymbol{\beta}$, $\mathbf{\Sigma}_B$, $\mathbf{\Sigma}_W$, $\mathbf{\Sigma}_{xy}$, and $\mathbf{\Sigma}_{xx}$ to be some basic set of parameters $\boldsymbol{\gamma}' = (\gamma_1, \gamma_2, \ldots, \gamma_k)$.

For example, assume the following pattern for the matrices $\mathbf{\Sigma}_W$ and $\mathbf{\Sigma}_B$, where $\mathbf{\Sigma}_W$ refers to the within (level-1) covariance matrix and $\mathbf{\Sigma}_B$ to the between (level-2) covariance matrix:

$$\begin{aligned}\mathbf{\Sigma}_W &= \mathbf{\Lambda}_W \mathbf{\Psi}_W \mathbf{\Lambda}_W' + \mathbf{D}_W \\ \mathbf{\Sigma}_B &= \mathbf{\Lambda}_B \mathbf{\Psi}_B \mathbf{\Lambda}_B' + \mathbf{D}_B.\end{aligned} \qquad (10)$$

Factor analysis models typically have the covariance structures defined by (10).

Consider a confirmatory factor analysis model with 2 factors and assume $p = 6$.

$$\mathbf{\Lambda}_W = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix}, \quad \mathbf{\Psi}_W = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix},$$

and

$$D_W = \begin{bmatrix} \theta_{11} & & \\ & \ddots & \\ & & \theta_{11} \end{bmatrix}.$$

If we restrict all the parameters across the level-1 and level-2 units to be equal, then

$$\gamma^{'} = [\lambda_{11}, \lambda_{21}, \ldots, \lambda_{44}, \psi_{11}, \psi_{22}, \theta_{11}, \ldots, \theta_{66}]$$

is the vector of unknown parameters.

3 Maximum Likelihood for General Means and Covariance Structures

In this section, we give a general framework for normal maximum likelihood estimation of the unknown parameters. In practice, the number of variables $(p + q)$ and the number of level-1 units within a specific level-2 unit may be quite large, which leads to Σ_i matrices of very high order. It is therefore apparent that further simplification of the likelihood function derivatives and Hessian is required if the goal is to implement the theoretical results in a computer program. These aspects are addressed in du Toit and du Toit (2008).

Denote the expected value and covariance matrix of $\mathbf{y}_i$ by $\boldsymbol{\mu}_i$ and Σ_i respectively (see (8) and (9)). The log-likelihood function of $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$ may then be expressed as

$$\ln L = -\frac{1}{2} \sum_{i=1}^{N} \{n_i \ln 2\pi + \ln |\Sigma| + tr\Sigma_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i)(\mathbf{y}_i - \boldsymbol{\mu}_i)^{'}\} \qquad (11)$$

Instead of maximizing $\ln L$, maximum likelihood estimates of the unknown parameters are obtained by minimizing $-\ln L$ with the constant term omitted, *i.e.*, by minimizing the following function

$$F(\gamma) = \frac{1}{2} \sum_{i=1}^{N} \{\ln |\Sigma_i| + tr\Sigma_i^{-1}\mathbf{G}_y\}, \qquad (12)$$

where

$$\mathbf{G}_{y_i} = (\mathbf{y}_i - \boldsymbol{\mu}_i)(\mathbf{y}_i - \boldsymbol{\mu}_i)^{'}. \qquad (13)$$

8

Its minimum $\dfrac{\partial F(\gamma)}{\partial \gamma} = \mathbf{0}$ yields the normal maximum likelihood estimator $\overset{\wedge}{\gamma}$ of the unknown vector of parameters γ.

Unless the model yields maximum likelihood estimators in closed form, it will be necessary to make use of an iterative procedure to minimize the discrepancy function. The optimization procedure (Browne and du Toit, 1992) is based on the so-called Fisher scoring algorithm, which in the case of structured means and covariances may be regarded as a sequence of Gauss-Newton steps with quantities to be fitted as well as the weight matrix changing at each step. Fisher scoring algorithms require the gradient vector and an approximation to the Hessian matrix.

4 Fit Statistics and Hypothesis Testing

The multilevel structural equations model, $M(\gamma)$, and its assumptions imply a covariance structure $\Sigma_B(\gamma)$, $\Sigma_W(\gamma)$, $\Sigma_{xy}(\gamma)$, $\Sigma_{xx}(\gamma)$ and mean structure $\mu(\gamma)$ for the observable random variables where γ is a $k \times 1$ vector of parameters in the statistical model. It is assumed that the empirical data are a random sample of N level-2 units and $\sum_{i=1}^{N} n_i$ level-1 units, where n_i denotes the number of level-1 units within the i-th level-2 unit. From this data, we can compute estimates of μ, Σ_B, ..., Σ_{xx} if no restrictions are imposed on their elements. The number of parameters for the unrestricted model is

$$k^* = m + 2\left[\frac{1}{2}p(p+1)\right] + pq + \frac{1}{2}q(q+1)$$

and is summarized in the $k^* \times 1$ vector π. The unrestricted model $M(\pi)$ can be regarded as the "baseline" model.

To test the model $M(\gamma)$, we use the likelihood ratio test statistic

$$c = -2\ln L(\overset{\wedge}{\gamma}) - 2\ln L(\overset{\wedge}{\pi}) \qquad (14)$$

If the model $M(\gamma)$ holds, c has a χ^2-distribution with $d = k^* - k$ degrees of freedom. If the model does not hold, c has a non-central χ^2-distribution with d degrees of freedom and non-centrality parameter λ that may be estimated as (see Browne and Cudeck, 1993):

$$\overset{\wedge}{\lambda} = \max\{(c-d), 0\} \qquad (15)$$

These authors also show how to set up a confidence interval for λ.

It is possible that the researcher has specified a number of competing models $M_1(\gamma_1), M_2(\gamma_2), \ldots, M_k(\gamma_k)$. If the models are nested in the sense that $\gamma_j : k_j \times 1$ is a subset of $\gamma_i : k_i \times 1$, then one may use the likelihood ratio test with degrees of freedom $k_i - k_j$ to test $M(\gamma_j)$ against $M(\gamma_i)$.

Another approach is to compare models on the basis of some criteria that take parsimony as well as fit into account. This approach can be used regardless of whether or not the models can be ordered in a nested sequence. Two strongly related criteria are the AIC measure of Akaike (1987) and the CAIC of Bozdogan (1987).

$$AIC = c + 2d \tag{16}$$

$$CAIC = c + (1 + \ln \sum_{i=1}^{N} n_i)d \tag{17}$$

The use of c as a central χ^2-statistic is based on the assumption that the model holds exactly in the population. A consequence of this assumption is that models that hold approximately in the population will be rejected in large samples.

Steiger (1990) proposed the root mean square error of approximation (RMSEA) statistic that takes particular account of the error of approximation in the population

$$RMSEA = \sqrt{\frac{\hat{F}_0}{d}}, \tag{18}$$

where $\hat{F}_0$ is a function of the sample size, degrees of freedom and the fit function. To use the RMSEA as a fit measure in multilevel SEM, we propose

$$\hat{F}_0 = \max\left\{ \left[\frac{c - d}{N} \right], 0 \right\} \tag{19}$$

Browne and Cudeck (1993) suggest that an RMSEA value of 0.05 indicates a close fit and that values of up to 0.08 represent reasonable errors of approximation in the population.

5 Starting Values and Convergence Issues

In fitting a structural equations model to a hierarchical data set, one may encounter convergence problems unless good starting values are provided. A procedure that appears to work well in practice is to start the estimation procedure by fitting the unrestricted model to the data. The first step is therefore to obtain estimates of the fixed components $(\boldsymbol{\beta})$ and the variance components $(\boldsymbol{\Sigma}_B, \boldsymbol{\Sigma}_{xy}, \boldsymbol{\Sigma}_{xx}$ and $\boldsymbol{\Sigma}_W)$. Our experience with the Gauss-Newton algorithm (see, for example, Browne and du Toit, 1992) is that convergence is usually obtained within less than 15 iterations, using initial estimates $\boldsymbol{\beta} = \boldsymbol{0}, \boldsymbol{\Sigma}_B = \mathbf{I}_p, \boldsymbol{\Sigma}_{xy} = \boldsymbol{0}, \boldsymbol{\Sigma}_{xx} = \mathbf{I}_q$ and $\boldsymbol{\Sigma}_W = \mathbf{I}_p$. At convergence, the value of $-2\ln L$ is computed.

Next, we treat

$$\mathbf{S}_B = \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_B & \hat{\boldsymbol{\Sigma}}_{yx} \\ \hat{\boldsymbol{\Sigma}}_{xy} & \hat{\boldsymbol{\Sigma}}_{xx} \end{bmatrix} \text{ and } \mathbf{S}_W = \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_W & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$

as sample covariance matrices and fit a two group structural equations model to the between- and within-groups. Parameter estimates obtained in this manner are used as the elements of the initial parameter vector $\boldsymbol{\gamma}_0$.

In the third step, the iterative procedure is restarted and $\boldsymbol{\gamma}_k$ updated from $\boldsymbol{\gamma}_{k-1}$, $k = 1, 2, \ldots$ until convergence is reached.

The following example illustrates the steps outlined above. The data set used in this section forms part of the data library of the Multilevel Project at the University of London and comes from the Junior School Project (Mortimore *et al*, 1988). Mathematics and language tests were administered in three consecutive years to more than 1000 students from 49 primary schools, which were randomly selected from primary schools maintained by the Inner London Education Authority.

The following variables were selected from the data file:

- *School*   School code (1 to 49)
- *Math1*   Score on mathematics test in year 1 (score 1 - 40)
- *Math2*   Score on mathematics test in year 2 (score 1 - 40)
- *Math3*   Score on mathematics test in year 3 (score 1 - 40)

The school number (*School*) is used as the level-2 identification.
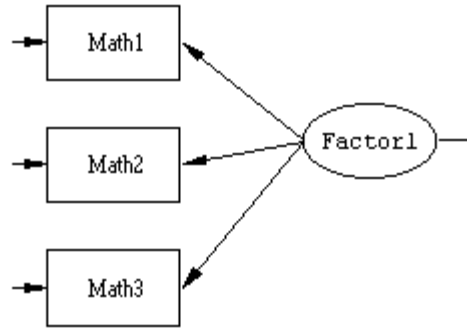
Figure 1: Confirmatory factor analysis model

A simple confirmatory factor analysis model (see Figure 1) is fitted to the data:

$$\mathbf{\Sigma}_B = \mathbf{\lambda}\mathbf{\Psi}\mathbf{\lambda}' + \mathbf{D}_B,$$
$$\mathbf{\Sigma}_W = \mathbf{\lambda}\mathbf{\Psi}\mathbf{\lambda}' + \mathbf{D}_W,$$

where

$$\mathbf{\lambda}' = (1, \lambda_{21}, \lambda_{31})$$

and $\mathbf{D}_B$ and $\mathbf{D}_W$ are diagonal matrices with diagonal elements equal to the unique (error) variances of *Math1*, *Math2* and *Math3*. The variance of the factor is denoted by Ψ. Note that we assume equal factor loadings and factor variances across the between- and within-groups, leading to a model with 3 degrees of freedom. The SIMPLIS (see Jöreskog and Sörbom, 1993) syntax file to fit the factor analysis model is shown below. Note that the between- and within-groups covariance matrices are the estimated $\mathbf{\Sigma}_B$ and $\mathbf{\Sigma}_W$ obtained in the first step by fitting the unrestricted model.

```
Group 1: Between Schools JSP data (Level 2)
Observed Variables: Math1 Math2 Math3
Covariance matrix
3.38885
2.29824    5.19791
 2.31881    3.00273    4.69663
 Sample Size=24   ! Taken as (n1+n2+...nN)/N rounded to
          ! nearest integer
Latent Variables: Factor1
Relationships
Math1=1*Factor1
Math2-Math3=Factor1
```

Group 2: Within Schools JSP data (Level 1)
Covariance matrix
47.04658
38.56798   55.37006
30.81049   36.04099   40.71862
Sample Size=1192 ! Total number of pupils
! Set the Variance of Factor1 Free ! Remove comment to
                      !free parameter
Set the Error Variance of Math1 Free
Set the Error Variance of Math2 Free
Set the Error Variance of Math3 Free
Path Diagram
LISREL OUTPUT ND=3
End of Problem

Table 1: Parameter estimates and standard errors for factor analysis model

	SIMPLIS		Multilevel SEM	
	Estimate	Standard error	Estimate	Standard error
Factor loadings				
λ_{11}	1.000	-	1.000	-
λ_{21}	1.173	0.031	1.177	0.032
λ_{31}	0.939	0.026	0.947	0.028
Factor variance				
Ψ	32.109	1.821	31.235	1.808
Error variances (between)				
Math1	1.640	0.787	1.656	0.741
Math2	2.123	1.059	2.035	0.942
Math3	1.868	0.779	1.840	0.734
Error variances (within)				
Math1	14.114	0.810	14.209	0.890
Math2	10.274	0.884	10.256	0.993
Math3	11.910	0.699	11.837	0.806
Chi-square	36.233		46.56	
Degrees of freedom	3		3	

Table 1 shows the parameter estimates, estimated standard errors and χ^2-statistic values obtained from the SIMPLIS output and from the multilevel SEM output respectively.

Remarks:

1. The between-groups sample size of 26 used in the SIMPLIS syntax file was computed as $\frac{1}{N}\sum_{i=1}^{N}n_i$, where N is the number of schools and n_i the number of children within school i. Since this value is only used to obtain starting values, it is not really crucial how the between-group sample size is computed. See, for example, Muthén (1990,1991) for an alternative formula.

2. The within-group sample size of 1192 used in the SIMPLIS file syntax is equal to the total number of school children.

3. The number of missing values per variable is as follows:

   Math1: 38
   Math2: 63
   Math3: 239

   The large percentage missing for the Math3 variable may partially explain the relatively large difference in χ^2-values from the SIMPLIS and multilevel SEM outputs.

4. If one allows for the factor variance parameter to be free over groups, the χ^2 fit statistic becomes 1.087 at 2 degrees of freedom. The total number of multilevel SEM iterations required to obtain convergence equals eight.