

Factor Analysis

Factor Analysis

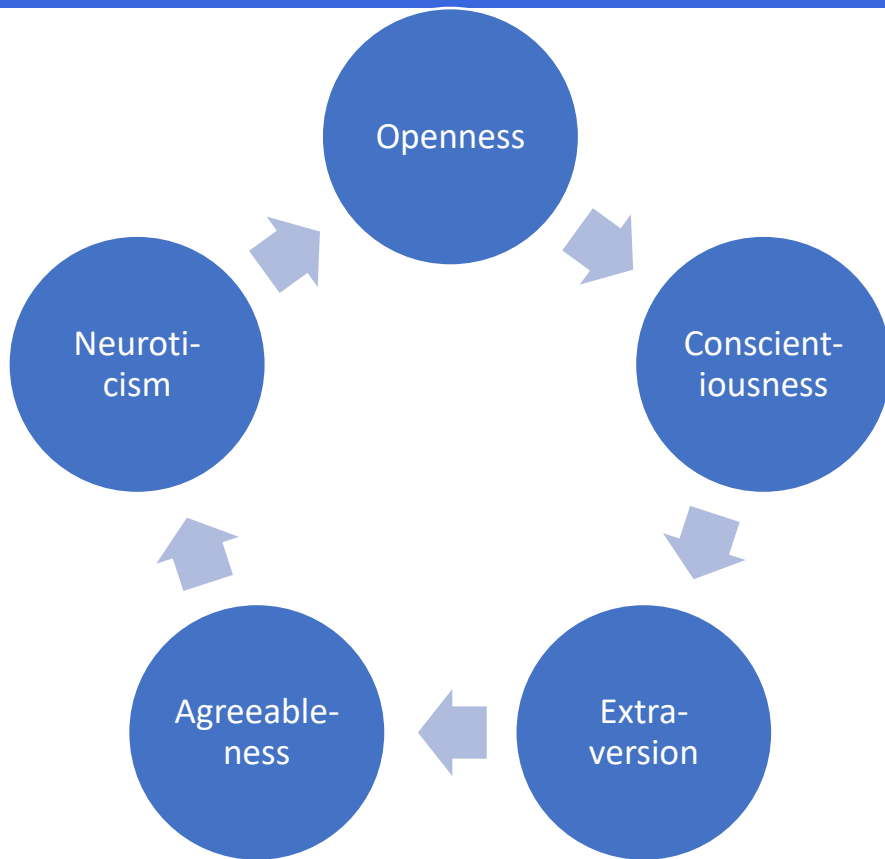
Introduction

- Problem
 - Data with many variables – curse of dimensionality
 - Finding patterns between observations hard
 - Visualisation not possible
- Main objective
 - Identify and measure factors that are not directly measurable
- FA
 - Unsupervised learning technique
 - Very similar to PCA
 - Based on assumption of existing latent (hidden, not observable) variables
 - As many latent factors as variables in data

Factor Analysis

Latent Factors

Example: **Big Five**
personality traits



Factor Analysis

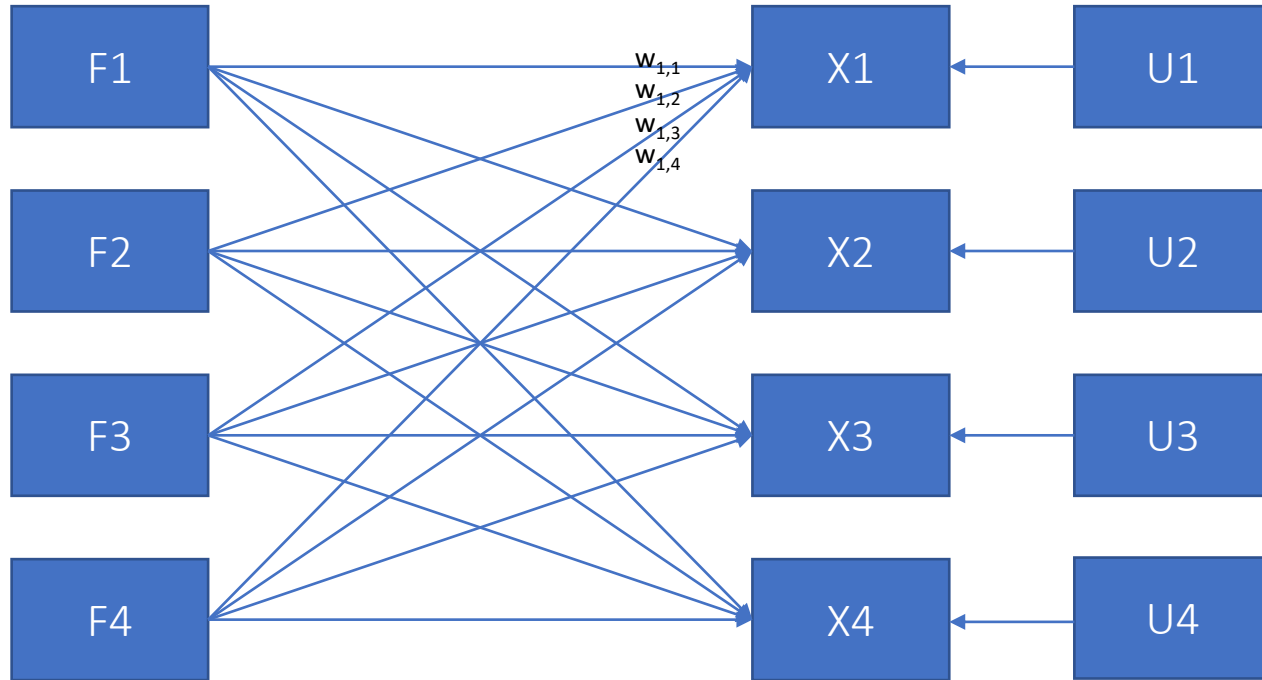
Principle



- Required: (observed) correlated variables
- FA transforms correlated variables into unobservable, latent factors

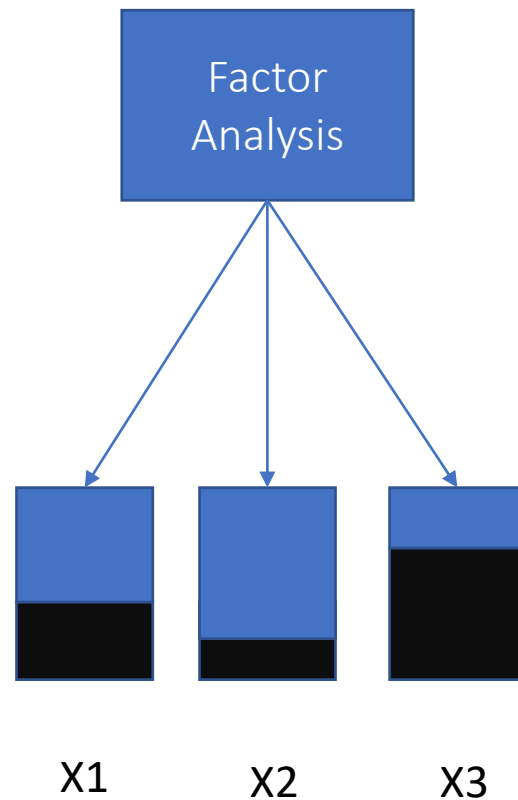
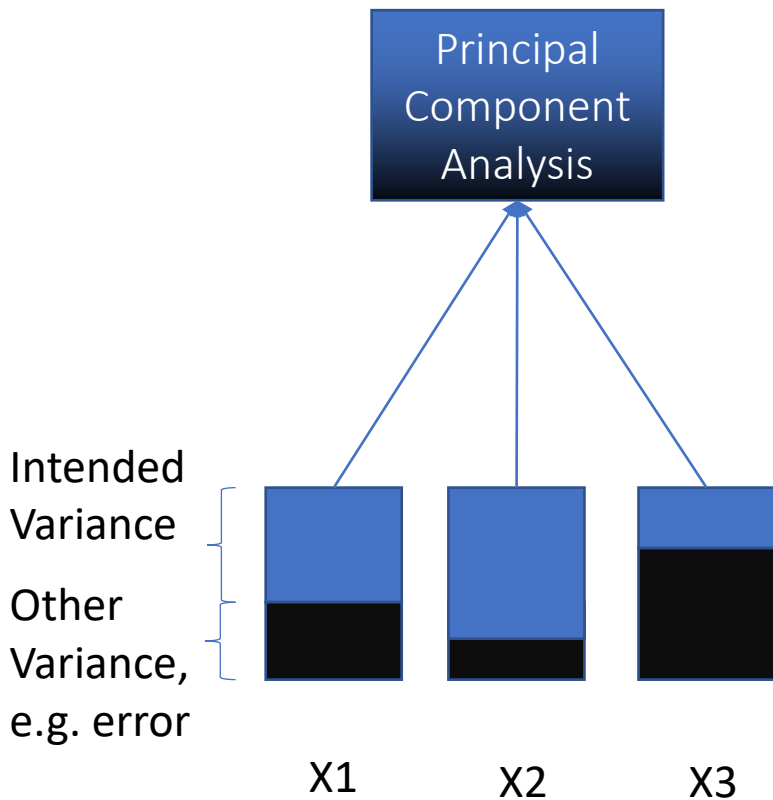
Factor Analysis

Principle



Factor Analysis

PCA and FA



Factor Analysis

KMO Criterion

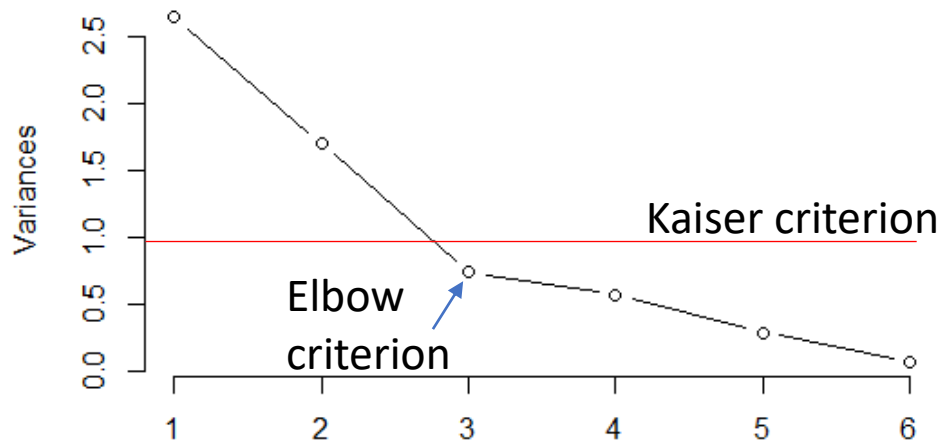
Kaiser-Meyer-Olkin

- Measures how good variables can be grouped to factors
- Range 0 to 1
- Usually >0.5 should be used to proceed
- If <0.5 results of FA possibly not useful

Factor Analysis

Number of Components

- Two or three for visualisation
- Scree plot
 - find „elbow“
- Kaiser criterion
 - Eigenvalues > 1



Factor Analysis

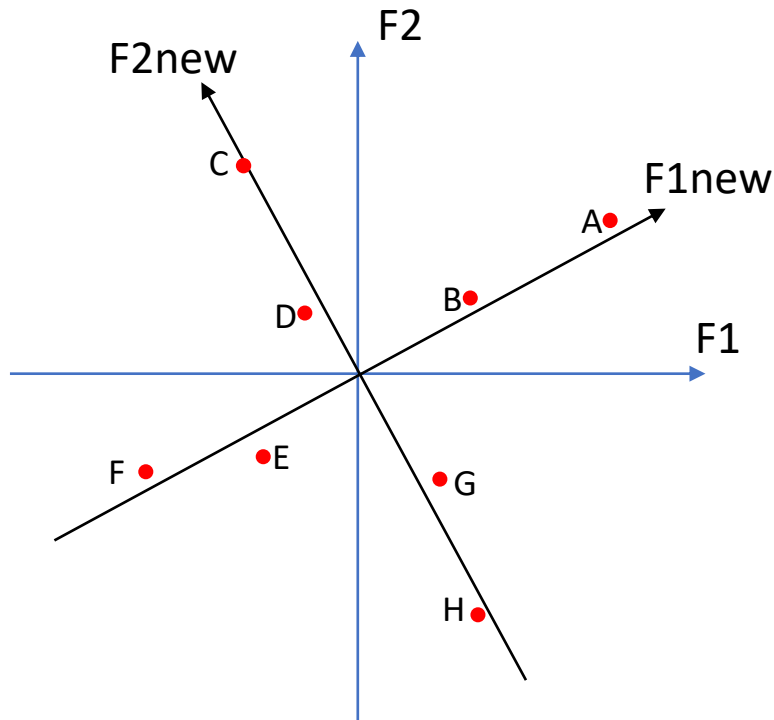
Rotation

Base point

- All observations load partially on F1 and F2
- Hard to interpret

After rotation

- Observations load on either F1 or F2
- Does not affect explained variance of components
- Most common rotation: varimax
- Helps to interpret the model



Factor Analysis

Advantages / Disadvantages



- can test a theory on latent factors
- Applicable for feature selection
- Well established



- Requires a theory that you want to test
- Subjective interpretation of factors
- Factors only represent what was measured