SEM training using R

Session 1: Exploratory Factor Analysis

04 August 2021

Factor analysis process

Stage 1: Objectives of factor analysis

Stage 2: Designing an Exploratory factor analysis

Stage 3: Assumptions in Exploratory factor analysis

Stage 4: Deriving factors and assessing overall fit

Stage 5: Interpreting the factors

Stage 1 : Objectives of factor analysis

Types of factor analysis

Exploratory factor analysis

- Use when you do not have a well-developed theory
- Estimate all possible variable/ factor relationships
- Looking for patterns in the data

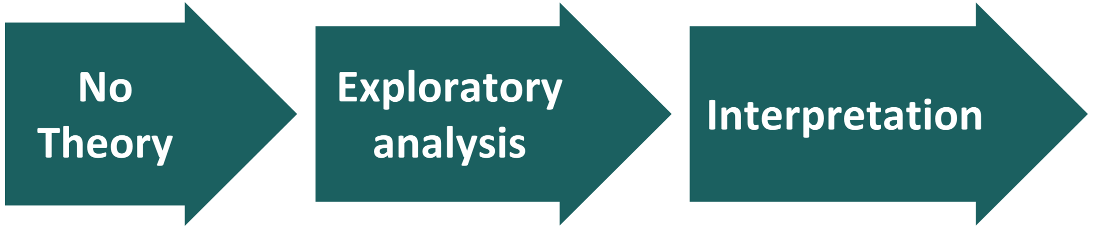
Confirmatory factor analysis

- Testing a theory that you know in advance
- Only specified variables/factor relationships

Types of factor analysis

Exploratory factor analysis

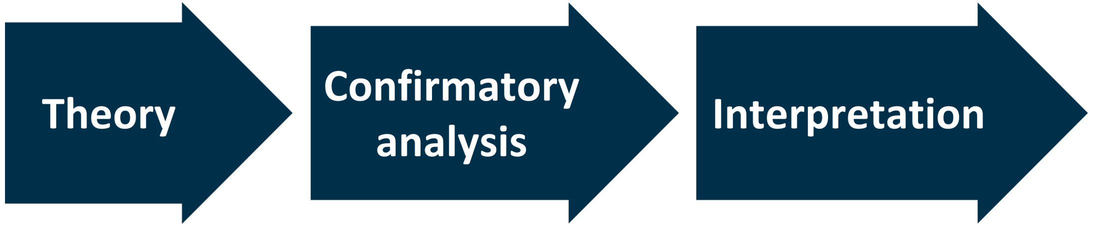
- Difficult to interpret without a theory.
- factor loadings: meanings can sometimes be inferred from patterns.



Types of factor analysis

Confirmatory factor analysis

- Model fit: how well the hypothesized model fits the data.
- Factor loadings: how well items measure their corresponding constructs.



Stage 2: Designing an EFA

Variable selection and measurement issues

1. What types of variables can be used in factor analysis?
    - *Primary requirement: a correlation value can be calculated among all variables.*
    - *e.g., metric variables, scale items, dummy variables to represent nonmetric variables.*
2. How many variables should be included?
    - *Five or more per factor for scale development.*
    - *Three or more per factor for factor measurement (based on how degrees of freedom is computed).*

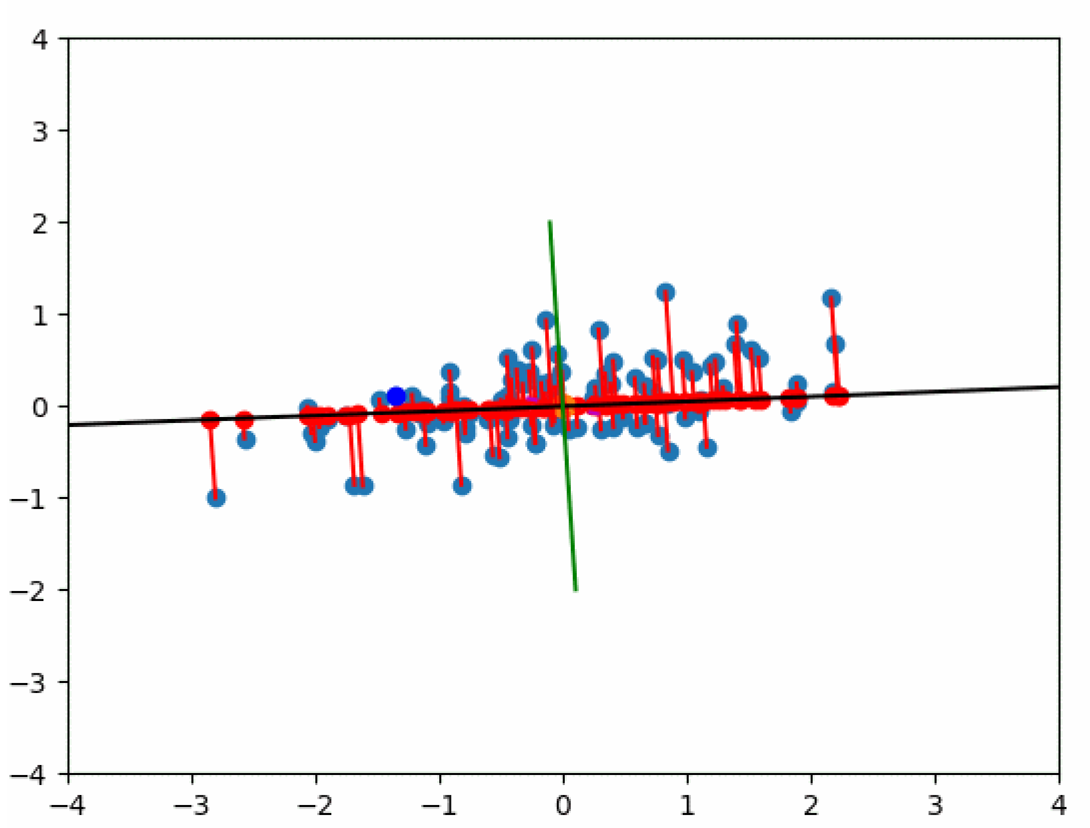
Sample size

Some recommended guidelines:

1. absolute size of the dataset
    - *should not fewer than 50 observation*
    - *preferably 100 and larger*
    - *200 and larger as the number of variables and expected factors incerases*
2. ratio of cases to variables
    - *observation is 5x as the number of variables*
    - *sample size is 10:1 ratio*
    - *some proposes 20 cases per variables*

Stage 3: Assumptions in EFA

How many dimensions do you have?



Dataset

(Info here!)

x6	x7	x8	x9	x10	x11	x12
<dbl+lbl>	<dbl+lbl>	<dbl+lbl>	<dbl+lbl>	<dbl+lbl>	<dbl+lbl>	<dbl+lbl>
8.5	3.9	2.5	5.9	4.8	4.9	6.0
8.2	2.7	5.1	7.2	3.4	7.9	3.1
9.2	3.4	5.6	5.6	5.4	7.4	5.8
6.4	3.3	7.0	3.7	4.7	4.7	4.5
9.0	3.4	5.2	4.6	2.2	6.0	4.5
6.5	2.8	3.1	4.1	4.0	4.3	3.7
6.9	3.7	5.0	2.6	2.1	2.3	5.4
6.2	3.3	3.9	4.8	4.6	3.6	5.1
5.8	3.6	5.1	6.7	3.7	5.9	5.8
6.4	4.5	5.1	6.1	4.7	5.7	5.7

1-10 of 100 rows | 1-7 of 11 col... Previous **1** 2 3 4 5 6 ... 10 Next

Source: J.F. Hair (2019): Multivariate data analysis.

Overall measures of intercorrelation

1. Bartlett Test
2. Measure of Sampling Adequacy

Overall measures of intercorrelation

1. Bartlett Test

- Examines the entire correlation matrix
- Test the hypothesis that correlation matrix is an identity matrix.
- A significant reulst signifies data are appropriate for FA

```
BARTLETT(data, N= nrow(data))
```

```
v The Bartlett's test of sphericity was significa
  These data are probably suitable for factor ana

  <U+0001D712>²(55) = 619.27, p < .001
```

Overall measures of intercorrelation

2. Kaiser-Meyen-Olkin (KMO Test)

- Measure of sampling adequacy
- Indicate the proportion of variance explained by the underlying factor.
- Guidelines:
    - ≥ 0.90 - marvelous
    - ≥ 0.80 - meritorious
    - ≥ 0.70 - middling
    - ≥ 0.60 - mediocre
    - ≥ 0.50 - miserable
    - < 0.50 - unacceptable

Overall measures of intercorrelation

2. Kaiser-Meyen-Olkin (KMO Test)

```
library(psych)
KMO(data)
```

```
-- Kaiser-Meyer-Olkin criterion (KMO) ------------------------------------------------

! The overall KMO value for your data is mediocre.
  These data are probably suitable for factor analysis.

  Overall: 0.653

  For each variable:
   x6    x7    x8    x9   x10   x11   x12   x13   x14   x16   x18
0.509 0.626 0.519 0.787 0.779 0.622 0.622 0.753 0.511 0.760 0.666
```

Overall measures of intercorrelation

2. Kaiser-Meyen-Olkin (KMO Test)

- When overall MSA is less than 0.50
  - Identify variables with lowest MSA subject for deletion.
  - Recalculate MSA
  - Repeat unitl overall MSA is 0.50 and above
- Deletion of variables with MSA under 0.50 means variable's correlation with other variables are poorly representing the extracted factor.

Overall measures of intercorrelation

2. Kaiser-Meyen-Olkin (KMO Test)

```
# Deselecting X15
data_deselect <- data %>%
  select(-x15, -x17)

KMO(data_deselect)
```

Let's practice!

Selecting factor extraction method

Partitioning the variance of a variable

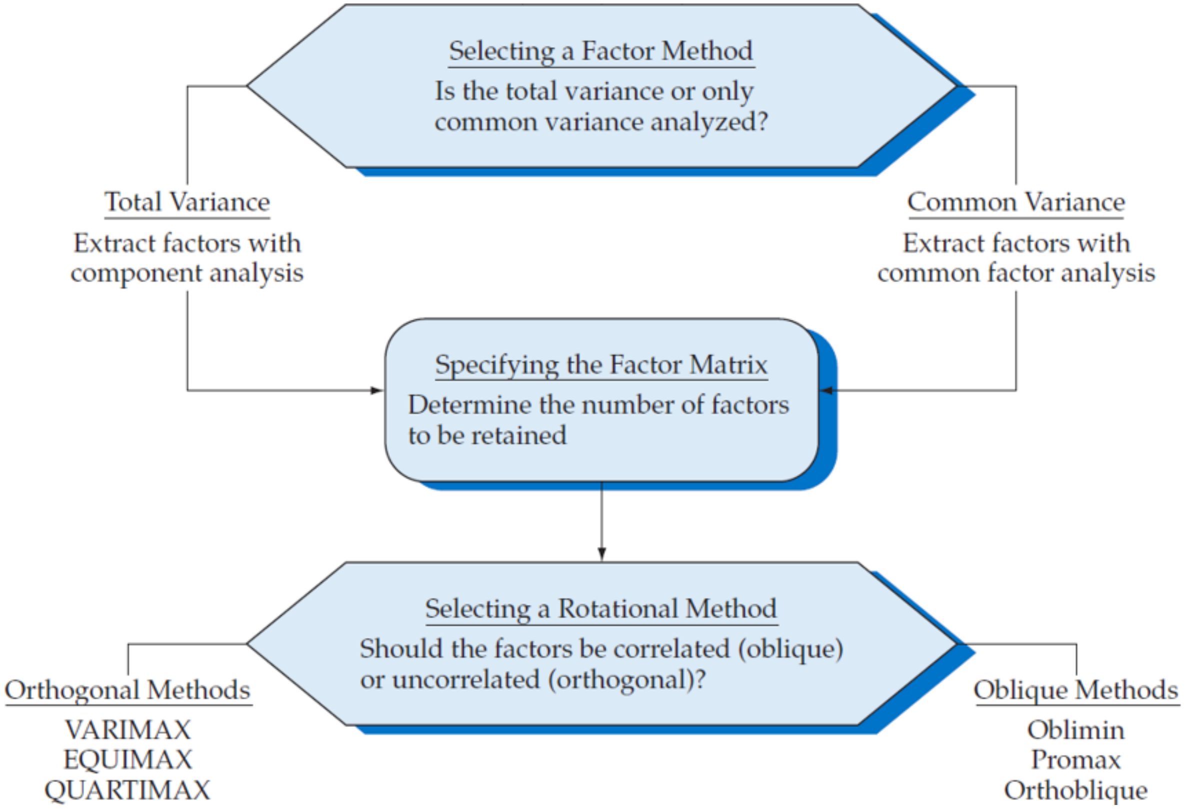
Unique variance

- Variance associated with only a specific variable.
- Not represented in the correlations among variables.
- *Specific variance*
  - associated uniquely with a single variable.
- *Error variance*
  - May be due to unreliability of data gathering process, measurement error, or a random component in the measured phenomenom.

Common variance

- Shared variance with all other variables.
- High common variance are more amenable for factor analysis.
- Derived factors represents the shared or common variance among the variables.

Partitioning the variance of a variable



Source: JF Hair et al. (2019) Multivariate data analysis.

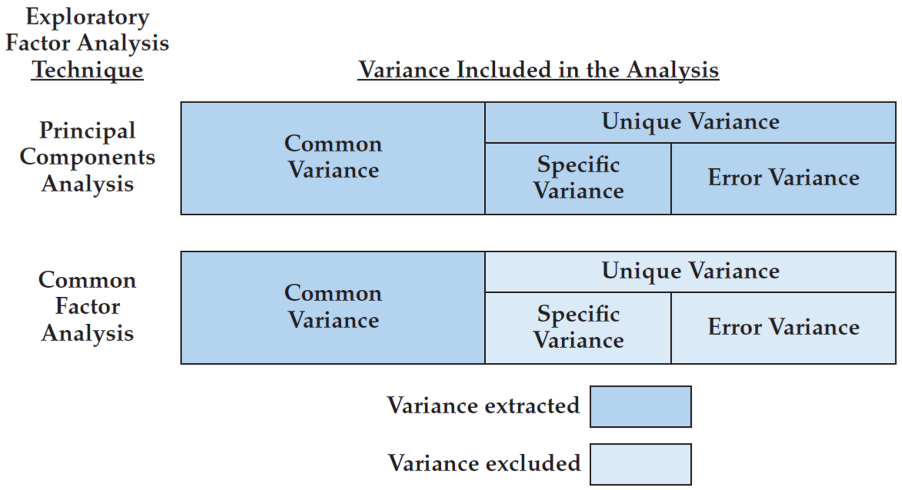
PCA vs Common factor analysis

Principal component analysis (PCA)

- Considers the total variance
- data reduction is a primary concern

Common factor analysis

- Considers only the common variance or shared variance
- Primary objective si to identify the laten dimensions or constructs



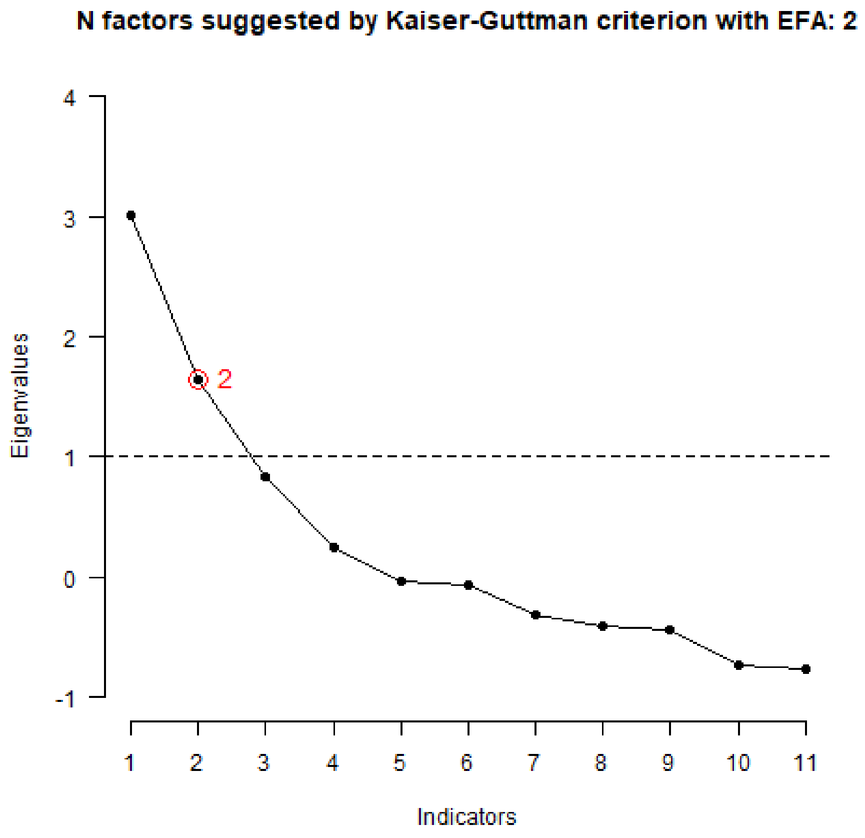
Source: JF Hair et al. (2019) Multivariate data analysis.

Exploring possible factor

1. Kaiser-Guttman Criterion

- Only consider factors whose eigenvalues is greater than 1.
- Rationale is that factor should account for the variance of at least a single variable if it is to be retained for interpretation.

```r
library(EFAtools)
KGC(Data, eigen_type = "EFA")
```



Exploring possible factors

2. Scree test

- Identify the optimum number of factors that can be extracted before the amount of unique variance begins to dominate the common variance.
- Inflection point or the "elbow"

```
library(psych)
scree(data)
```

Exploring possible factors

3. Parallel Test

- Generates a large number of simulated dataset.
- Each simulated dataset is factor analyzed.
    - Results is the average eigenvalues across simulation.
    - Values are then compared to the eigenvalues extracted from the original dataset.
    - All factors with eigenvalues above those average eigenvalues are retained.

```r
library(psych)
fa.parallel(data, fa = "fa")
```

Let's practice!

Three process of factor intepretation

1. Factor extraction

2. Factor rotation

3. Factor interpretation and re-specification

Factor extraction

Loadings

- Correlation of each variable and the factor.
- Indicate the degree of correspondence between variable and factor.
- Higher loadings making the variable representative of the factor.

```
fa_unrotated <- fa(r = data, nfactors = 4,rotate
print(fa_unrotated$loadings)
```

```
Loadings:
    MR1    MR2    MR3    MR4   
x6   0.201 -0.408         0.463
x7   0.290  0.656  0.267  0.210
x8   0.278 -0.382  0.744 -0.169
x9   0.862        -0.255 -0.184
x10  0.287  0.456         0.127
x11  0.689 -0.454 -0.141  0.316
x12  0.398  0.807  0.348  0.255
x13 -0.231  0.553        -0.287
x14  0.378 -0.322  0.730 -0.151
x16  0.747        -0.176 -0.181
x18  0.895        -0.304 -0.198

                 MR1   MR2   MR3   MR4
SS loadings    3.215 2.226 1.500 0.679
Proportion Var 0.292 0.202 0.136 0.062
Cumulative Var 0.292 0.495 0.631 0.693
```

Factor extraction

Loadings

- $\leq \pm 0.10 \approx$ zero
- ± 0.10 to ± 0.40 meet the minimal level
- $\geq \pm 0.50$ practically significant
- $\geq \pm 0.70 \approx$ well-defined structure

SS loadings

- Eigenvalues - column sum of squared factor loadings.
- Relative importance of each factor in accounting for the variance associated with the set of variables.

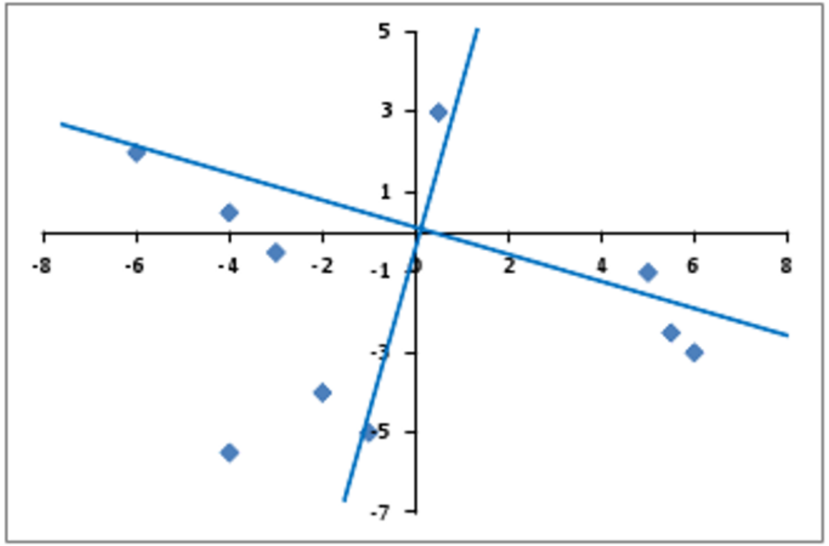
```r
fa_unrotated <- fa(r = data, nfactors = 4,rotate
print(fa_unrotated$loadings)
```

```

Loadings:
    MR1    MR2    MR3    MR4   
x6   0.201 -0.408         0.463
x7   0.290  0.656  0.267  0.210
x8   0.278 -0.382  0.744 -0.169
x9   0.862        -0.255 -0.184
x10  0.287  0.456         0.127
x11  0.689 -0.454 -0.141  0.316
x12  0.398  0.807  0.348  0.255
x13 -0.231  0.553        -0.287
x14  0.378 -0.322  0.730 -0.151
x16  0.747        -0.176 -0.181
x18  0.895        -0.304 -0.198

                 MR1   MR2   MR3   MR4
SS loadings    3.215 2.226 1.500 0.679
Proportion Var 0.292 0.202 0.136 0.062
Cumulative Var 0.292 0.495 0.631 0.693
```

Factor rotation

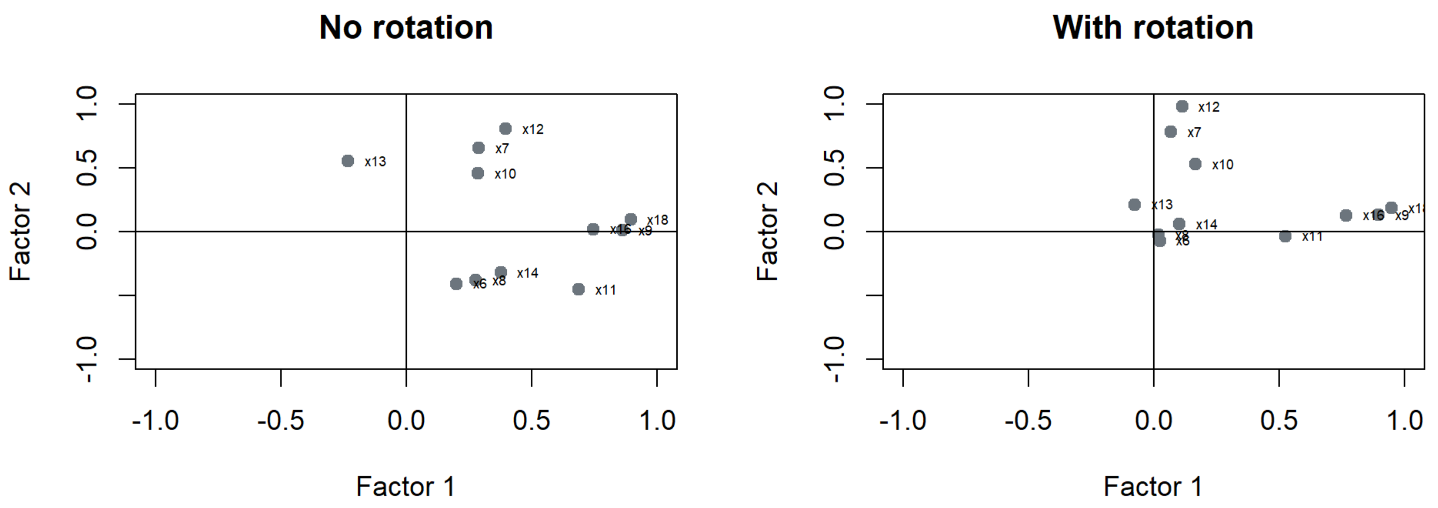
Why do factor rotation?

- To simplify the complexity of factor loadings.
- Distribute the loadings more clearly into the factors.
- Facilitate interpretation.



Factor rotation

```
par(mfrow = c(1, 2))
plot(fa_unrotated$loadings[,
     xlab = "Factor 1", ylab
     ylim = c(-1, 1), xlim =
     main = "No rotation",
     pch = 19, col = "#6c757
     abline(h=0, v=0)
     text(fa_unrotated$loadi
          labels = rownames(
          pos = 4, cex = 0.5

plot(fa_rotated$loadings[,1]
     xlab = "Factor 1", ylab
     ylim = c(-1, 1), xlim =
     main = "With rotation",
     pch = 19, col = "#6c757
     abline(h=0, v=0)
     text(fa_rotated$loading
          labels = rownames(
          pos = 4, cex = 0.5
```

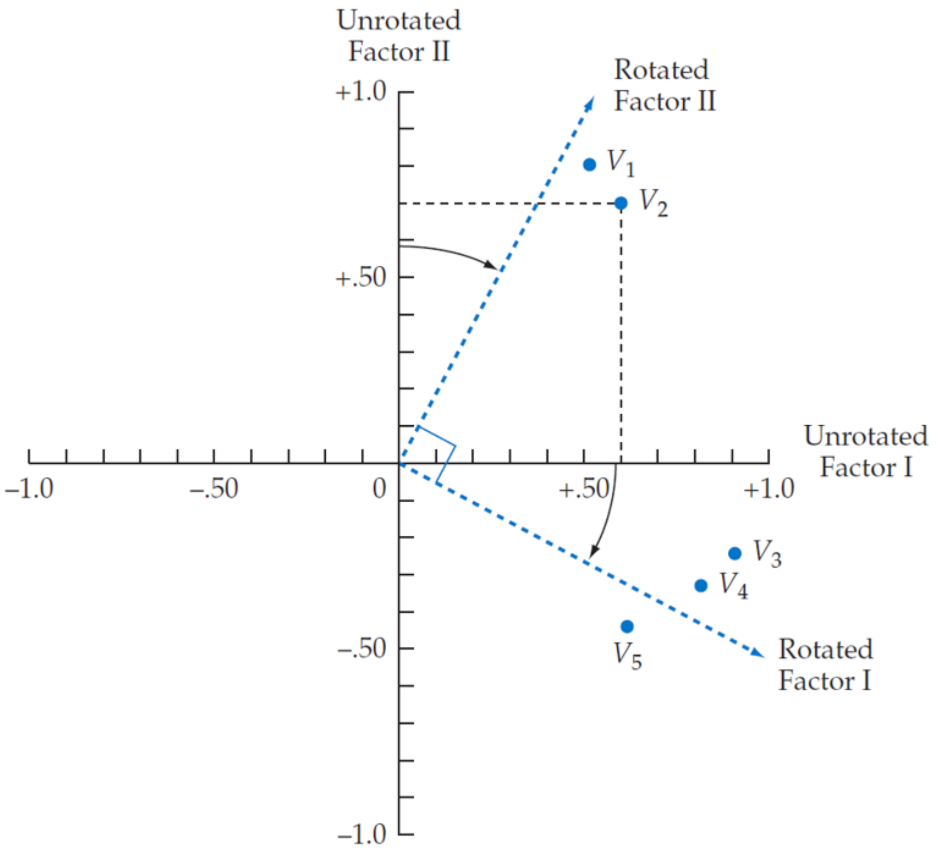


Factor rotation

Factor rotation

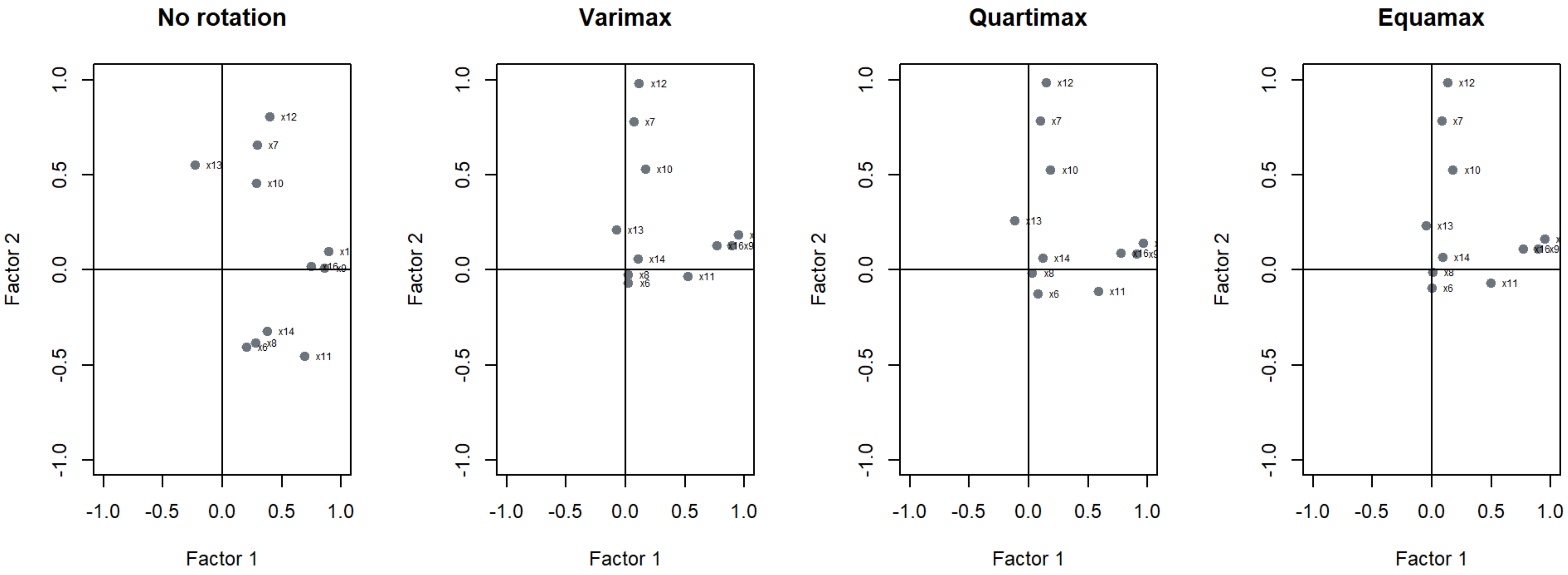
Orthogonal rotation

- axes are maintained at 90 degrees
- orthogonal rotation methods
  - Varimax - most commonly used
  - Quartimax
  - Equimax
- Check-out some of these references
  - IBM
  - Factor analysis



Factor rotation

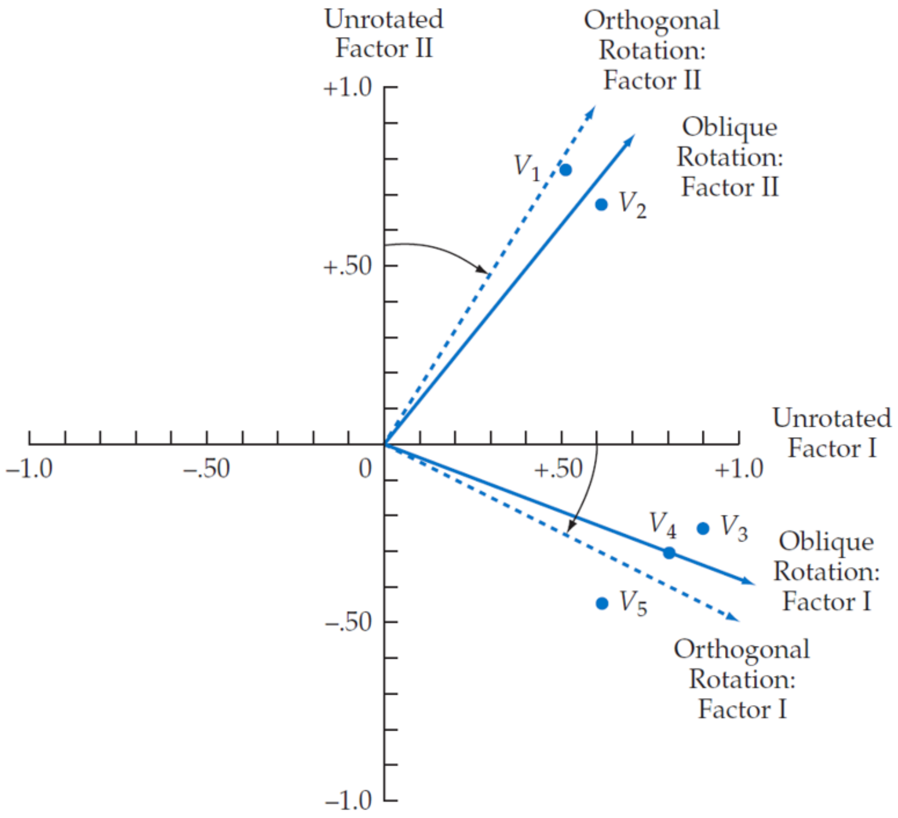
Orthogonal rotation



Factor rotation

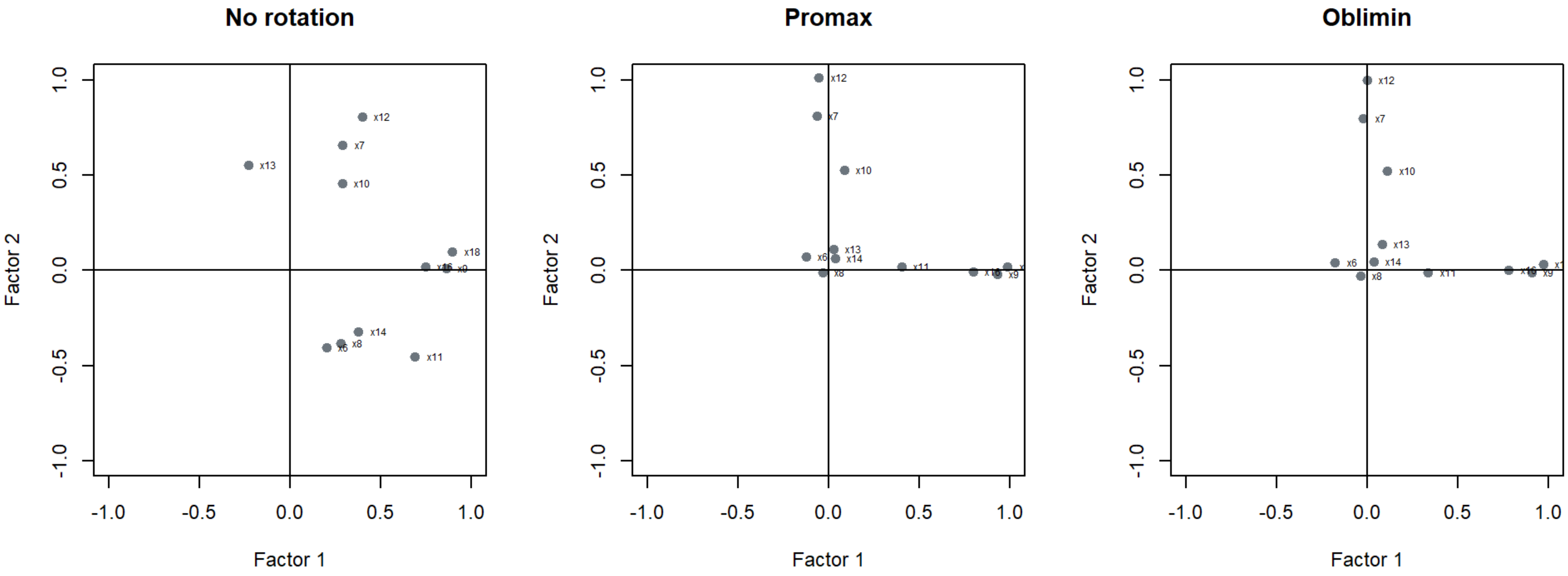
Oblique rotation rotation

- allow correlated factors
- suited to the goal of theoretically meaningful constracts
- oblique rotation methods
  - Promax
  - Oblimin
- Note: no specific rules in selecting a between rotation method.



Factor rotation

Oblique rotation



Let's practice!