

Factor Analysis

Good cite:

Larry Hatcher and Edward Stepanski,
“Principal Component Analysis” (Ch. 14) in
A Step-by-Step Approach to Using the SAS
System for Univariate and Multivariate
Statistics

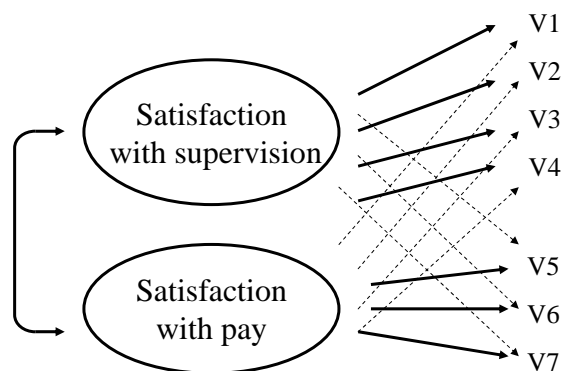
Factor Analysis: Purposes

- 1) Revealing patterns of inter-relationships
- 2) Detecting clusters of strongly correlated variables
- 3) Reducing large number of variables to smaller number of factors***
(Way to handle multicollinearity)
***We'll focus on (3)—scale construction

Factor Analysis

- Incorporate different variables to measure an underlying dimension
- Assumes an underlying causal structure

Factor Analysis



e.g., V1 = "My supervisor treats me with consideration"

V5 = "My pay is fair"

Source: Hatcher & Stepanski, p. 457

Factor Analysis: Basic principles

- Start with large number of variables
- Find minimum number of underlying factors (principal components) that together account for the pattern of intercorrelations among observed variables
- *Note:* variables must be coded in similar ways (high = high)

Factor Analysis: properties of principal components

- Component = linear combination
- Orthogonal to each other
- 1st principal component = largest variance of any linear function of original variables
- 2nd principal component = second largest of remaining variance

Eigenvalues

“Characteristic root of the matrix”

- Used to calculate variance of the *factors*
- Largest eigenvalue represents amount of variance in data explained by 1st factor
- Each factor has one eigenvalue
- Keep factors with eigenvalues > 1.0
- Wide variation in eigenvalues = severe multicollinearity

Eigenvectors

- AKA “*factor loadings*” (or, factor structure)
- Coefficients for linear transformation of the standardized variables
- Used to interpret structure of variables
- Used to compute correlations between original variables and new factor variables
- SAS: coefficients in factor pattern matrix and rotated factor pattern matrix

Eigenvectors (cont.)

Each observed variable viewed as determined by underlying factors:

$$\text{e.g., Var1} = \text{weight*F1} + \text{weight*F2} \\ + \text{weight*F3} \dots + \text{weight*\mu}$$

Interpretation: largest weight = most important determinant

Eigenvectors (example)

$$\text{Var1} = .889\text{F1} + .078\text{F2} + .032\text{F3} + \text{d}\mu$$

- (weight)² = proportion of variance in variable accounted for by factor
- (.889)² = 79% of variance in Var1 explained by F1

Eigenvectors (example)

h^2 (communality of variable):

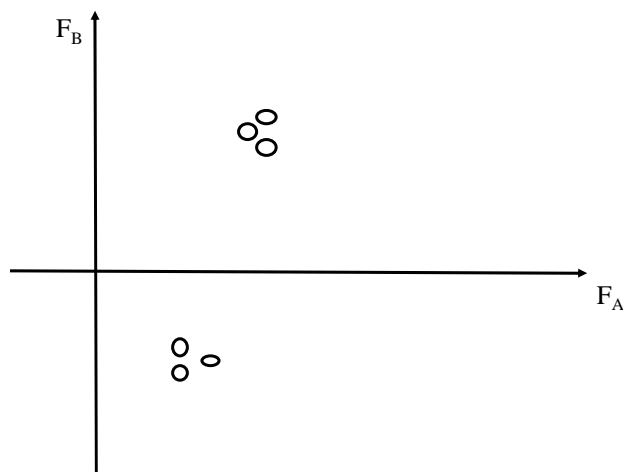
$$h^2 = (.889)^2 + (.078)^2 + (.032)^2$$

$h^2 = 79.7\%$ of variance in Var1
explained by F1 to F3

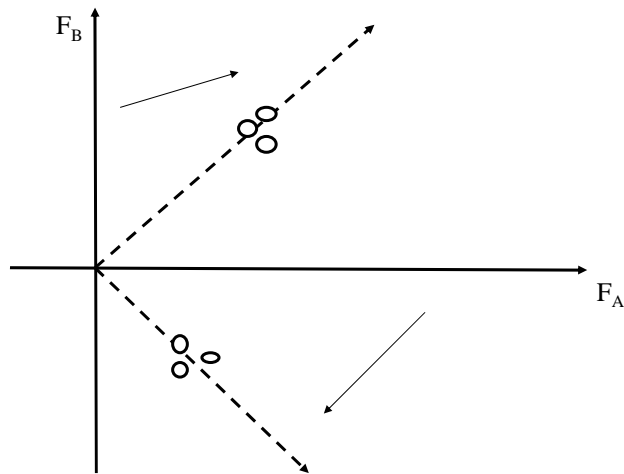
$1-h^2 =$ proportion left unexplained by F1 to F3

$$\sqrt{1-h^2} = d \quad (\text{weight for residual term})$$

Rotation



Rotation: (Varimax: maximizes contrast)



Evaluate

- Factors are orthogonal, whether rotated or not
- Rotation makes clustering more obvious
- We'll use "Varimax Rotation": maximizes contrast between factors
- Evaluate loadings: criteria = .4 or .5

Evaluate

Principal component analysis \neq factor analysis:

- Factor analysis assumes an underlying causal structure
- Principal component analysis makes no assumption of underlying causal model