Factor Analysis

Jerome Reboulleau[1]

[1] HEC, Lausanne

Quantitative Method for Management

Outline

1. Introduction to Factor Analysis
2. Principal Component Analysis
3. Factor Analysis
4. Factor Rotation

PART I: INTRODUCTION TO FACTOR ANALYSIS

1. **Introduction to Factor Analysis**
   - Overview: Factors, Factor Loadings
   - Two Methods: Factor analysis vs Component Analysis
2. Principal Component Analysis
3. Factor Analysis
4. Factor Rotation

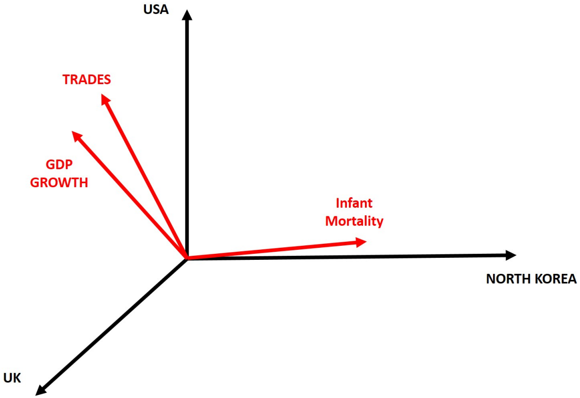
Factor Analysis

Factor Analysis

- Factor analysis is defined generally as a method for simplifying complex sets of data **summarizing data** with several variables.
- Generally speaking, the method is associated to correlation.
- Mutiple variables can have common patterns. Factor analysis helps understanding the role of each factor and identifying common variables effects and separated variables effects.
- The main weakness is that there is often an infinite number of solutions when we are willing to **summarize data**.
- We will go through 2 methods: **Principal Component Analysis (PCA)** and **Factor Analysis**.

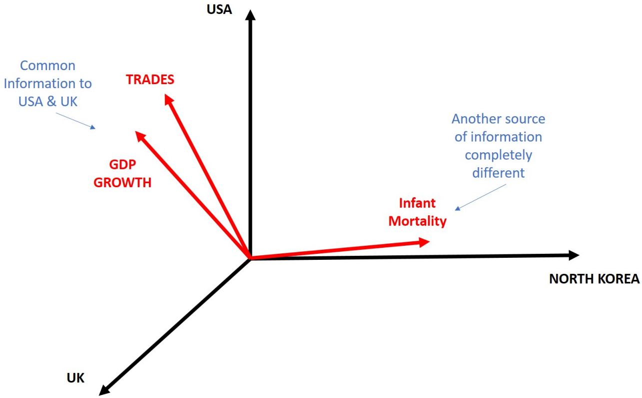
Factors and Factor Loadings

- We define a **Factor** as a dimension which is a condensed statement of the relationships between a set of variables.



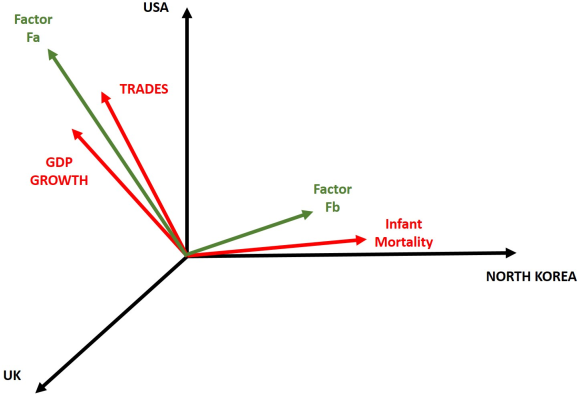
Factors and Factor Loadings

- We define a **Factor** as a dimension which is a condensed statement of the relationships between a set of variables.



Factors and Factor Loadings

- These factors can be "apparent" or "hidden" through the combination of underlying information.



Factors and Factor Loadings

- Remarks
  1. These factors will be related to the variability of the variables **Trades, GDP Growth, Mortality...**, so related to the variance-covariance matrix.
  2. Because we are summarizing the variability, we will have a loss in capturing the variance.
  3. Compared to our previous Chapters, we are not interested in a response variable y.

Factors and Factor Loadings

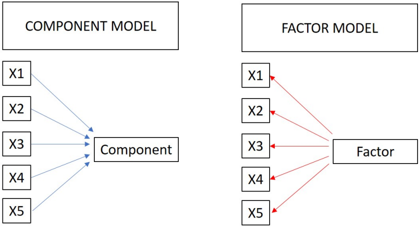
- We define a **Factor** as a dimension which is a condensed statement of the relationships between a set of variables.
- A Factor is defined through the **Factor Loadings**.
- **Factor Loadings** are the correlations of the variables with a factor.

Factor vs Component

- We will go through 2 methods to work on Factor Analysis.
- The first one is Principal Component Analysis.
- The second one is called Factor Analysis.
- Before covering each method, we briefly express the key difference.

Factor vs Component

- The difference between the 2 methods can be summarized is the following graph



- Principal Component analysis: components are linear sums of variables.
- Factor analysis: the variables are expressed as a linear combinations of the factors.

PART II:
PRINCIPAL COMPONENT ANALYSIS

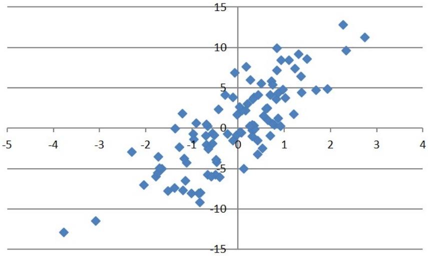
1 Introduction to Factor Analysis

2 Principal Component Analysis
- Principal Component Analysis: Method introduction
- Principal Component Analysis: Method Geometrical interpretation
- Principal Component Analysis: Method Algebric interpretation
- Principal Component Analysis: Method interest
- Principal Component Analysis: Example 1
- Principal Component Analysis: Example 2
- Principal Component Analysis: Maps
- How many components to choose ?

3 Factor Analysis

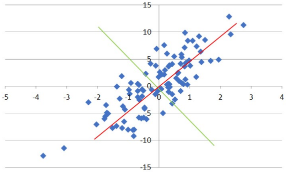
Principal Component Analysis: Method Explanation

- Suppose we have the following 2 dimensions situation



Principal Component Analysis: Method Explanation

- We are trying to find the green and red lines, that are going to be our Principal Components



- Geometrically, it is equivalent to rotating the axis x_1, x_2.
- Mathematically it is equivalent to a coordinate change from a dimension x_1, x_2 to Prin Comp$_1$,Prin Comp$_2$.
- We are thus going to "modify" our Variance-Covariance Matrix into a new coordinates Prin Comp$_1$,Prin Comp$_2$.
- Prin Comp$_1$ will capture most of the variation while Prin Comp$_2$ will get the rest of it, not related to Prin Comp$_1$ (orthogonality).

Method Presentation

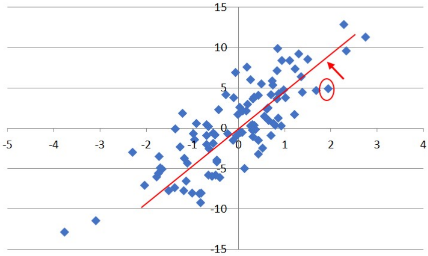
- X represents our p variables with N data for each x_j. To simplify we suppose all x_j have mean 0.
- The sample covariance matrix of our data X is $\frac{1}{N} X^T X$.
- The Eigen decomposition of $X^T X$ is

$$X^T X = VD^2 V^T$$

- The matrix V is composed by p vectors v_j. These vectors v_j are called principal components directions of X.
- V is organized such that his 1st vector v_1 is the one with the largest eigenvalue d_1^2 of $X^T X$.
- The vector $z_1 = X v_1$ is called the 1st principal component of X.
- z_1 has the largest sample variance amongst all linear combination of the columns of X: $Var(z_1) = \frac{1}{N} d_1^2$
- The 2nd principal component z_2 is the eigenvector corresponding to the 2nd largest eigenvalue d_2^2.
- And so on ...

Principal Component Analysis: Method Explanation

- If we choose to use only the 1st Principal Component, we can summarize the date in a reasonable manner.
- The circle data can be represented by its neighbor on the principal component. It can be enough to say where it is along the principal component.



- By doing so, we have simplified from 2 coordinates to only 1 without loosing too much information.

Principal Component Analysis: Method Explanation

- We have simplified from 2 coordinates to only 1 without loosing too much information.
- When we want to increase information, we add principal components: Photo source Penn State online course.

Principal Component Analysis: Example 1

- To illustrate the method, we work with the simple following dataset.
- A 12-years old girl made five ratings on a 9-point sementic differential scale for each of seven of her acquaintances. These ratings were based on the five adjectives
- The six variables are
    1. Kind
    2. Intelligent
    3. Happy
    4. Likeable
    5. Just

Principal Component Analysis: Step 1 Data analysis

- We start by analyzing the covariance matrix

```
> cor(mypcaexample2[,-1])
                 Kind Intelligent       Happy  Likeable      Just
Kind        1.0000000  0.29553589  0.88057207 0.9954293 0.5445672
Intelligent 0.2955359  1.00000000 -0.02174427 0.3261640 0.8372882
Happy       0.8805721 -0.02174427  1.00000000 0.8666667 0.1303372
Likeable    0.9954293  0.32616404  0.86666667 1.0000000 0.5440163
Just        0.5445672  0.83728820  0.13033724 0.5440163 1.0000000
```

- Thanks to the correlation matrix, we see that the variable 1, 3 and 4 are related one to each other.
- Variables 2 and 5 constitutes the other group of data.

Principal Component Analysis: Step 2 Principal Component Result

- Applying the method in R

```
> summary(princomp(mypcaexample2[,-1]))
Importance of components:
                          Comp.1    Comp.2    Comp.3      Comp.4       Comp.5
Standard deviation     5.793697 2.7433564 0.7527716 0.508692585 3.827548e-08
Proportion of Variance 0.800769 0.1795395 0.0135183 0.006173145 3.494918e-17
Cumulative Proportion  0.800769 0.9803086 0.9938269 1.000000000 1.000000e+00
```

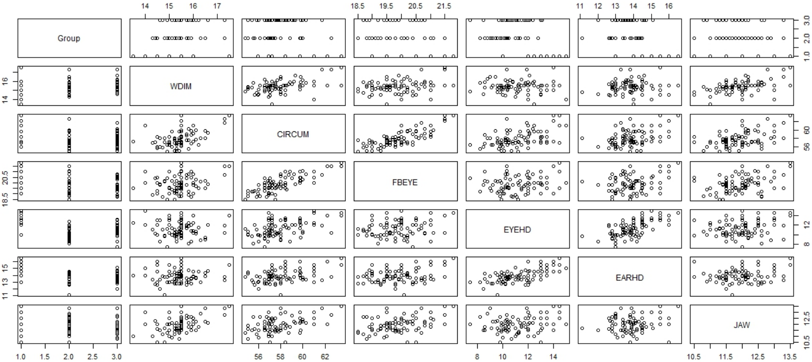
- We see that the first component explain 80% of the overall variation.
- The second component explain a further 18%.
- The rest of the components (Components 3 to 5) explain less than 2% overall.

Principal Component Analysis: Example 2

- Data were collected as part of a preliminary study of a possible link between football helmet design and neck injuries.
- Six measurements were made on each subjects. 3 groups of 30 subjects each were considered (Group 1: high school football players, Group 2: College players, Group 3: Non footfall players).
- The six variables are
  1. WDIM = head width at widest dimension,
  2. CIRCUM = head circumference,
  3. FBEYE = front-to-back measurement at eye level,
  4. EYEHD = eye-to-top-of-head measurement,
  5. EARHD = ear-to-top-of-head measurement,
  6. JAW = jaw width.

Principal Component Analysis: Step 1 Data analysis

- As usual, we start by taking a look at the data globally



Principal Component Analysis: Step 1 Data analysis

- We will exclude data from Group 1 and mix data from Group 2 and 3.
- We get the following Covariance matrix for our 6 variables.

```
> cov(mypcadatainuse)
```

	WDIM	CIRCUM	FBEYE	EYEHD	EARHD	JAW
WDIM	0.37016949	0.6020339	0.14881356	0.04440678	0.10711864	0.20932203
CIRCUM	0.60203390	2.6528136	0.80827119	0.66450847	0.10186441	0.37962712
FBEYE	0.14881356	0.8082712	0.45825989	0.01126554	-0.01322034	0.11984181
EYEHD	0.04440678	0.6645085	0.01126554	1.47371751	0.25220339	-0.05438418
EARHD	0.10711864	0.1018644	-0.01322034	0.25220339	0.48800847	-0.03559322
JAW	0.20932203	0.3796271	0.11984181	-0.05438418	-0.03559322	0.32368362

- We see almost 2 groups of variables:
  1. First Group with WDIM, CIRCUM, FBEYE and JAW
  2. Second Group with EYEHD and EARHD.

Principal Component Analysis: Step 2 Principal Component Result

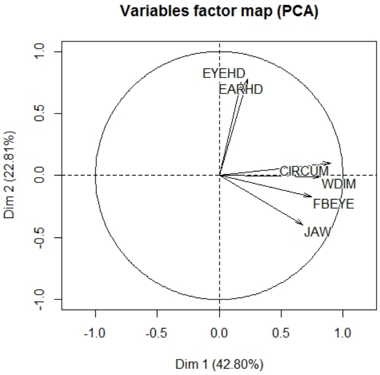
- Applying in R the method

```
> summary(PCA(mypcadatainuse))$call

Call:
PCA(X = mypcadatainuse)


Eigenvalues
                       Dim.1   Dim.2   Dim.3   Dim.4   Dim.5   Dim.6
Variance               2.568   1.369   0.934   0.678   0.321   0.131
% of var.             42.796  22.810  15.563  11.300   5.348   2.183
Cumulative % of var.  42.796  65.606  81.169  92.469  97.817 100.000
```

- We see that the first component explain 43% of the overall variation.
- The second component explain a further 23%.
- Using the first 2 components brings a cumulative proportion of variance to 65%.
- The rest of the components (Components 3 to 6) explain less than 35% overall.

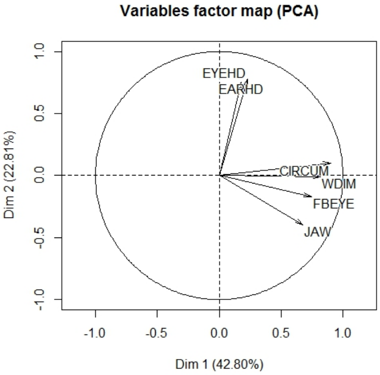
Principal Component Analysis: Variable Factor Map

- We can produce a variable factor map to better understand the role of each factor.
- The variable factor map show the variables and organized them along dimensions.
- Here the first two dimensions are represented.



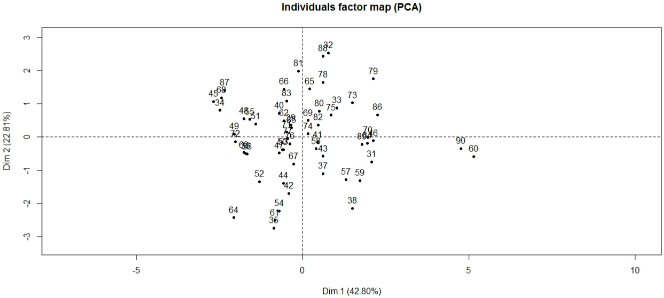
Principal Component Analysis: Variable Factor Map

- Dimension 1 (x-axis): highly correlated to CIRCUM, FBEYE, JAW and WDIM.
- Dimension 1 is moderately correlated to JAW
- Dimension 1 is poorly correlated to EYEHD and EARHD.
- Dimension 2 is well correlated to EYEHD and EARHD.
- It seems that we have 2 groups of variables playing a different role.



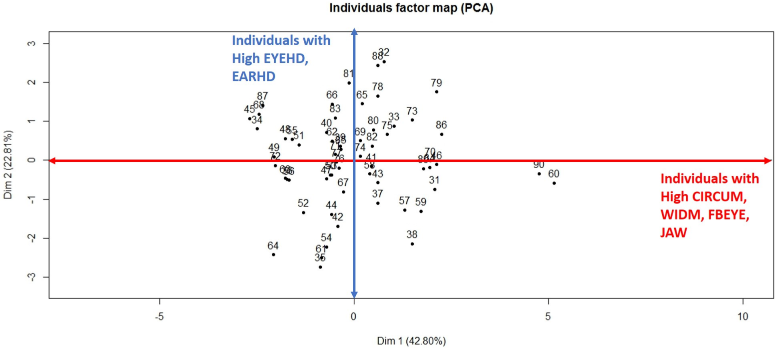
Principal Component Analysis: Individual Factor Map

- We can produce the individual factor map, showing each observations along the principal component.
- The idea is to start analyzing groups of data.



Principal Component Analysis: Variable Factor Map

- For our dataset, we get the following

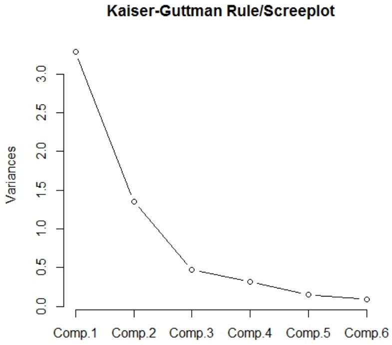


Principal Component Analysis: How many components ?

- Generally speaking we would prefer to work with 2 components as it is easier to represent graphically.
- However, 2 components are not always sufficient.
- A first rule of thumb is to stop adding components whent the total variance explained exceeds a high value, like 80% for example.
- Another rule is the **Kaiser-Guttman rule**.
- The Kaiser-Guttman rule states that components with an eigenvalue greater to 1 should be retained.
- The reason for this is we have p variables so the sum of the eigenvalues is p. A value above 1 is above average.

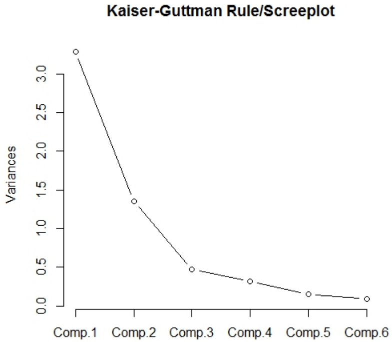
Principal Component Analysis: Screeplot

- A screeplot represents the values of each eigenvalue.
- According to the Kaiser-Guttman rule, we shold stop at Component 2.



Principal Component Analysis: Screeplot

- A screeplot represents the values of each eigenvalue.
- According to the Kaiser-Guttman rule, we shold stop at Component 2.



Principal Component Analysis: Executive Summary

- A principal component is a linear combination of the variables.
- Principal component analysis explains the total variance.
- The level of explanations depends on the number of components used, i.e. the percentage of variance explained.
- There is no assumptions behind the method.
- We end up the process by analyzing the role of each variable and work with groups (see dedicated chapter).

PART III: FACTOR ANALYSIS

1 Introduction to Factor Analysis

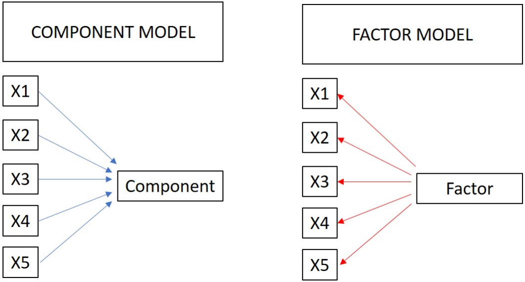
2 Principal Component Analysis

3 Factor Analysis
- Reminder: Factor vs Component
- Factor Analysis: the new meaning of our data
- Factor Analysis: Common Variance and Specific Variance
- Factor analysis: Fitting Idea
- Factor analysis: Model assumptions
- Factor Analysis: Maximum Likelihood Estimate

4 Factor Rotation

Factor vs Component

- We remember the key slide presented at the beginning of the Chapter:



- We have to think the converse way.

Factor Analysis: Our data are representing our factors

- Suppose we have our p variables $X_1, \ldots, X_p$.
- We are looking for a certain number of Factors F_i (for example m factors F_i here).
- Each variable X_j is a linear combination of these factors F_i plus a residual.

$$X_j = a_{j1} \cdot F_1 + a_{j2} \cdot F_2 + \ldots + a_{jm} \cdot F_m + \epsilon_j$$

- What we see, i.e. our data X_j, are in fact representing our Factors F_i.

Factor Analysis: Factor loadings

$$X_j = a_{j1} \cdot F_1 + a_{j2} \cdot F_2 + \ldots + a_{jm} \cdot F_m + \epsilon_j$$

- The factor loading a_{j1} is the factor loading of the j^{th} on the 1st factor.
- As in PCA, the factors loading give us an idea about how much the variable has contributed to the factor.
- A first problem is in how many m factors. Of course, if we change the number of "underlying" factors, the fit and the loadings will change.

Factor Analysis: Common and Specific Variance

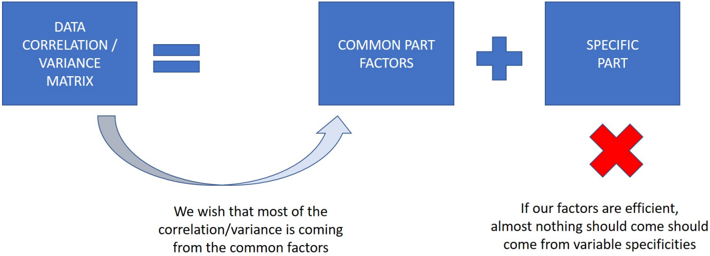
$$X_j = a_{j1} \cdot F_1 + a_{j2} \cdot F_2 + \ldots + a_{jm} \cdot F_m + \epsilon_j$$

- The above equation also mean that we can structure the information in 2 parts:
  1. The one coming from the **common factors**, i.e. the $a_{j1} \cdot F_1 + a_{j2} \cdot F_2 + \ldots + a_{jm} \cdot F_m$. It leads to the **common variance**.
  2. Common variance is also called **communalities**.
  3. The one coming on top of the **common factors**, ϵ_j, i.e. **specific** to the variable X_j, leading to the **specific variance**.
  4. Specific variance is also called **uniqueness**.

Factor analysis: Fitting Idea

- The key idea to find our unknown factors is the following

$$X_j \quad = a_{j1} \cdot F_1 + a_{j2} \cdot F_2 + \ldots + a_{jm} \cdot F_m \quad + \epsilon_j$$



Factor analysis: Consequences

As a result, we have to note the following consequences

- Common and unique factors are uncorrelated, otherwise the model does not make sense

$$\mathrm{Cor}(F_j, \epsilon_i) = 0$$

- Unique factors are all uncorrelated (and with mean 0).

$$\mathrm{Cor}(\epsilon_i, \epsilon_j) = 0$$

If the common factor model perfectly holds, then the correlation between data should fully come from the Factors, not the residuals. There is no partial correlation in the following sense

$$\mathrm{Cor}(x_i, x_j | \mathrm{Factors}) = \mathrm{Cor}(\epsilon_i, \epsilon_j) = 0$$

Factor analysis: Maximum Likelihood Estimate

- We assume that all ϵ_j are following a normal distribution.
- We apply the Maximum Likelihood method to our data.
- The Maximum Likelihood method may not converge.
- However, it is the only method indicating if the number of factors is sufficient through a χ^2 test.
- The hypothesis tested are

$$\begin{aligned} H_0 : &= \text{k factors are sufficient} \\ H_1 : &= \text{More than k factors are requested} \end{aligned}$$

- If the p-value is less than α (say 5%) we have to take more factors into account.
- The weakness of this test is the one coming from the χ^2 test, i.e. its sensitivity to the number of data.

Factor analysis: Maximum Likelihood Estimate

```
> n.factors <- 2
>
> myfactofit <- factanal(mypcadatainuse,
+                 n.factors,              # number of factors to extract
+                 rotation="none")
>
> print(myfactofit, digits=2, cutoff=.3, sort=TRUE)

Call:
factanal(x = mypcadatainuse, factors = n.factors, rotation = "none")

Uniquenesses:
  WDIM CIRCUM  FBEYE  EYEHD  EARHD    JAW
  0.48   0.00   0.46   0.83   0.97   0.00

Loadings:
       Factor1 Factor2
WDIM    0.72
CIRCUM  0.84    0.54
FBEYE   0.62    0.39
JAW     0.84   -0.54
EYEHD           0.38
EARHD

               Factor1 Factor2
SS loadings       2.34    0.91
Proportion Var    0.39    0.15
Cumulative Var    0.39    0.54

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 25.69 on 4 degrees of freedom.
The p-value is 3.65e-05
```

Factor analysis: Maximum Likelihood Estimate

```
> n.factors <- 2
>
> myfactofit <- factanal(mypcadatainuse,
+                 n.factors,              # number of factors to extract
+                 rotation="none")
>
> print(myfactofit, digits=2, cutoff=.3, sort=TRUE)

Call:
factanal(x = mypcadatainuse, factors = n.factors, rotation = "none")

Uniquenesses:
  WDIM CIRCUM  FBEYE  EYEHD  EARHD    JAW
  0.48   0.00   0.46   0.83   0.97   0.00

Loadings:
       Factor1 Factor2
WDIM    0.72
CIRCUM  0.84    0.54
FBEYE   0.62    0.39
JAW     0.84   -0.54
EYEHD           0.38
EARHD

               Factor1 Factor2
SS loadings       2.34    0.91
Proportion Var    0.39    0.15
Cumulative Var    0.39    0.54

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 25.69 on 4 degrees of freedom.
The p-value is 3.65e-05
```

- In our case, we should increase the number of factors to use more than 3 factors.
- However, in our case, we reach a problem as taking 4 factors to represent 6 variables is not really of interest.
- Generally, after deciding about the number of factors, an analysis of them is conducted and a name is attributed to each factor.

PART IV: FACTOR ROTATION

1. Introduction to Factor Analysis
2. Principal Component Analysis
3. Factor Analysis
4. **Factor Rotation**
   - Factor Analysis: Factors Rotation
   - Factor Analysis: Factors Rotation and Varimax
   - Factor Analysis: Oblique Factors Rotation
   - Factor Analysis: Oblique Factors Rotation, Promax Method
   - Factor Analysis: Promax vs Varimax

Factor analysis: Factors Rotation

- We have been in position to determine a set of factors that are maximizing the variance explained.
- However, this set of factors is not unique. A lot of **rotations** of our factors can lead to the same amount of variance explained.
- If the solution is satisfying from the point of view of the variance explained, the factors computed are may be not the best solution.
- An improved solution can be achieved through rotating the factors.
- The target of the improvment would be to get with a cleaner (simpler) interpretation by "eliminating" the role of many variables.
- Rotated and Unrotated solutions can differ greatly despite it does not affect at all the goodness of fit of our data.

Factor analysis: Factors Rotation and Varimax

- In order to make the interpretation of factors as simple as possible, each factor should have major loadings on only a few variable and the rest near 0.
- The most popular method is the Varimax method.
- The idea of the method is to work with squared loadings.
- In that way, during the optimization, we reinforce the difference between the lowst loadings (near 0) and the highest one.

Factor analysis: Rotated vs Unrotated Factors

We compare the Varimax solution with the Unrotated one

ROTATED VARIMAX SOLUTION

```
> n.factors <- 3
>
> myfactofitvarimax <- factanal(mypcadatainuse,
+                     n.factors,              # number of factors to extract
+                     rotation="varimax")
>
> print(myfactofitvarimax, digits=2, cutoff=.3, sort=TRUE)

Call:
factanal(x = mypcadatainuse, factors = n.factors, rotation = "varimax")

Uniquenesses:
  WDIM CIRCUM  FBEYE  EYEHD  EARHD    JAW
  0.00   0.00   0.35   0.24   0.77   0.59

Loadings:
       Factor1 Factor2 Factor3
CIRCUM  0.90    0.35
FBEYE   0.77
WDIM            0.94
JAW             0.59
EYEHD                   0.84
EARHD                   0.42

               Factor1 Factor2 Factor3
SS loadings       1.58    1.46    1.00
Proportion Var    0.26    0.24    0.17
Cumulative Var    0.26    0.51    0.67

The degrees of freedom for the model is 0 and the fit was 0.0741
```

2.

1.

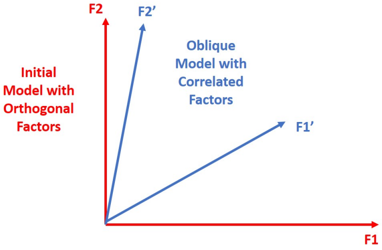
UNROTATED SOLUTION

```
> n.factors <- 3
>
> myfactofit <- factanal(mypcadatainuse,
+                     n.factors,              # number of factors to extract
+                     rotation="none")
>
> print(myfactofit, digits=2, cutoff=.3, sort=TRUE)

Call:
factanal(x = mypcadatainuse, factors = n.factors, rotation = "none")

Uniquenesses:
  WDIM CIRCUM  FBEYE  EYEHD  EARHD    JAW
  0.00   0.00   0.35   0.24   0.77   0.59

Loadings:
       Factor1 Factor2 Factor3
WDIM    0.89           -0.44
CIRCUM  0.90            0.44
FBEYE   0.61   -0.32    0.42
JAW     0.57
EYEHD           0.78    0.32
EARHD           0.40

               Factor1 Factor2 Factor3
SS loadings       2.38    0.91    0.75
Proportion Var    0.40    0.15    0.12
Cumulative Var    0.40    0.55    0.67

The degrees of freedom for the model is 0 and the fit was 0.0741
```

- We observe that the total variance explained is unchanged.
- The variance contribution of each factor has been rearranged and spread differently.
- A clean up of the loadings of each factor with 2-3 coordinates per factor

Factor analysis: Oblique Factors Rotation

- Another idea to simplify and improve results is to relax one of the strongest method assumptions.
- One can sometimes achieve a simpler structure in the factor by allowing the factors to be correlated.
- In order to relax this correlation assumption, we have to consider **oblique** rotations, and no more orthogonal factors.



Factor analysis: Oblique Factors Rotation, Promax Method

- An equivalent method to the Varimax in the context of oblique rotations exists and is called the Promax method
- The algorithms starts with a Varimax and then allow correlation between factors.
- If having correlated factors does not bring anything to our problem, then the solutions Varimax and Promax will be very closed.

Factor analysis: Varimax vs Promax

We compare the Varimax solution with the Promax

ROTATED VARIMAX SOLUTION

```
> n.factors <- 3
>
> myfactofitvarimax <- factanal(mypcadatainuse,
+                     n.factors,                 # number of factors to extract
+                     rotation="varimax")
>
> print(myfactofitvarimax, digits=2, cutoff=.3, sort=TRUE)

Call:
factanal(x = mypcadatainuse, factors = n.factors, rotation = "varimax")

Uniquenesses:
  WDIM CIRCUM  FBEYE  EYEHD  EARHD    JAW
  0.00   0.00   0.35   0.24   0.77   0.59

Loadings:
       Factor1 Factor2 Factor3
CIRCUM  0.90    0.35
FBEYE   0.77                                2.
WDIM            0.94
JAW             0.59
EYEHD                   0.84
EARHD                   0.42

               Factor1 Factor2 Factor3
SS loadings       1.58    1.46    1.00
Proportion Var    0.26    0.24    0.17
Cumulative Var    0.26    0.51    0.67    1.

The degrees of freedom for the model is 0 and the fit was 0.0741
```

ROTATED PROMAX SOLUTION

```
> myfactofitpromax <- factanal(mypcadatainuse,
+                     n.factors,                 # number of factors to extract
+                     rotation="promax")
>
> print(myfactofitpromax, digits=2, cutoff=.3, sort=TRUE)

Call:
factanal(x = mypcadatainuse, factors = n.factors, rotation = "promax")

Uniquenesses:
  WDIM CIRCUM  FBEYE  EYEHD  EARHD    JAW
  0.00   0.00   0.35   0.24   0.77   0.59

Loadings:
       Factor1 Factor2 Factor3
CIRCUM  0.90
FBEYE   0.82
WDIM            0.97
JAW             0.54
EYEHD                   0.87
EARHD           0.33    0.35

               Factor1 Factor2 Factor3
SS loadings       1.59    1.38    0.98
Proportion Var    0.27    0.23    0.16
Cumulative Var    0.27    0.49    0.66

Factor Correlations:
        Factor1 Factor2 Factor3
Factor1    1.00   0.470   0.104
Factor2    0.47   1.000  -0.027
Factor3    0.10  -0.027   1.000
```

- Overall variance explained remained the same.
- Relaxing the factors correlation assumptions does not bring any relevant