



## Multiple groups: testing equality of factor structures

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, and control groups, etc. In fact, there 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.

Any LISREL model may be specified and fitted for each group of data. However, LISREL assumes by default that the models are identically the same over groups, i.e., all relationships and all parameters are the same in each group. Thus, only differences between groups need to be specified. Our first example of a multi-group analysis clarifies how this is done.

In the table below, observed covariance matrices are given for two samples ( $N_1 = 865$ ,  $N_2 = 900$ , respectively) of candidates who took the Scholastic Aptitude Test in January 1971. The four measures are, in order,

VERBAL40 = a 40-item verbal aptitude section,

VERBAL50 = a separately timed 50-item verbal aptitude section,

MATH35 = a 35-item math aptitude section, and

MATH25 = a separately timed 25-item math aptitude section.

**Table 1: Covariance matrices for SAT Verbal and Math sections**

*Covariance matrix for group 1*

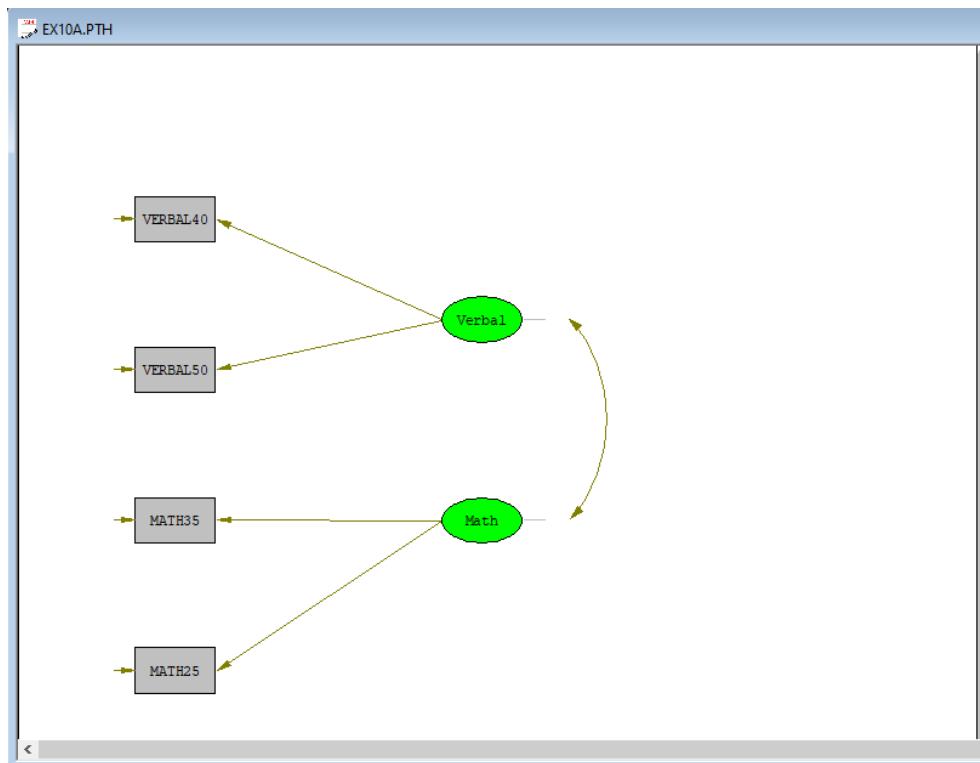
Tests	VERBAL40	VERBAL50	MATH35	MATH25
VERBAL40	63.382			
VERBAL50	70.984	110.237		
MATH35	41.710	52.747	60.584	
MATH25	30.218	37.489	36.392	32.295

*Covariance matrix for group 2*

Tests	VERBAL40	VERBAL50	MATH35	MATH25
VERBAL40	67.898			
VERBAL50	72.301	107.330		
MATH35	40.549	55.347	63.203	
MATH25	28.976	38.896	39.261	35.403

Werts, Rock, Linn & Jöreskog (1977) used these data to test various assumptions about the psychometric properties of the tests both within and between groups. Werts, Rock, Linn, & Jöreskog (1976) showed how to compare correlations, variances, covariances and regression coefficients between groups. Here we use the data to illustrate how one can test equality of factor structures in a confirmatory factor analysis model.

We regard VERBAL40 and VERBAL50 as indicators of a latent variables Verbal and MATH35 and MATH25 as indicators of a latent variable Math. The model we consider is shown below. The model is a simple confirmatory factor analysis model.



There are three sets of parameters in the model:

1. The four factor loadings corresponding to the paths from Verbal and Math to the observed variables,
2. The correlation between Verbal and Math, and
3. The four error variances of the observed variables.

We want to investigate to what extent each of these sets of parameters are invariant over groups.

We begin by constructing a data file **EX10.COV** containing the two covariance matrices. In free format, each covariance matrix can be written on one line. File **EX10.COV** looks like this:

```
63.382 70.984 110.237 41.710 52.747 60.584 30.218 37.489 36.392 32.295
67.898 72.301 107.330 40.549 55.347 63.203 28.976 38.896 39.261 35.403
```

In the first model (Model A) we assume all parameters are the same in both groups. The input file **EX10.SPL** (see the **SIMPLIS examples** folder) is:

```

Group 1: Testing Equality Of Factor Structures
Model A: Factor Loadings, Factor Correlation, Error Variances Invariant
Observed Variables: VERBAL40 VERBAL50 MATH35 MATH25
Covariance Matrix from File EX10.COV
Sample Size = 865
Latent Variables: Verbal Math
Relationships:
    VERBAL40 VERBAL50 = Verbal
    MATH35 MATH25 = Math

```

Group 2: Testing Equality Of Factor Correlations  
Covariance Matrix from File EX10.COV  
Sample Size = 900  
Path Diagram  
End of Problem

The new element here is that we are reading specifications for two groups in the same input file. For each group, the input is written as described in previous examples. To specify the beginning of each group, the first word in the first title line for each group begins with the word Group. The specifications for groups 2, 3, 4, ..., may be considerable simplified, because of the general principle that everything is the same as in the previous group unless otherwise stated. Thus, in the input above, only two things differ between group 1 and group 2 – namely, the data in the covariance matrices and the sample sizes. The names of the variables, observed as well as latent, are the same, and the model is the same. Notice that no relationships are specified from Group 2, which implies that they are the same for Group 1.

The output file reveals two solutions, one for each group, but the parameter estimates are all identical. The value of chi-square is reported only after the second group. In a multi-sample analysis, the chi-square is a measure of fit of all models in the group, and, in general, this chi-square cannot be decomposed into a chi-square for each group separately. For our example, the chi-square is

Degrees of Freedom for (C1)-(C2)	11
Maximum Likelihood Ratio Chi-Square (C1)	34.915 (P = 0.0003)

so we must seek a less constrained model. Suppose therefore that we allow the factor loadings to be different for the two groups, retaining the invariance of the factor correlation and the error variances. The input for such a model is (**EX10B.SPL**):

Group 1: Testing Equality Of Factor Structures  
Model B : Factor Correlation and Error Variances Invariant  
Observed Variables: VERBAL40 VERBAL50 MATH35 MATH25  
Covariance Matrix from File EX10.COV  
Sample Size = 865  
Latent Variables: Verbal Math  
Relationships:  
VERBAL40 VERBAL50 = Verbal  
MATH35 MATH25 = Math

Group 2: Testing Equality Of Factor Correlations  
Covariance Matrix from File EX10.COV  
Sample Size = 900  
Relationships:  
VERBAL40 VERBAL50 = Verbal  
MATH35 MATH25 = Math  
Path Diagram  
End of Problem

The only difference between this and the previous input is that the relationships of the model are also specified in the second group. Note that this implies that the coefficients in the relationships will be estimated for the second group independently of the first. But since nothing is specified about the factor correlation and the error variances, these are assumed to be the same for both groups. The chi-square for Model B is

Degrees of Freedom for (C1)-(C2)	7
Maximum Likelihood Ratio Chi-Square (C1)	29.687 ( $P = 0.0001$ )

which also does not indicate a good fit. Continuing the example, let us assume that, in addition to the factor loadings, the error variances may also be different over groups, still retaining the assumption that the factor correlation is the same. To do so, add the line

Set the Error Variances of VERBAL40 - MATH25 free

before the last line (see **EX10C.SPL**). This allows the error variances for the second group to be different from those of the first group.

Model C gives a chi-square of 4.03 with three degrees of freedom ( $P = 0.26$ ), which indicates a good fit. The difference in chi-square between Models B and C is 25.64 with four degrees of freedom, and that between Models A and B is 5.23 with four degrees of freedom. This suggests that it is really the error variances that differ between groups and not the factor loadings. We may therefore want to do one final analysis of this data, namely one in which factor loadings and factor correlations are the same for both groups and only the error variances are allowed to be different. This may be done with the following input file (**EX10D.SPL**):

Group 1: Testing Equality Of Factor Structures  
Model D: Factor Loadings and Factor Correlation Invariant  
Observed Variables: VERBAL40 VERBAL50 MATH35 MATH25  
Covariance Matrix from File EX10.COV  
Sample Size = 865  
Latent Variables: Verbal Math  
Relationships:  
VERBAL40 VERBAL50 = Verbal  
MATH35 MATH25 = Math

Group 2: Testing Equality Of Factor Correlations  
Covariance Matrix from File EX10.COV  
Sample Size = 900  
Set the Error Variances of VERBAL40 - MATH25 free  
Path Diagram  
End of Problem

The chi-square for this model is 10.87 with 7 degrees of freedom ( $P = 0.14$ ). Note that this model fits the data, whereas Model B – with the same degrees of freedom – does not.