Mild-Restricted Confirmatory Factor Analysis

R Code manual

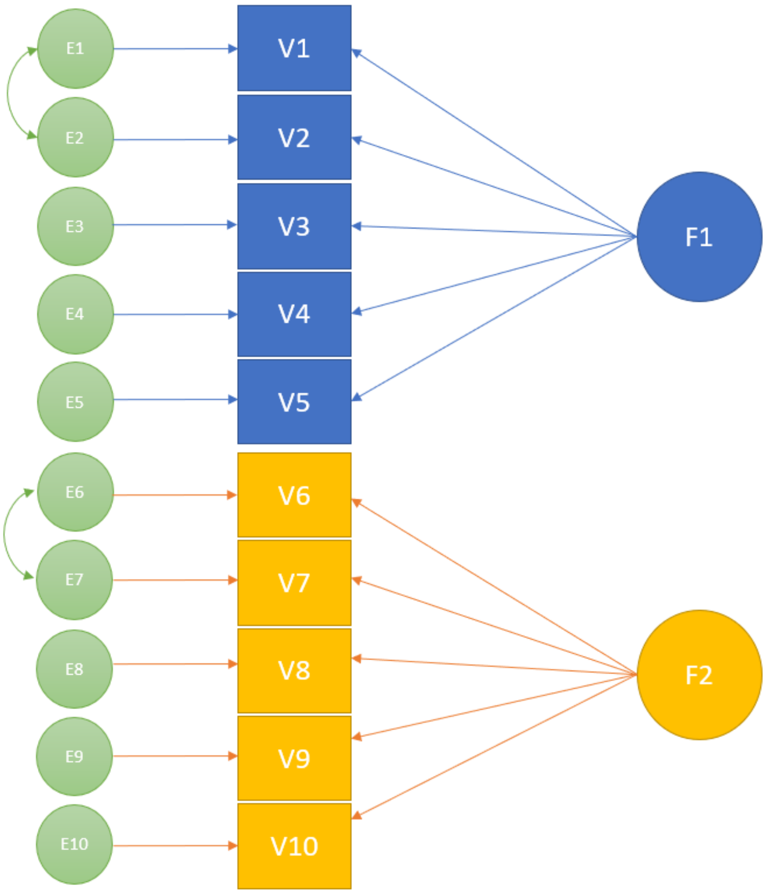
Urbano Lorenzo-Seva, 2024

Mild-restricted confirmatory factor analysis is a method for computing confirmatory factor analysis (CFA). The aim in CFA is to assess if a hypothesized factor model can be expected to be true at the population level. In CFA, it is usual that each variable in the model is expected to be a *pure indicator of a factor* at the population level: a variable that shows a single salient loading value related to the factor, and any other loading value of the variable is exactly zero. Frequently, even if a variable is strongly related to a single factor (i.e., the variable has only a large loading value associated with a factor in the model), it cannot be considered a pure indicator because the other loading values of the variable are not exactly zero, but close to zero. A variable that meets this situation can be said to be *a close indicator of a factor*. From a practical point of view, researchers tend to interpret *close indicators* as if they were *pure indicators* (i.e., secondary loadings close to zero are not given a substantive interpretation). It means that when applied researchers propose a factor model based on pure indicators, they would also accept a factor model based on close indicators. However, a CFA based on restricted factor model will reject a factor model based on pure indicators if the condition related to the secondary loadings values equal to zero is not meet. The same reasoning can be applied to other model parameters usually expected be zero at the population level: inter-factor correlations and correlated errors. A CFA based on mild-restricted factor analysis helps to assess if a factor model can be expected to exists at population level even if the variables are close indicators, or none the inter-factor correlations or the correlated errors are not exactly zero. Mild-restricted factor analysis is a three-step method. First, all the parameters of the model are estimated using unrestricted factor analysis based on Morgana. Second, parameters are restricted to the proposed values in the hypothetical model. And, third, the goodness-of-fit of the model is computed using the chi-square test statistic scaled with LOSEFER.

In this document, I explain how to assess a factor model expected at the population model using R using an illustrative example.

Illustrative example

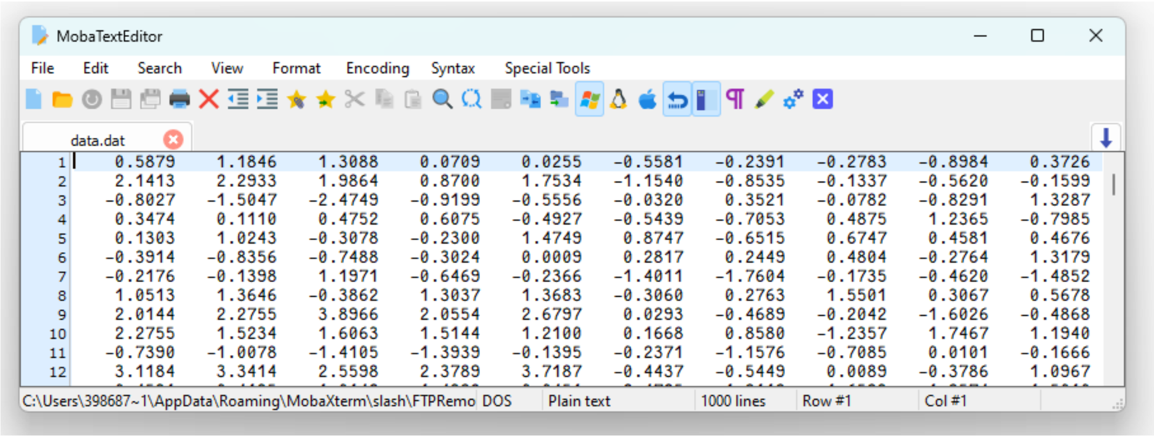
The factor model expected at the population model is the following. Ten variables are expected to define two factors. A set of first five variables are supposed to be related to the first factor, and the second set of 5 variables are expected to define the second factor. The two factors are expected to be orthogonal (i.e., the correlation among them is expected to be zero). Two pairs of correlated errors are expected to be different from zero: the pair composed by variables 1 and 2, and the pair composed by variables 6 and 7. This figure represents the model to be assessed at the population model.



While the relationship between variables and factors is expected to be in terms of pure variable indicators of the factors, for this model we can admit that the variables could be close indicators of the factors.

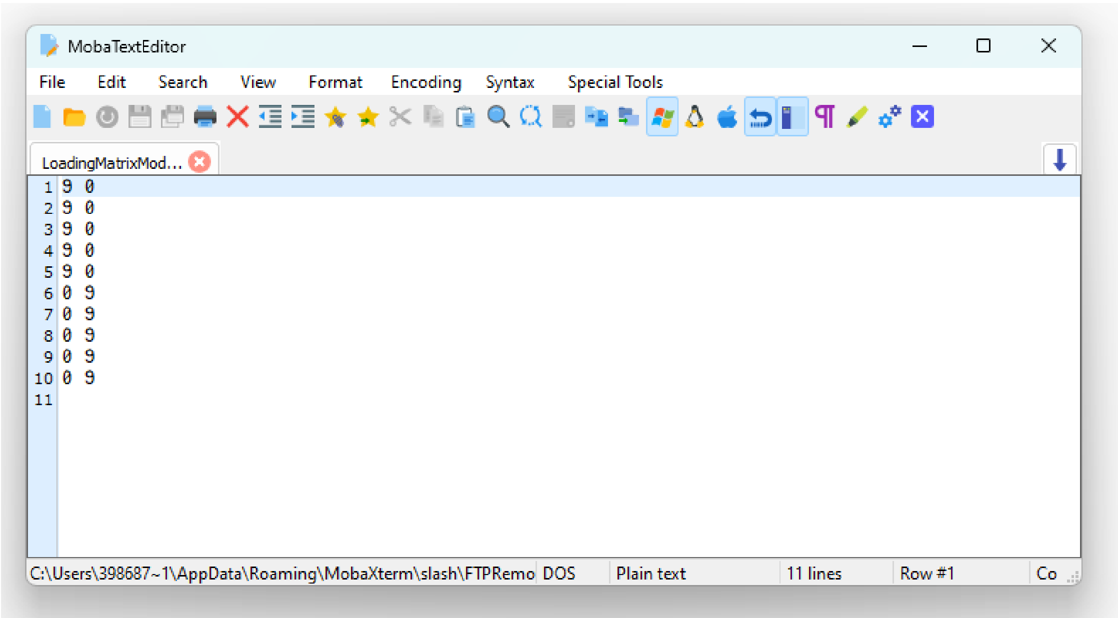
Data file

For the ten variables, a sample of 1,000 observations were obtained from the population, and stored in a text file named *data.dat*.



Specification of loading matrix parameters

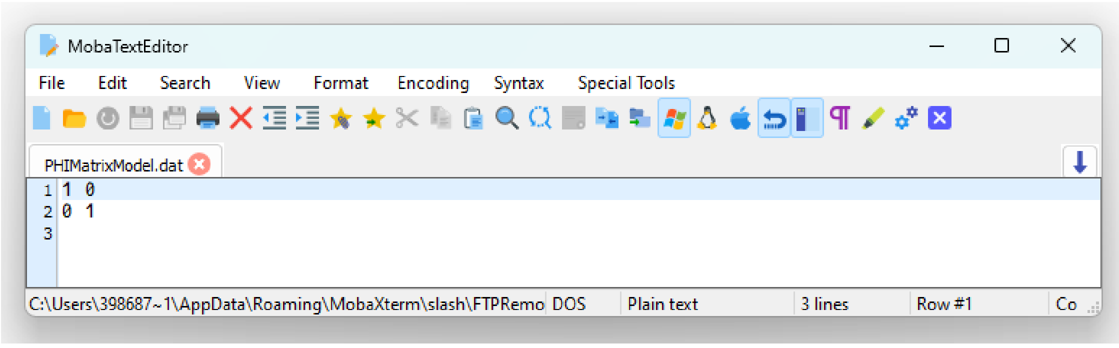
Mild-restricted factor analysis needs to be informed which loading values needs to be estimated (i.e., they are free values in the model), and which loading values are specified to a fixed value. The value specified can be zero or some other value. In this example, the specified values are set to zero. In order to inform the loading values that need to be estimated a value of 9 is used. This information is introduced in a matrix and stored in a text file named LoadingMatrixModel.dat.



Please note that the matrix must contain as many columns as factors in the model, and as many rows as variables in the model.

Specification of inter-factor correlations

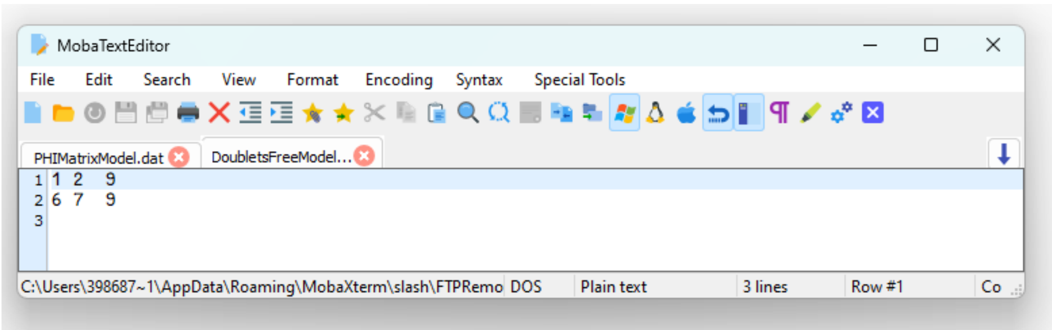
Mild-restricted factor analysis needs to be informed which inter-factor correlation values needs to be estimated (i.e., they are free values in the model), and which inter-factor correlation values are specified to a fixed value. The value specified can be zero or some other value in the range [-1; 1]. In this example, factors are expected to be orthogonal, so a inter-factor correlation of zero is specified. This information is introduced in a matrix and stored in a text file named PHIMatrixModel.dat.



Please note that the matrix must be square matrix, and the number of columns has to correspond to the number of factors in the model. The matrix also needs to be symmetrical: the same values out of the diagonal of the matrix. Finally, all the diagonal elements need to have a value of 1.

Specification of correlated errors

By default, mild-restricted factor analysis specifies that not correlated errors are to be observed in the model. In case of correlated errors, it has to be informed which pairs of items need to be estimated using a value for the pair of 9. If a specific value for a correlated error wants to be tested, the specific value can be proposed and has to be in the range [-1; 1]. In this example, two doublets are proposed to be estimated: pair of variables 1 and 2, and pair of variables 6 and 7. This information is introduced in a matrix and stored in a text file named DoubletsFreeModel.dat.



To compute Mild-restricted confirmatory factor analysis

We aim to assess if the hypothesized model presented in this illustrative analysis can be expected to be true at the population level. In order to assess the hypothesized model using mild-restricted factor analysis, we prepared the text file:

Example_MildRestrictedConfirmatoryFactorAnalysis.r

in which we introduced the name of the text files previously prepared. The most important lines of this file are shown in this figure:

```
#################################################################
#
#        MILD-RESTRICTED  FACTOR  ANALYSIS     R   CODE
#
#                 AUTHOR:   URBANO LORENZO-SEVA
#
#                     URV, TARRAGONA (SPAIN)
#                       DATE: 20/11/2023
#
# DATA EXAMPLE: 10 ITEMS AND N = 1,000
#
# DATA TEXT FILE USED:      data.dat
#
# LOADING MATRIX MODEL:     LoadingMatrixModel.dat
#
# PHI MATRIX MODEL:         PHIMatrixModel.dat
#
# DOUBLEST FREE IN MODEL:   DoubletsFreeModel.dat
#
# EXECUTE THE THIS EXAMPLES AS
#
# source("Example_MildRestrictedFactorAnalysis.r")
#
#################################################################

#################################################################
# UPDATE HERE THE NAME OF YOUR INPUT DATA FILE
#
filein ← "data.dat"
#
# UPDATE HERE THE FILE NAME WITH THE SPECIFIED LOADING MATRIX
# OF THE MODEL.
#
fileinP ← "LoadingMatrixModel.dat"
#
# UPDATE HERE THE FILE NAME WITH THE SPECIFIED PHI MATRIX OF THE MODEL.
#
 fileinPHI ← "PHIMatrixModel.dat"
#
# UPDATE HERE THE FILE NAME WITH THE DOUBLEST FREE IN THE MODEL
#
 fileinDoublets ← "DoubletsFreeModel.dat"
#
# UPDATE HERE THE NUMBER OF RANDOM SAMPLES FOR LOSEFER (AT LEAST 1000)
#
K ← 1000
#
#################################################################
source("MildRestrictedFA.rlib")
MildRestrictedFactorAnalysis(filein, fileinP, fileinPHI, fileinDoublets, K)
```

It must be noted that a data file (line 29, in the figure) and a specified loading matrix has (line 34, in the figure) are mandatory. If factors are supposed to be all of them oblique among them (i.e., all the inter-factor correlations need to be estimated), or if the model considers a single factor, then not a file needs to be prepared and line 38 in the figure could be:

```
fileinPHI <- ""
```

In addition, if not doublets are supposed in the model, then not a file needs to be prepared and line 42 in the figure could be:

```
fileinDoublets <- ""
```

Line 46 in the figure defines the number of samples used to estimate LOSEFER chi-square. A minimum of 1,000 is advised. Please not that using larger samples will help to obtain a best estimate of the statistic, but might make the computing time longer.

Once the information related to the datafiles are updated and the file saved, it can be executed using the following R line command:

```
> source("Example_MildRestrictedConfirmatoryFactorAnalysis.r")
```

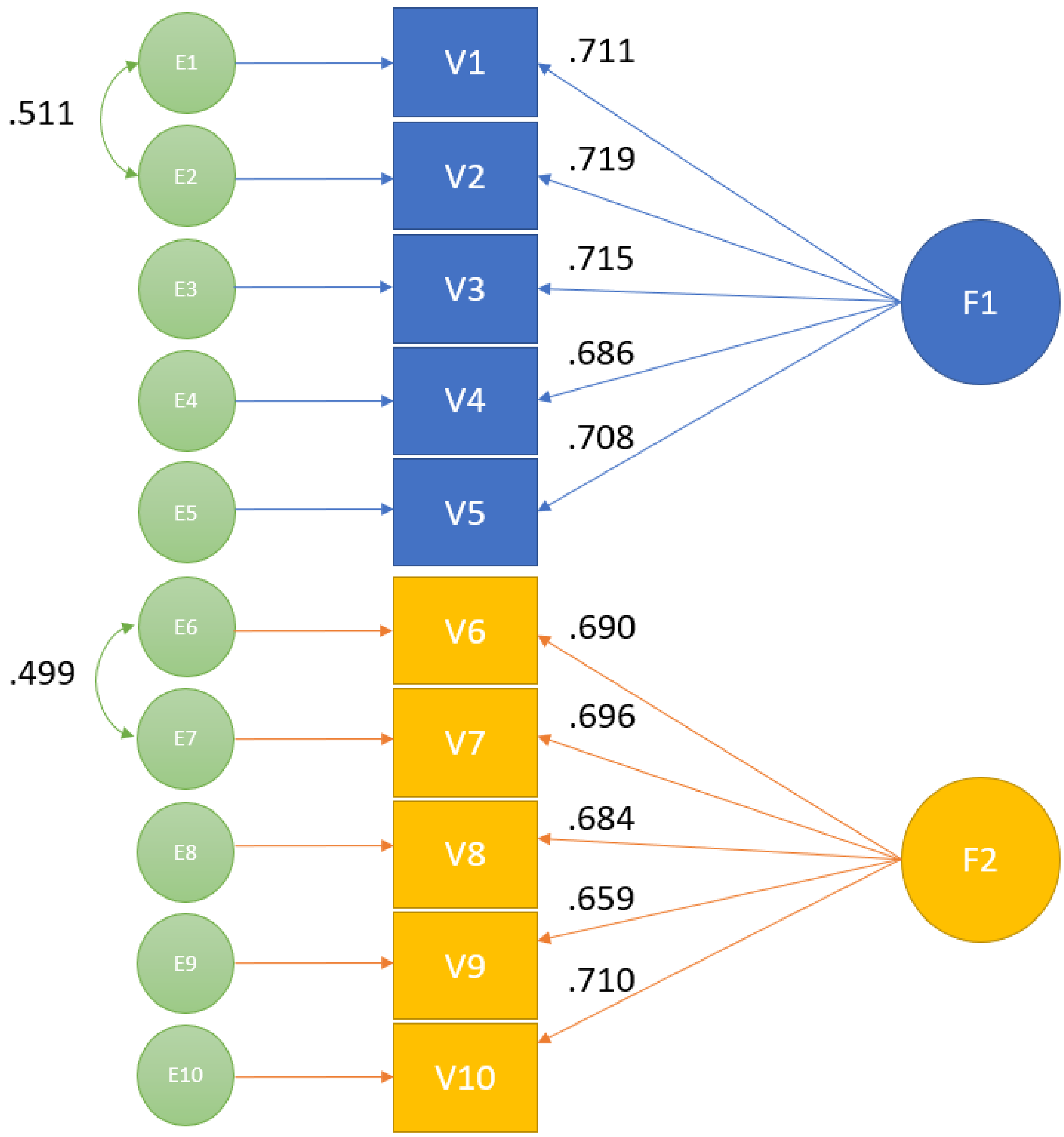
After executing this command line, the analysis will start. Please note that, depending on the characteristics of the dataset analyzed, it might take a few minutes up to a few hours.

Outcome of the analysis

The computing time for this dataset in our computer was 4 minutes and 54 seconds. The outcome obtained is shown in this figure:

```
#########################################################################
#                                                                       #
#        MILD-RESTRICTED   CONFIRMATORY   FACTOR  ANALYSIS              #
#                                                                       #
#               AUTHOR:   URBANO LORENZO-SEVA                           #
#                                                                       #
#                    URV, TARRAGONA (SPAIN)                             #
#                                                                       #
#                       DATE: 13/12/2024                                #
#                                                                       #
#########################################################################
#
# Filein                :  data.dat
# Cases                 :  1000
# Variables             :  10
# Factors               :  2


GOODNESS OF FIT INDICES

Chi Square adjusted by Losefer with  33 degrees of freedom = 40.636

                              Goodness of Fit (Null Model) = 5600.45
                                        Degrees of freedom = 45

                           Chi Square / degrees of freedom = 1.231

           RMSEA (Root Mean Square Error of Approximation) = 0.015
                               CFI (Comparative Fit Index) = 0.999
                                NNFI (Non-Normed Fit Index) = 0.998
                               GFI (Goodness of Fit Index) = 0.998
                     AGFI (Adjusted Goodness of Fit Index) = 0.998

 Advised threshold values to refuse the hypothetical factor model:

       chi-square/df > 2 (Tabachnick & Fidell, 2007)
               RMSEA > .07 (Steiger, 2007)
                 CFI < .95 (Hu & Bentler, 1999)
           GFI, AGFI < .95 (Miles & Shevlin, 1998)


PARAMETER MODEL ESTIMATES

Loading matrix

Var     F 1    F 2

V 1    -0.711  0.000
V 2    -0.719  0.000
V 3    -0.715  0.000
V 4    -0.686  0.000
V 5    -0.708  0.000
V 6     0.000  0.690
V 7     0.000  0.696
V 8     0.000  0.684
V 9     0.000  0.659
V10     0.000  0.710


Inter-factor correlation matrix: orthogonal factors


Pairs of items with correlated residuals

Doublets     Residual Correlations

 1  &  2     0.511
 6  &  7     0.499



 Hu, L. T., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis:
             Conventional Criteria Versus New Alternatives. Structural Equation Modeling, 6 (1), 1-55.
 Miles, J. & Shevlin, M. (1998). Effects of sample size, model specification and factor loadings on the GFI
             in confirmatory factor analysis, Personality and Individual Differences, 25, 85-90.
 Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling.
             Personality and Individual Differences, 42 (5), 893-98.
 Tabachnick, B. G., & Fidell, L.S. (2007). Using Multivariate Statistics (5th ed.). New York: Allyn and Bacon.

#########################################################################
```

As can be observed, the values of goodness-of-fit indices suggest that the model can be expected to be found at the population level. The estimate of the loading values shows that items are strongly related with the factor expected. As the estimate loading values in the first factor are all of them negative, the factor can be reflected. Finally, the values of the two doublets that were modelled as free parameters to be estimated show that the pairs were indeed strongly correlated. The numerical outcome can be translated now to the model figure.



Code availability

The R code and the data used in this example can be downloaded from this site:

https://www.psicologia.urv.cat/ca/utilitats/