Confirmatory factor analysis

In exploratory factor analysis, the objective is to find, for a given set of response variables $x_1, ..., x_q$, a set of underlying factors $\xi_1, ..., \xi_n$, fewer in number than the observed variables. These factors are supposed to account for the intercorrelations of the response variables in the sense that when the factors are partialed out no correlation between variables should remain. The assumed model is

$$\mathbf{x} = \mathbf{\Lambda \xi} + \mathbf{\delta},$$

where $E(\mathbf{\xi}) = 0$ and $E(\mathbf{\delta}) = 0$, $\mathbf{\delta}$ being uncorrelated with $\mathbf{\xi}$.

If $\mathbf{\Phi} = E(\mathbf{\xi\xi}')$ is taken as a correlation matrix and $\mathbf{\Theta} = E(\mathbf{\delta\delta}')$ is diagonal, the covariance $\mathbf{\Sigma}$ of $\mathbf{x}$ is

$$\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Theta}.$$

If $(q-n)^2 < q+n$, this relationship can be tested statistically, unlike the first equation which involves hypothetical variables and cannot be verified directly.

When $n > 1$ there is an indeterminancy in the equation above arising from the fact that a nonsingular linear transformation of $\mathbf{\xi}$ changes $\mathbf{\Lambda}$ and in general also $\mathbf{\Phi}$ but leaves $\mathbf{\Sigma}$ unchanged. The usual way to deal with this indeterminancy in exploratory factor analysis is to choose $\mathbf{\Phi} = \mathbf{I}$ and $\mathbf{\Lambda}'\mathbf{\Phi\Lambda}$ or $\mathbf{\Lambda}'\mathbf{\Lambda}$ to be diagonal and to estimate the parameters in $\mathbf{\Lambda}$ and $\mathbf{\Theta}$ subject to these conditions. This leads to an arbitrary set of factors, which may then be subjected to a rotation or a linear transformation to another set of factors that may be given a more meaningful interpretation. Thurstone's principle of simple structure or Kaiser's varimax criterion are often used in these rotations.

In a confirmatory factor analysis, the investigator has such knowledge about the factorial nature of the variables that he or she is able to specify at least n^2 independent conditions on $\mathbf{\Lambda}$ and $\mathbf{\Phi}$. The most common way of doing this, assuming that $\mathbf{\Phi}$ is an unconstrained correlation matrix, is to set at least $n-1$ zeros in each column of $\mathbf{\Lambda}$. When there are many variables it usually suffices that the zeros are distributed over the rows of $\mathbf{\Lambda}$ in such a way that $\mathbf{\Lambda}$ has full column rank. The following example illustrates confirmatory factor analysis, including in particular the assessment of model fit and the use of the model modification index.

The table below shows correlations based on data for nine psychological tests administered to 145 seventh- and eighth-grade children (Holzinger & Swineford, (1939)). For this example, three common factors are hypothesized: visual perception, verbal ability, and speed. The first three variables are assumed to measure Visual, the next three to measure Verbal, and the last three to measure Speed.

The path diagram of the assumed factor model is shown below.

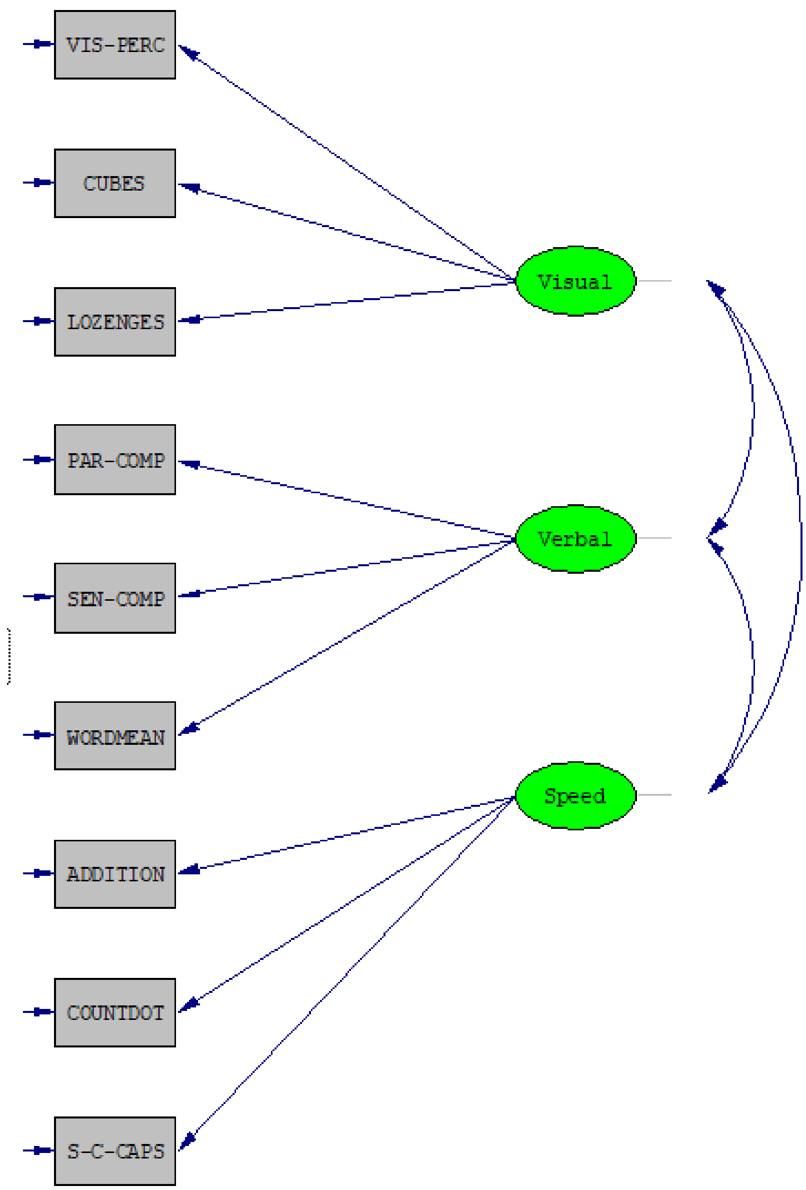

Table: Correlation matrix for nine psychological variables

VIS-PERC	1.000								
CUBES	0.318	1.000							
LOZENGES	0.436	0.419	1.000						
PAR-COMP	0.335	0.234	0.323	1.000					
SEN-COMP	0.304	0.157	0.283	0.722	1.000				
WORDMEAN	0.326	0.195	0.350	0.714	0.685	1.000			
ADDITION	0.116	0.057	0.056	0.203	0.246	0.170	1.000		
COUNTDOT	0.314	0.145	0.229	0.095	0.181	0.113	0.585	1.000	
S-C CAPS	0.489	0.239	0.361	0.309	0.345	0.280	0.408	0.512	1.000

The LISREL input for this analysis is (**EX34.LIS** in the **LISREL Examples** folder):

```
NINE PSYCHOLOGICAL VARIABLES - A CONFIRMATORY FACTOR ANALYSIS
DA NI=9 NO=145 MA=KM
LA
VIS-PERC CUBES LOZENGES PAR-COMP SEN-COMP WORDMEAN ADDITION
COUNTDOT S-C-CAPS
KM FI=EX34.COR
MO NX=9 NK=3 PH=ST
LK
Visual Verbal Speed
PA LX
3(1 0 0) 3(0 1 0) 3(0 0 1)
PL LX(9,1)
OU SE TV MI
```

Both labels and correlations are read in free format without any format lines. The new part of the input is in the lines following labels for the latent variables. Rather than a listing of all free elements of $\mathbf{\Lambda}$ on an FR command, a pattern matrix of 0’s and 1’s is entered. this is convenient because the first three rows of the pattern matrix are all equal to 1 0 0, then next three rows are all equal to 0 1 0, etc. One line contains the whole pattern matrix of order 9 x 3.

The PL command requests the program to plot the fitting function (ML in this case) against the parameter λ_{91}.

The OU command requests modification indices (MI) as well as standard errors and *t*-statistics.

The ML solution produced by this input looks reasonable except for the fact that χ^2 is rather large: 52.63 with 24 degrees of freedom.

Output for this model is shown below:

Modification Indices and Expected Change

Modification Indices for LAMBDA-X

	Visual	Verbal	Speed
VIS-PERC	- -	0.268	3.918
CUBES	- -	0.665	0.968
LOZENGES	- -	0.032	1.346
PAR-COMP	0.003	- -	0.694
SEN-COMP	0.342	- -	2.059
WORDMEAN	0.277	- -	0.303
ADDITION	10.503	0.178	- -
COUNTDOT	2.738	10.078	- -
S-C-CAPS	24.736	10.040	- -

Expected Change for LAMBDA-X

	Visual	Verbal	Speed
VIS-PERC	- -	0.067	0.251
CUBES	- -	-0.094	-0.116
LOZENGES	- -	0.023	-0.149
PAR-COMP	0.005	- -	-0.056
SEN-COMP	-0.052	- -	0.098
WORDMEAN	0.047	- -	-0.038
ADDITION	-0.367	0.035	- -
COUNTDOT	-0.204	-0.278	- -
S-C-CAPS	0.565	0.265	- -

No Non-Zero Modification Indices for PHI

Modification Indices for THETA-DELTA

	VIS-PERC	CUBES	LOZENGES	PAR-COMP	SEN-COMP	WORDMEAN
VIS-PERC	- -					
CUBES	0.628	- -				
LOZENGES	1.832	4.355	- -			
PAR-COMP	0.042	0.740	0.044	- -		
SEN-COMP	0.007	1.293	0.631	0.173	- -	
WORDMEAN	0.013	0.130	1.395	0.127	0.003	- -
ADDITION	4.111	0.414	4.506	0.621	0.866	0.082
COUNTDOT	0.374	0.169	0.000	3.764	0.210	0.193
S-C-CAPS	9.072	0.021	1.036	0.350	0.421	0.018

```
 Modification Indices for THETA-DELTA

           ADDITION   COUNTDOT   S-C-CAPS
           --------   --------   --------
 ADDITION      - -
 COUNTDOT    25.068       - -
 S-C-CAPS     4.026      8.327       - -


  Expected Change for THETA-DELTA

           VIS-PERC      CUBES   LOZENGES   PAR-COMP   SEN-COMP   WORDMEAN
           --------   --------   --------   --------   --------   --------
 VIS-PERC      - -
    CUBES    -0.065       - -
 LOZENGES    -0.155      0.174       - -
 PAR-COMP     0.009      0.040     -0.009       - -
 SEN-COMP    -0.004     -0.056     -0.037      0.041       - -
 WORDMEAN     0.005     -0.018      0.055     -0.034     -0.005       - -
 ADDITION    -0.119     -0.040     -0.125      0.034      0.041     -0.013
 COUNTDOT     0.035     -0.024     -0.001     -0.077      0.019     -0.018
 S-C-CAPS     0.176      0.009      0.059      0.025      0.029     -0.006

         Expected Change for THETA-DELTA

           ADDITION   COUNTDOT   S-C-CAPS
           --------   --------   --------
 ADDITION      - -
 COUNTDOT     0.625       - -
 S-C-CAPS    -0.195     -0.377       - -

 Maximum Modification Index is   25.07 for Element ( 8, 7) of THETA-DELTA
```

The largest modification index is 24.74 and occurs for λ_{91}. This is highly significant. The modification index is approximately a χ^2 with one degree of freedom. This modification index suggests directly that λ_{91} should be set free and predicts that, if this is done, the overall χ^2 will decrease by about 24.74 and that λ_{91} will be approximately 0.57. This result is also shown in a plot of the concentrated fit function printed in the program output. If the model is true, the sample large, and the asymptotic theory valid, the plotted curve should be quadratic around the minimum.

To modify the model, merely add the command

```
FR LX(9,1)
```

before the OU command, as is done in **EX34B.LIS**:

```
NINE PSYCHOLOGICAL VARIABLES - A CONFIRMATORY FACTOR ANALYSIS   SECOND RUN
DA NI=9 NO=145 MA=KM
LA
VIS-PERC CUBES LOZENGES PAR-COMP SEN-COMP WORDMEAN ADDITION
COUNTDOT S-C-CAPS
KM FI=EX34.COR
MO NX=9 NK=3 PH=ST
LK
Visual Verbal Speed
PA LX
3(1 0 0) 3(0 1 0) 3(0 0 1)
FR LX(9,1)
PL LX(9,1)
OU SE TV MI
```

Some goodness-of-fit statistics from the output are shown below:

```
                          Goodness-of-Fit Statistics

 Degrees of Freedom for (C1)-(C2)                            23
 Maximum Likelihood Ratio Chi-Square (C1)                    29.062 (P = 0.1783)
 Browne's (1984) ADF Chi-Square (C2_NT)                      28.873 (P = 0.1846)
```

The overall goodness-of-fit measure for the modified model is χ^2 = 28.86 with 23 degrees of freedom. The difference between the previous χ^2 and this one is 52.63 – 28.86 = 23.77, which is reasonably close to the value 24.74 predicted by the modification index. The reason for the discrepancy is that the fit function is not quite quadratic in a region of the parameter space around the first solution. The modification index is based on a quadratic approximation of the fit function.