



Getting Started in Factor Analysis (using Stata)

(ver. 1.0 beta/draft)

Oscar Torres-Reyna

Data Consultant otorres@princeton.edu

DEI

http://dss.princeton.edu/training/

Factor analysis: intro

Factor analysis is used mostly for data reduction purposes:

- To get a small set of variables (preferably uncorrelated) from a large set of variables (most of which are correlated to each other)
- To create indexes with variables that measure similar things (conceptually).

Two types of factor analysis

Exploratory

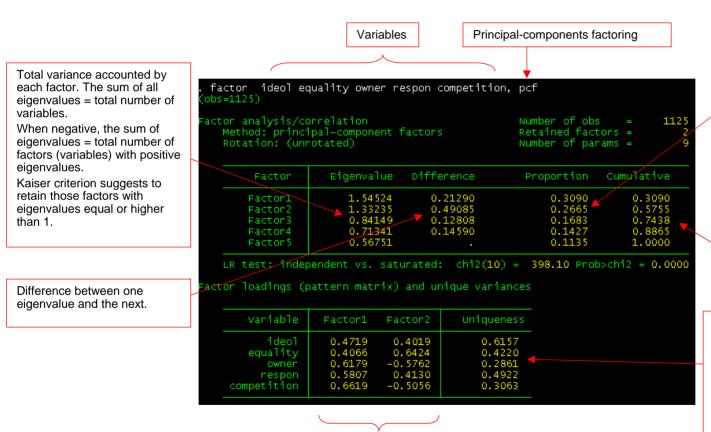
It is exploratory when you do not have a pre-defined idea of the structure or how many dimensions are in a set of variables.

Confirmatory.

It is confirmatory when you want to test specific hypothesis about the structure or the number of dimensions underlying a set of variables (i.e. in your data you may think there are two dimensions and you want to verify that).

Factor analysis: step 1

To run factor analysis use the command factor (type help factor for more details).



Since the sum of eigenvalues = total number of variables. Proportion indicate the relative weight of each factor in the total variance. For example, 1.54525/5=0.3090. The first factor explains 30.9% of the total variance

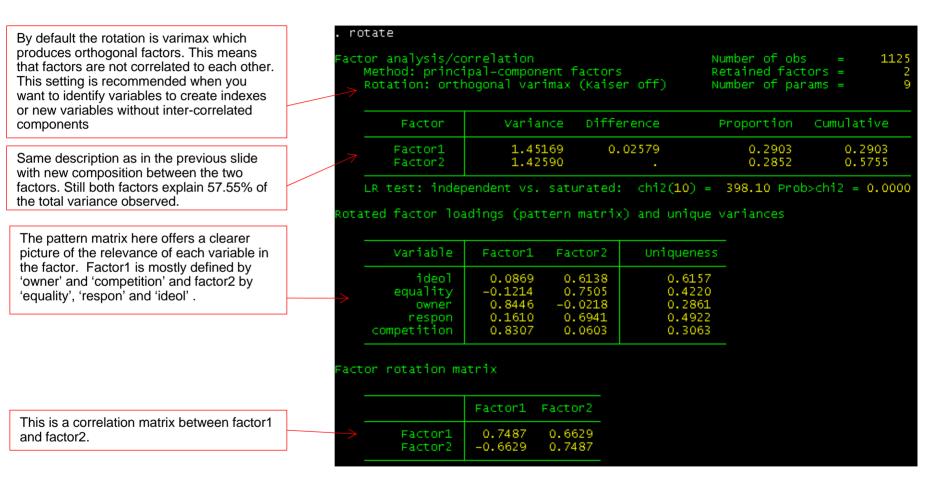
Cumulative shows the amount of variance explained by n+(n-1) factors. For example, factor 1 and factor 2 account for 57.55% of the total variance.

Uniqueness is the variance that is 'unique' to the variable and not shared with other variables. It is equal to 1 – communality (variance that is shared with other variables). For example, 61.57% of the variance in 'ideol' is not share with other variables in the overall factor model. On the contrary 'owner' has low variance not accounted by other variables (28.61%). Notice that the greater 'uniqueness' the lower the relevance of the variable in the factor model.

Factor loadings are the weights and correlations between each variable and the factor. The higher the load the more relevant in defining the factor's dimensionality. A negative value indicates an inverse impact on the factor. Here, two factors are retained because both have eigenvalues over 1. It seems that 'owner' and 'competition' define factor1, and 'equality', 'respon' and 'ideol' define factor2.

Factor analysis: step 2 (final solution)

After running factor you need to rotate the factor loads to get a clearer pattern, just type rotate to get a final solution.



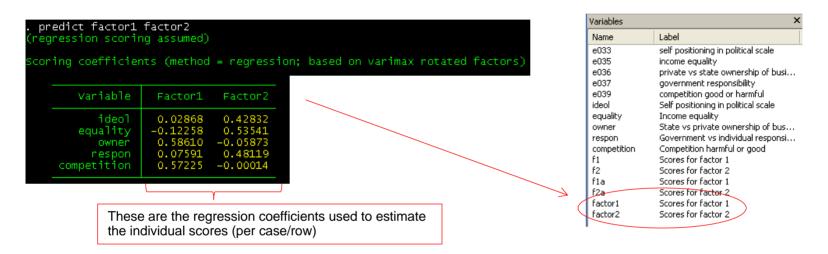
NOTE: If you want the factors to be correlated (oblique rotation) you need to use the option promax after rotate: rotate, promax

Type help rotate for details.

Factor analysis: step 3 (predict)

To create the new variables, after factor, rotate you type predict.

predict factor1 factor2 /*or whatever name you prefer to identify the factors*/



Another option could be to create indexes out of each cluster of variables. For example, 'owner' and 'competition' define one factor. You could aggregate these two to create a new variable to measure 'market oriented attitudes'. On the other hand you could aggregate 'ideol', 'equality' and 'respon' to create an index to measure 'egalitarian attitudes'. Since all variables are in the same valence (liberal for small values, capitalist for larger values), we can create the two new variables as

```
gen market = (owner + competition)/2
gen eqalitatiran = (ideol + equality + respon)/3
```

Factor analysis: sources/references

The main sources/references for this section are:

Books

- Factor Analysis in International Relations. Interpretation, Problem Areas and Application / Vincent, Jack. University of Florida Press, Gainsville, 1971.
- Factor Analysis. Statistical Methods and Practical Issues / Kim Jae-on, Charles W. Mueller, Sage publications, 1978.
- Introduction to Factor Analysis. What it is and How To Do It / Kim Jae-on, Charles W. Mueller, Sage publications, 1978.
- Statistics with STATA (updated for version 9) / Hamilton, Lawrence C. Thomson Books/Cole, 2006.

Online

- StatNotes: http://faculty.chass.ncsu.edu/garson/PA765/factor.htm
- StatSoft: http://www.statsoft.com/textbook/stfacan.html
- UCLA Resources: http://www.ats.ucla.edu/stat/stata/output/fa_output.htm