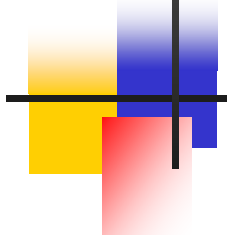




Exploratory Factor Analysis

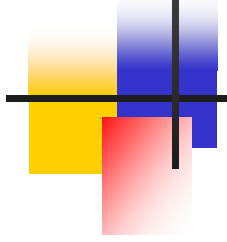
Gerrit Jan Kootstra

7 May 2004



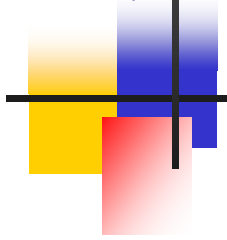
Contents

- Introduction
- Factor Analysis stepwise
- Application in SPSS (if time left)



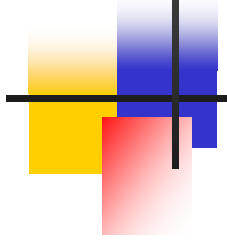
Goal of Factor Analysis

- “Reducing the dimensionality of the original space and giving an interpretation to the new space, spanned by a reduced number of new dimensions which are supposed to underlie the old ones”. (Rietveld & Van Hout, 1993)
- “To explain the variance in the observed variables in terms of underlying latent factors”. (Habing, 2003)



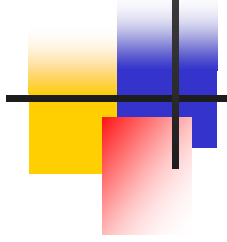
Application of FA

- A lot in research using questionnaires
- Testing of preschool children in order to examine relationship between various enabling skills and early reading performance. Factor analysis indicated three factors: phonemic awareness, naming, short-term memory. (Sprugevica & Høien, 2003)



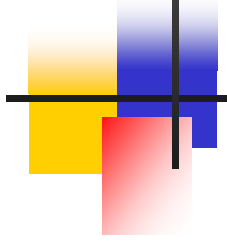
Steps in Factor Analysis

1. Reliable measurements? If so:
2. Correlation matrix
3. Factor Analysis or Principal Component Analysis?
4. Factor Analysis: how to estimate communalities?
Principal Component Analysis: communalities are 1
5. How many factors to be retained?
6. Factor rotation? Orthogonal or oblique?
7. Results: factor loadings and factor scores (use in subsequent analysis)
8. Interpretation by the researcher



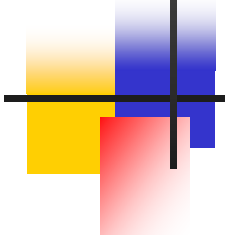
Reliable Measurements?

- Measurements should be at the interval level (at least);
- Variables should have roughly normal distributions;
- Sample size: the bigger the sample, the more stable the factor solution (but this also depends on the factor loadings and hence the communalities).



Correlation Matrix

- The variables have to be intercorrelated (substantial number >0.3 , Bartlett's test of sphericity, KMO-test)
- The variables should not correlate too highly (>0.9): that makes it difficult to determine the unique contribution of the variables to a factor (multicollinearity).



Correlation Matrix

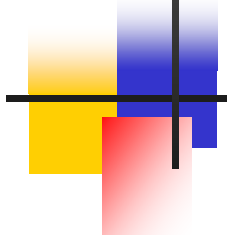
- Clusters of variables with high intercorrelations may be caused by underlying variables:

- $$\begin{pmatrix} 1.00 & & & & & & \\ \mathbf{0.77} & 1.00 & & & & & \\ \mathbf{0.66} & \mathbf{0.87} & 1.00 & & & & \\ 0.09 & 0.04 & 0.11 & 1.00 & & & \\ 0.12 & 0.06 & 0.10 & \mathbf{0.51} & 1.00 & & \\ 0.08 & 0.14 & 0.08 & \mathbf{0.61} & \mathbf{0.49} & 1.00 & \end{pmatrix}$$



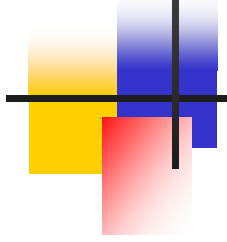
Factor Analysis or Principal Component Analysis?

- Difference lies in the amount of variance in the variables accounted for by the factors:
- PCA: explains the total variance in a variable by means of its components (communality = 1)
- FA: separates common variance from unique variance (communality to be estimated)



FA-analysis vs. PC-analysis

- Factor Analysis is more conservative;
- Difference decreases with increasing number of variables and magnitudes of factor loadings (correlation between original variable and factor);
- Estimating communality is difficult; PCA easier understandable.



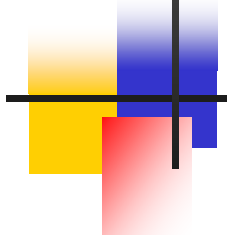
Principal Component Analysis

- From the correlation matrix the linear components are calculated, by means of determining the eigenvalues of the matrix;
- The number of positive eigenvalues represents the number of factors/components to be extracted
- The linear component is determined by a transformation matrix, which is determined by the eigenvectors.



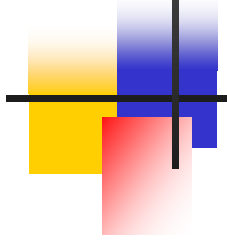
PCA -- criteria

- The first component should be chosen such that it accounts for the maximum part of the variance, second component maximum of the remaining part, and so on;
- The scores on the components are not correlated.



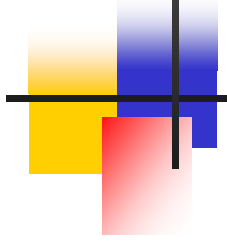
Factor Analysis

- Communality estimation (squared multiple correlations)
- Using these communalities FA is done and new communalities can be determined
- Estimated communalities on the diagonal of the reproduced R-matrix (goodness of fit, residuals)
- Important to know that the factor loadings do not represent the total variance of the variable.



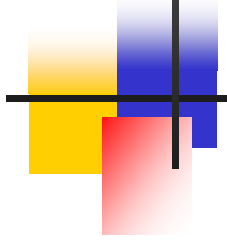
Factor extraction

- Depends on eigenvalues. Many rules of thumb:
 - Retain only those factors with an eigenvalue larger than 1 (Guttman-Kaiser)
 - Make a scree-plot; keep all factors before the break point
 - Keep the factors that account for about 70% / 80% of the variance altogether



Factor loadings

- Important for the interpretation of the factors
- When is a factor loading significant?
Stevens' table of critical values best way to determine that.
- The higher the loadings, the better interpretation possible. Therefore: rotation



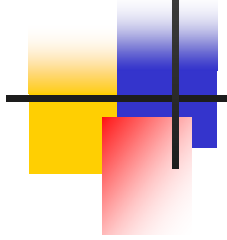
Rotation

- Orthogonal and oblique
- Difference is correlation between factors
- Which rotation is best? Do both types of rotation and look at the correlation between the factors in oblique rotation: negligible correlation or not?
- Oblique rotation only if you have good reasons to suppose a relation between underlying factors.
- Angle of rotation represented by transformation matrix.



Oblique rotation

- Results in:
 - Matrix with correlations between the factors
 - Matrix with correlation coefficients between variables and factors: structure matrix
 - Matrix with regression coefficients of the variables on the factors: pattern matrix.
Important one for interpretation of factors.



Results: loadings and scores

- Loadings: interpretation of factor
- Scores:
 1. Looking for cluster of subjects scoring similarly
 2. Use of factor scores in cases of multicollinearity in multiple regression
- Scores can be used in further research; loadings are for interpreting the content of the factor.