



Multi-sample analyses: analysis based on covariance matrices

LISREL can be used to analyze data from several samples simultaneously according to LISREL models for each group with some or all parameters constrained to be equal over groups.

Consider a set of G populations. These may be different nations, states, or regions, culturally or socioeconomically different groups, groups of individuals selected on the basis of some known or unknown selection variables, groups receiving different treatments, etc. In fact, they may be any set of mutually exclusive groups of individuals that are clearly defined. It is assumed that a number of variables have been measured on a number of individuals from each population. This approach is particularly useful in comparing a number of treatment and control groups, regardless of whether individuals have been assigned to the groups randomly or not.

It is assumed that a LISREL model of the form

$$\begin{aligned}\boldsymbol{\eta} &= \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \\ \mathbf{y} &= \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon} \\ \mathbf{x} &= \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta}\end{aligned}$$

holds in each group. The model for group g refers to the g -th group, $g = 1, 2, \dots, G$. Each of these matrices may contain fixed, free, and constrained parameters as before. If there are no constraints across groups, each group can be analyzed separately. However, if there are constraints across groups, the data from all groups must be analyzed simultaneously to get fully efficient estimates of the parameters.

Multi-sample LISREL analysis can be used to test whether the covariance or correlation matrices of the observed variables are equal for different groups.

To test the equality of covariance matrices of \mathbf{x} , one specifies $\boldsymbol{\Lambda}_x^{(g)} = \mathbf{I}$ and $\boldsymbol{\Theta}_\delta^{(g)} = \mathbf{0}$ for all groups and that $\boldsymbol{\Lambda}_x^{(g)}$ is a diagonal matrix of standard deviations of \mathbf{x} for the g -th group. The test of equality of correlation matrices is then equivalent to the specification that

$$\boldsymbol{\Phi}^{(1)} = \boldsymbol{\Phi}^{(2)} = \dots = \boldsymbol{\Phi}^{(G)},$$

all Φ -matrices having ones in the diagonal. The standard deviations in the diagonal of the Λ_x -matrices are estimated from the data and their estimates are not necessarily equal to the sample standard deviations in each group.

One can also test various forms of less strong equalities. For examples, with measurement models of the form given at the top of this document, if the measurement properties of the observed variables are the same in all groups, one would postulate that

$$\begin{aligned}\Lambda_y^{(1)} &= \Lambda_y^{(2)} = \dots = \Lambda_y^{(G)}, \\ \Lambda_x^{(1)} &= \Lambda_x^{(2)} = \dots = \Lambda_x^{(G)},\end{aligned}$$

and perhaps also that

$$\begin{aligned}\Theta_\varepsilon^{(1)} &= \Theta_\varepsilon^{(2)} = \dots = \Theta_\varepsilon^{(G)}, \\ \Theta_\delta^{(1)} &= \Theta_\delta^{(2)} = \dots = \Theta_\delta^{(G)}.\end{aligned}$$

The possible differences between groups would then be represented by differences in the distributions of the latent variables, i.e., by $\Phi^{(g)}$ and $\Psi^{(g)}$. By postulating

$$\begin{aligned}\mathbf{B}^{(1)} &= \mathbf{B}^{(2)} = \dots = \mathbf{B}^{(G)}, \\ \Gamma^{(1)} &= \Gamma^{(2)} = \dots = \Gamma^{(G)},\end{aligned}$$

one can test the hypothesis that the structural relations are also invariant over groups.

In general, any degree of invariance can be tested, from the one extreme where all parameters are assumed to be invariant over groups to the other extreme when there are no constraints across groups. For example, if $\Phi^{(g)}$, $g = 1, 2, \dots, G$, are matrices of order 3×3 , one can specify that $\phi_{31}^{(g)}$, $\phi_{32}^{(g)}$, and $\phi_{33}^{(g)}$ should be equal across groups whereas $\phi_{11}^{(g)}$, $\phi_{21}^{(g)}$, and $\phi_{22}^{(g)}$ are free parameters for each group.

To estimate all the models simultaneously, LISREL minimizes the fit function

$$F = \sum_{g=1}^G (N_g / N) F_g (\mathbf{S}^{(g)}, \Sigma^{(g)}, \mathbf{W}^{(g)}),$$

where F_g is any of the fit functions defined for ULS, GLS, ML, DWLS, and WLS. Here N_g is the sample size in group g and $N = N_1 + N_2 + \dots + N_G$ is the total sample size; \mathbf{S}_g and Σ_g are the sample and population covariance matrices in group g , and $\mathbf{W}^{(g)}$ is the weight matrix for group g . When $G = 1$, the function shown above reduces to the fit function

$$\begin{aligned}F(\boldsymbol{\theta}) &= (\mathbf{s} - \boldsymbol{\sigma})' \mathbf{W}^{-1} (\mathbf{s} - \boldsymbol{\sigma}) \\ &= \sum_{g=1}^k \sum_{h=1}^g \sum_{i=1}^k \sum_{j=1}^i w^{h,ij} (s_{gh} - \sigma_{gh})(s_{ij} - \sigma_{ij}),\end{aligned}$$

where

$$\mathbf{s}' = (s_{11}, s_{21}, s_{22}, s_{31}, \dots, s_{kk}),$$

is a vector of the elements in the lower half, including the diagonal, of the covariance matrix \mathbf{S} of order $k \times k$ used to fit the model to the data.

$$\boldsymbol{\sigma}' = (\sigma_{11}, \sigma_{21}, \sigma_{22}, \sigma_{31}, \dots, \sigma_{kk}),$$

is the vector of corresponding elements of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ reproduced from the model parameters $\boldsymbol{\theta}$; and $w^{gh,ij}$ is a typical element of a positive definite matrix \mathbf{W}^{-1} of order $u \times u$, where $u = k(k+1)/2$.

Initial estimates are first computed by IV and TSLS for each group separately, as before, ignoring all equality constraints among groups. Parameters which are specified to be equal across groups are then replaced by their mean value.

As before, the χ^2 goodness-of-fit measure is defined as N times the minimum of F . This is a measure of the fit of all LISREL models in all groups, including all constraints, to the data from all groups. The degrees of freedom are

$$d = \frac{1}{2}G(p+q)(P+q+1) - t,$$

where t is the total number of independent parameters estimated in all groups. Thus, in a multi-sample analysis, only one χ^2 goodness-of-fit measure is given, whereas the GFI, AGFI, and RMR measures are given for each group.

The command files for multi-sample analyses are stacked after each other. They are set up as usual, with the following additional rules:

- NG must be defined on the DA command for the first group.
- For each group g ($g = 2, 3, \dots, G$), every option or keyword that has the same value as in the previous group may be omitted.
- Pattern matrices and non-zero fixed values as well as starting values are defined as before. A matrix element such as BE(4,3), with one or two indices, refers to the element in the current group. To refer to an element in another group, one must use three indices where the first one refers to the group number. For example, BE(2,4,3) refers to β_{43} in $\mathbf{B}^{(2)}$.
- To define equality constraints between groups, one specifies the constrained elements as free for the first group, and equality constraints in each of the other groups. For example, if β_{43} is to be invariant over groups, one specifies
in group 1: FR BE(4,3)
in group 2: EQ BE(1,4,3) BE(4,3)
in group 3: EQ BE(1,4,3,) BE (4,3), etc.
- If a matrix is specified as ID or ZE in group 1, it must not be specified as DI, FU, or SY in subsequent groups. Similarly, if a matrix is specified as DI in group 1, it must not be specified as FU or SY in any subsequent group.

- In addition to the matrix specifications on the MO command, the following specifications are possible on the MO command for groups 2, 3, ..., G:
 - SP means that the matrix has the *same pattern* of fixed and free elements as the corresponding matrix in the previous group.
 - SS means that the matrix will be given the *same starting values* as in the corresponding matrix in the previous group.
 - PS means *same pattern and starting values* as the corresponding matrix in the previous group.
 - IN means that the matrix is *invariant* over groups, i.e., all parameter matrices have the same pattern of fixed and free elements, and all elements which are defined as free in group 1 are supposed to be equal across groups.

In principle, NY, NX, NE, and NK must all be the same in all groups. However, if the number of variables are different in different groups, it is possible to introduce pseudo-variables (observed or latent) so as to make the number of variables equal in all groups. These pseudo-variables are artificial variables that, if chosen properly, have no effects on anything.

It is important to note that the number of variables and the form (ZE, ID, DI, SY, FU) of each parameter matrix is specified on the MO command for the first groups. These specifications must not be contradicted by a different specification on the MO command for subsequent groups. This does not refer to the specification of a fixed or free matrix, the matrix mode.

We illustrate by considering an example of the testing of equality of factor structures (Sörbom (1976)). The covariance matrices in the table below are based on scores on the ETS Sequential Test of Educational Progress (STEP) for two groups of boys who took the test in both Grade 5 and Grade 7. The two groups were defined accordingly to whether or not they were in the academic curriculum in Grade 12.

Table: Covariance matrices for STEP Reading and Writing for academic and non-academic boys

| Boys Academic ($N = 373$) | | | | |
|---------------------------------|---------|---------|---------|---------|
| STEP Reading, Grade 5 | 281.349 | | | |
| STEP Writing, Grade 5 | 184.219 | 182.821 | | |
| STEP Reading, Grade 7 | 216.739 | 171.699 | 283.289 | |
| STEP writing, Grade 7 | 198.376 | 153.201 | 208.837 | 246.069 |
| Boys Non-Academic ($N = 249$) | | | | |
| STEP Reading, Grade 5 | 174.485 | | | |
| STEP Writing, Grade 5 | 134.468 | 161.869 | | |
| STEP Reading, Grade 7 | 129.840 | 118.836 | 228.449 | |
| STEP writing, Grade 7 | 102.194 | 97.767 | 136.058 | 180.460 |

A: Test the hypothesis $H_{\Sigma} : \Sigma^{(1)} = \Sigma^{(2)}$

B: Assuming that a measurement (factor analysis) model of the form $\mathbf{x} = \Lambda_x \boldsymbol{\xi} + \boldsymbol{\delta}$ holds in both groups, test the hypothesis that there are two correlated common factors in both groups with a factor pattern of the form

$$\Lambda_x = \begin{bmatrix} * & 0 \\ * & 0 \\ 0 & * \\ 0 & * \end{bmatrix}$$

C: Assuming **B**, test the hypothesis $H_{\Lambda} : \Lambda_x^{(1)} = \Lambda_x^{(2)}$

D: Assuming **C**, test the hypothesis $H_{\Lambda\Theta} : \Theta_{\delta}^{(1)} = \Theta_{\delta}^{(2)}$

E: Assuming **D**, test the hypothesis $H_{\Lambda\Phi\Theta} : \Phi^{(1)} = \Phi^{(2)}$

All five problems may be solved by using a LISREL Submodel 1, i.e.,

$$\mathbf{x} = \Lambda_x \boldsymbol{\xi} + \boldsymbol{\delta}.$$

In problem **A**, we set $\Lambda_x^{(1)} = \Lambda_x^{(2)} = \mathbf{I}$ and $\Theta_{\delta}^{(1)} = \Theta_{\delta}^{(2)} = \mathbf{0}$. This means that $\mathbf{x} = \boldsymbol{\xi}$ so that $\Sigma^{(g)} = \Phi^{(g)}$ for $g = 1, 2$. The hypothesis $\Sigma^{(1)} = \Sigma^{(2)}$ is therefore the same as $\Phi^{(1)} = \Phi^{(2)}$. The command file **EX91A.LIS** given in the **LISREL Examples** folder is extremely simple, as follows (the data file **EX91.DAT** contains both labels and covariance matrices).

```
TESTING EQUALITY OF FACTOR STRUCTURES.  HYPOTHESIS A.  GROUP: BOYS ACADEMIC
DA NG=2 NI=4 NO=373
LA FI=EX91.DAT
CM FI=EX91.DAT
MO NX=4 NK=4 LX=ID TD=ZE
OU
TESTING EQUALITY OF FACTOR STRUCTURES.  HYPOTHESIS A.  GROUP: BOYS NON-ACADEMIC
DA NO=249
LA FI=EX91.DAT
CM FI=EX91.DAT
MO PH=IN
OU
```

In group 1, Φ is free by default. In group 2, Φ is declared invariant (PH = IN). Note that on the MO command for the second group, the parameters NX, NL, LX, and TD need not be given since they are the same for the first group.

In problem **B**, we assume a common factor model with two common factors and leave $\Phi(2 \times 2)$ to be a free covariance matrix in both groups. To fix the scale for the two factors ξ_1 and ξ_2 , we fix the elements λ_{11} and λ_{32} of Λ_x equal to one for both groups. The other elements λ_{21} , λ_{42} and the diagonal elements of Θ_{δ} are free in both groups. The hypothesis in problem **B** does not impose any equality constraints on parameters; it only states that the *number* of factors is the same for both groups. The command file for problem **B** is (**EX91B.LIS**):

```
TESTING EQUALITY OF FACTOR STRUCTURES.  HYPOTHESIS B.  GROUP: BOYS ACADEMIC
DA NG=2 NI=4 NO=373
LA FI=EX91.DAT
```

```

CM FI=EX91.DAT
MO NX=4 NK=2
FR LX 2 1 LX 4 2
VA 1 LX 1 1 LX 3 2
OU
TESTING EQUALITY OF FACTOR STRUCTURES. HYPOTHESIS B. GROUP: BOYS NON-ACADEMIC
DA NO=249
LA FI=EX91.DAT
CM FI=EX91.DAT
MO LX=PS
OU

```

The matrices $\Phi^{(1)}$ and $\Phi^{(2)}$ are both free by default. The matrices $\Theta_{\delta}^{(1)}$ and $\Theta_{\delta}^{(2)}$ are diagonal and free, also by default. The elements λ_{21} and λ_{42} are declared free in group one. Starting values for these are estimated by the program using two-stage least-squares. In group 2, Λ_x is specified by the same pattern and the same starting values as for group 1 (LX = PS).

As no equality constraints across groups are imposed, the overall χ^2 -value obtained in problem **B** is the sum of the two χ^2 's that would be obtained if the two groups were analyzed separately.

In problems **C**, **D**, and **E**, the input for group 1 is the same as in problem **B**. In problem **C**, the input for group 2 differs from that of problem **B** only in that Λ_x is declared invariant, i.e., LX = IN instead of only LX = PS. In problem **D**, TD = IN in addition and in problem **E**, PH = IN in addition. The command files for the three problems may be stacked together to let LISREL analyze hypotheses **C**, **D**, and **E** in one run.

In the output file from a multi-sample analysis, all requested parts of the output are given for each group. Note that the GFI and root mean squared residual are given for each group, but the χ^2 measure is only given for the last group. This χ^2 is a measure of the overall fit of *all* models in *all* groups.

The results of the tests are given in the table below. The most reliable hypothesis to retain is H_{Λ} . The two groups have an invariant factor pattern, but there is some evidence that they differ in error variances and in factor covariance matrices.

Table: Summary of results

| Problem | Hypothesis | χ^2 | d.f. | p-value | Decision |
|----------|-------------------------|----------|------|---------|----------|
| A | H_{Σ} | 38.142 | 10 | 0.000 | Rejected |
| B | $H_{n=2}$ | 1.52 | 2 | 0.468 | Accepted |
| C | H_{Λ} | 8.79 | 4 | 0.067 | Accepted |
| D | $H_{\Lambda\Theta}$ | 21.581 | 8 | 0.006 | Rejected |
| E | $H_{\Lambda\Phi\Theta}$ | 38.28 | 11 | 0.000 | Rejected |