Exploratory Factor Analysis (EFA)

Outline

- Introduction
- Orthogonal Factor Model with m Common Factors
- Model Estimation:
  - **The Principal Component (and Principal Factor) Method**
  - **The Maximum Likelihood Method**
- Factor Rotation
- Estimation of Factor Scores:
  - **The Weighted Least Squares Method**
  - **The Unweighted (Ordinary) Least Squares Method**
  - **The Regression Method**
- Major Steps in Exploratory Factor Analysis (EFA)
- Perspectives and A Strategy for Factor Analysis
- Application with R

Introduction

Factor analysis is a theory driven statistical data reduction technique used to explain **covariance** among observed random variables in terms of fewer unobserved random variables named factors/ latent variable.

There are two kinds of factor analysis:

- Exploratory Factor Analysis (EFA)
- Confirmatory Factor Analysis (CFA)

Exploratory:

- summarize data
- describe correlation structure between variables
- generate hypotheses

Confirmatory:

- Testing correlated measurement errors (Test a priori theory)
- Test reliability of measures
- Redundancy test of one-factor vs. multi-factor models

Exploratory Factor Analysis (EFA)

- The first stage consists of the estimation of the parameters in the model and the rotation of the factors, followed by an (often heroic) attempt to interpret the fitted model.
- The second stage is concerned with estimating latent variable scores for each individual in the data set

Orthogonal Factor Model with m Common Factors

$$
\begin{aligned}
X_1 - \mu_1 &= \ell_{11}F_1 + \ell_{12}F_2 + \cdots + \ell_{1m}F_m + \varepsilon_1 \\
X_2 - \mu_2 &= \ell_{21}F_1 + \ell_{22}F_2 + \cdots + \ell_{2m}F_m + \varepsilon_2 \\
&\vdots \\
X_p - \mu_p &= \ell_{p1}F_1 + \ell_{p2}F_2 + \cdots + \ell_{pm}F_m + \varepsilon_p
\end{aligned}
$$

$$
\underset{(p\times 1)}{\mathbf{X}} = \underset{(p\times 1)}{\boldsymbol{\mu}} + \underset{(p\times m)}{\mathbf{L}}\;\underset{(m\times 1)}{\mathbf{F}} + \underset{(p\times 1)}{\boldsymbol{\varepsilon}}
$$

μ_i = *mean* of variable i

ε_i = ith *specific factor*

F_j = jth *common factor*

ℓ_{ij} = *loading* of the ith variable on the jth factor

$$
\underset{(p\times 1)}{\mathbf{X} - \boldsymbol{\mu}} = \underset{(p\times m)}{\mathbf{L}}\;\underset{(m\times 1)}{\mathbf{F}} + \underset{(p\times 1)}{\boldsymbol{\varepsilon}}
$$

The unobservable random vectors $\mathbf{F}$ and $\boldsymbol{\varepsilon}$ satisfy the following conditions:

$\mathbf{F}$ and $\boldsymbol{\varepsilon}$ are independent

$E(\mathbf{F}) = \mathbf{0}$, $\mathrm{Cov}(\mathbf{F}) = \mathbf{I}$

$E(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}$, where $\boldsymbol{\Psi}$ is a diagonal matrix

Covariance Structure for The Orthogonal Factor Model

1. $\mathrm{Cov}(\mathbf{X}) = \mathbf{LL}' + \mathbf{\Psi}$

   or

$$\mathrm{Var}(X_i) = \ell_{i1}^2 + \cdots + \ell_{im}^2 + \psi_i$$

$$\mathrm{Cov}(X_i, X_k) = \ell_{i1}\ell_{k1} + \cdots + \ell_{im}\ell_{km}$$

2. $\mathrm{Cov}(\mathbf{X}, \mathbf{F}) = \mathbf{L}$

   or

$$\mathrm{Cov}(X_i, F_j) = \ell_{ij}$$

$$\underbrace{\sigma_{ii}}_{\mathrm{Var}(X_i)} = \underbrace{\ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{im}^2}_{\text{communality}} + \underbrace{\psi_i}_{\text{specific variance}}$$

$$h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{im}^2$$

$$\sigma_{ii} = h_i^2 + \psi_i, \qquad i = 1, 2, \ldots, p$$

- Communality of X_i: h_i^2
  = % variance of X_i explained by $F_1, F_2, \ldots, F_m$

- Uniqueness of X_i (specific variance)= residual variance of X_i

$$\psi_i = \sigma_{ii} - h_i^2$$

or

$$\psi_i = 1 - h_i^2 \text{ (for standardize form)}$$

```
Factor Analysis using method =  minres
Call: fa(r = data, nfactors = 3, rotate = "oblimin", fm = "minres")
Standardized loadings (pattern matrix) based upon correlation matrix
                      MR1   MR2   MR3    h2   u2 com
Price                0.19  0.52  0.02 0.298 0.70 1.3
Safety               0.27 -0.14 -0.16 0.124 0.88 2.2
Exterior_Looks       0.22 -0.05 -0.17 0.075 0.93 2.0
Space_comfort        0.85 -0.06 -0.01 0.737 0.26 1.0
Technology           0.32  0.01  0.09 0.120 0.88 1.1
After_Sales_Service  0.39  0.13  0.21 0.246 0.75 1.8
Resale_Value        -0.13  0.73  0.03 0.579 0.42 1.1
Fuel_Type            0.56  0.09 -0.06 0.303 0.70 1.1
Fuel_Efficiency      0.13  0.29  0.53 0.484 0.52 1.7
Color               -0.21  0.03  0.60 0.373 0.63 1.3
Maintenance          0.07  0.49  0.30 0.417 0.58 1.7
Test_drive           0.16 -0.09  0.37 0.173 0.83 1.5
Product_reviews      0.21  0.10  0.42 0.278 0.72 1.6
Testimonials         0.05 -0.45  0.45 0.311 0.69 2.0
```

Factors

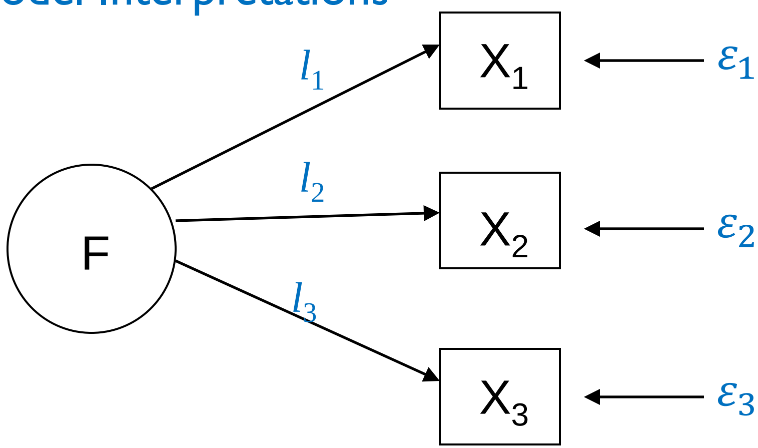
Loadings

Communality of X_i

Uniqueness of X_i

One Common Factor Model:

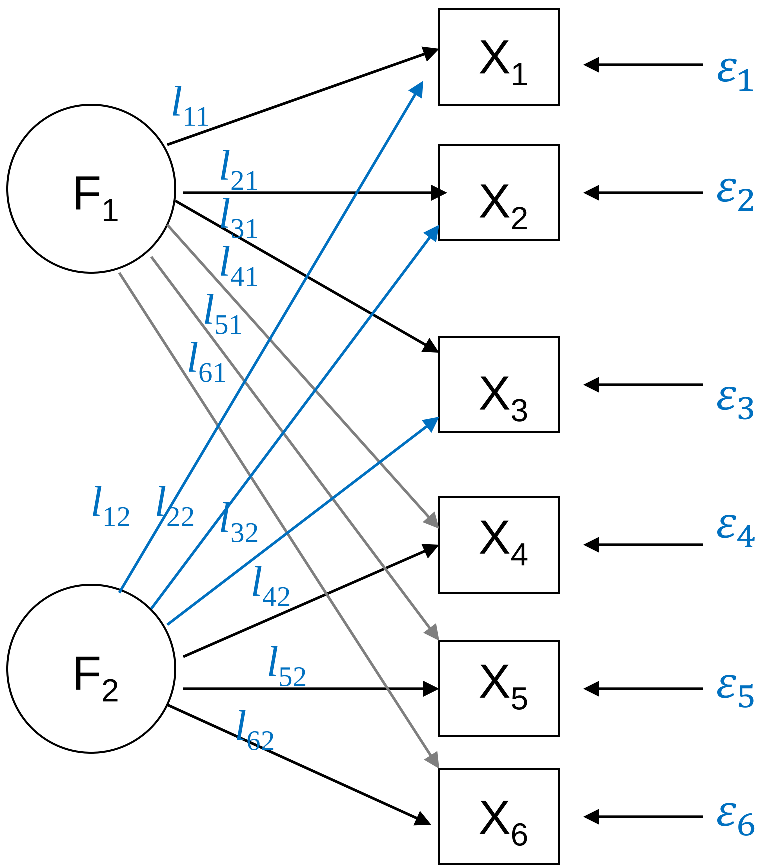
Model Interpretations



$$X_1 = \mu_1 + l_1 F + \varepsilon_1$$
$$X_2 = \mu_2 + l_2 F + \varepsilon_2$$
$$X_3 = \mu_3 + l_3 F + \varepsilon_3$$

Example:
Spearman considered a sample of children's examination marks in three subjects, Classics (X1), French (X2), and English (X3).
In this example, the underlying latent variable or common factor F, might possibly be equated with intelligence or general intellectual ability.

Two-Common Factor Model (Orthogonal):



Orthogonal (F_1 & F_2 independent):
$cov(F_1, F_2) = 0$

Model Interpretation

Given all variables in **standardized form**, i.e. $Var(X_i) = Var(F_i) = 1$

Factor loadings: l_{ij}

$l_{ij} = corr(X_i, F_j)$

Communality of X_i: h_i^2

$h_i^2 = l_{i1}^2 + l_{i2}^2$

= % variance of X_i explained by F_1 & F_2

Uniqueness of X_i (specific variance): $1 - h_i^2$

= residual variance of X_i

Model Estimation

- Goal: Does the factor model, with a small number of factors, adequately represent the data?
- If the **off-diagonal elements of sample covariance matrix $\boldsymbol{S}$ are small** or those of the sample correlation matrix R essentially zero, the variables are not related, and a factor analysis will not prove useful.

- If covariance matrix $\boldsymbol{\Sigma}$ appears to deviate significantly from a diagonal matrix, then a factor model can be entertained, and the initial problem is one of estimating the factor loadings l_{ij} and specific variances ψ_i.
- We shall consider two of the most popular methods of parameter estimation,
  - the principal component (and the related principal factor) method
  - and the maximum likelihood method.
  - The solution from either method can be rotated in order to simplify the interpretation of factors
  - It is always prudent to try more than one method of solution, if the factor model is appropriate for the problem at hand, the solutions should be consistent with one another.

The Principal Component (and Principal Factor) Method

Let $\boldsymbol{\Sigma}$ have eigenvalue-eigenvector pairs $(\lambda_i, \boldsymbol{e_i})$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$, then

$$\boldsymbol{\Sigma} = \lambda_1\mathbf{e}_1\mathbf{e}_1' + \lambda_2\mathbf{e}_2\mathbf{e}_2' + \cdots + \lambda_p\mathbf{e}_p\mathbf{e}_p'$$

$$= \left[\sqrt{\lambda_1}\,\mathbf{e}_1 \mid \sqrt{\lambda_2}\,\mathbf{e}_2 \mid \cdots \mid \sqrt{\lambda_p}\,\mathbf{e}_p\right] \begin{bmatrix} \sqrt{\lambda_1}\,\mathbf{e}_1' \\ \hdashline \sqrt{\lambda_2}\,\mathbf{e}_2' \\ \hdashline \vdots \\ \hdashline \sqrt{\lambda_p}\,\mathbf{e}_p' \end{bmatrix}$$

This fits the prescribed covariance structure for the factor analysis model having as many factors as variables ($m = p$) and specific variances $\psi_i = 0$ for all i.

$$\underset{(p\times p)}{\boldsymbol{\Sigma}} = \underset{(p\times p)}{\mathbf{L}}\;\underset{(p\times p)}{\mathbf{L}'} + \underset{(p\times p)}{\mathbf{0}} = \mathbf{L}\mathbf{L}'$$

We prefer models that explain the covariance structure in terms of just a few common factors, ($m < p$), thus

$$\boldsymbol{\Sigma} \doteq \boldsymbol{L}\boldsymbol{L}' + \boldsymbol{\Psi}$$

PRINCIPAL COMPONENT SOLUTION OF THE FACTOR MODEL

The principal component factor analysis of the sample covariance matrix **S** is specified in terms of its eigenvalue–eigenvector pairs $(\hat{\lambda}_1, \hat{\mathbf{e}}_1), (\hat{\lambda}_2, \hat{\mathbf{e}}_2), \ldots, (\hat{\lambda}_p, \hat{\mathbf{e}}_p)$, where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p$. Let $m < p$ be the number of common factors. Then the matrix of estimated factor loadings $\{\tilde{\ell}_{ij}\}$ is given by

$$\tilde{\mathbf{L}} = \left[\sqrt{\hat{\lambda}_1}\,\hat{\mathbf{e}}_1 \mid \sqrt{\hat{\lambda}_2}\,\hat{\mathbf{e}}_2 \mid \cdots \mid \sqrt{\hat{\lambda}_m}\,\hat{\mathbf{e}}_m\right] \tag{9-15}$$

The estimated specific variances are provided by the diagonal elements of the matrix $\mathbf{S} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}'$, so

$$\tilde{\boldsymbol{\Psi}} = \begin{bmatrix} \tilde{\psi}_1 & 0 & \cdots & 0 \\ 0 & \tilde{\psi}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{\psi}_p \end{bmatrix} \quad \text{with} \quad \tilde{\psi}_i = s_{ii} - \sum_{j=1}^{m} \tilde{\ell}_{ij}^2 \tag{9-16}$$

Communalities are estimated as

$$\tilde{h}_i^2 = \tilde{\ell}_{i1}^2 + \tilde{\ell}_{i2}^2 + \cdots + \tilde{\ell}_{im}^2 \tag{9-17}$$

The principal component factor analysis of the sample correlation matrix is obtained by starting with **R** in place of **S**.

- Ideally, the contributions of the first few factors to the sample variances of the variables should be large.
- The contribution to the sample variance s_{ii} from the first common factor is ${\tilde{l}_{i1}}^2$.
- The contribution to the total sample variance, $s_{11} + s_{22} + \cdots + s_{pp} = tr(\boldsymbol{S})$, from the first common factor is then

$$
{\tilde{l}_{11}}^2 + {\tilde{l}_{21}}^2 + \cdots + {\tilde{l}_{p1}}^2 = \left(\sqrt{\hat{\lambda}_1}, \hat{\boldsymbol{e}}_{\mathbf{1}}\right)' \left(\sqrt{\hat{\lambda}_1}, \hat{\boldsymbol{e}}_{\mathbf{1}}\right) = \hat{\lambda}_1
$$

since the eigenvector $\hat{\boldsymbol{e}}_{\mathbf{1}}$ has unit length (length=1).

- In general,

$$
\begin{pmatrix} \text{Proportion of total} \\ \text{sample variance} \\ \text{due to } j\text{th factor} \end{pmatrix} = \begin{cases} \dfrac{\hat{\lambda}_j}{s_{11} + s_{22} + \cdots + s_{pp}} & \text{for a factor analysis of } \mathbf{S} \\[2ex] \dfrac{\hat{\lambda}_j}{p} & \text{for a factor analysis of } \mathbf{R} \end{cases}
$$

The Maximum Likelihood Method

- If the common factors F and the specific factors $\boldsymbol{\varepsilon}$ can be assumed to be normally distributed, then maximum likelihood estimates of the factor loadings and specific variances may be obtained.
- When $\boldsymbol{F}_j$ and $\boldsymbol{\varepsilon}_j$ are jointly normal, the observations $\boldsymbol{X}_j - \boldsymbol{\mu} = \boldsymbol{L}\boldsymbol{F}_j + \boldsymbol{\varepsilon}_j$ are then normal, and the likelihood is

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-np/2} |\boldsymbol{\Sigma}|^{-n/2} \exp\left[-\frac{1}{2} tr\left(\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' + (\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})'\right)\right)\right]$$

  which depends on $\boldsymbol{L}$ and $\boldsymbol{\Psi}$ through $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}' + \boldsymbol{\Psi}$.
- The maximum likelihood estimates $\hat{\boldsymbol{L}}$ and $\widehat{\boldsymbol{\Psi}}$ must be obtained by numerical maximization of this likelihood.
- It is desirable to make $\boldsymbol{L}$ well defined by imposing the computationally convenient uniqueness condition

$$\boldsymbol{L}'\boldsymbol{\Psi}^{-1}\boldsymbol{L} = \Delta \qquad \text{is a diagonal matrix}$$

Let $\boldsymbol{X_1}, \boldsymbol{X_2}, \dots, \boldsymbol{X_n}$ be a random sample from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \boldsymbol{LL}' + \boldsymbol{\Psi}$ is the covariance matrix for the m common factor model of $\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{LF} + \boldsymbol{\varepsilon}$. The maximum likelihood estimators $\hat{\boldsymbol{L}}$, $\widehat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ maximize likelihood $L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ subject to $\hat{\boldsymbol{L}}'\widehat{\boldsymbol{\Psi}}^{-1}\hat{\boldsymbol{L}}$ being diagonal.

- So the maximum likelihood estimates of the communalities are

$$\hat{h}_i^{\;2} = \tilde{l}_{i1}^{\;2} + \tilde{l}_{i2}^{\;2} + \cdots + \tilde{l}_{im}^{\;2}, \qquad i = 1,2, \dots, p$$

$$\begin{pmatrix} \text{Proportion of Total sample} \\ \text{variance due to } j\text{th factor} \end{pmatrix} = \frac{\tilde{l}_{1j}^{\;2} + \tilde{l}_{2j}^{\;2} + \cdots + \tilde{l}_{pj}^{\;2}}{s_{11} + s_{22} + \cdots + s_{pp}}$$

- If the variables $\boldsymbol{X}$ are standardized so that $\boldsymbol{Z} = \boldsymbol{V}^{-\frac{1}{2}}(\boldsymbol{X} - \boldsymbol{\mu})$ the covariance matrix will be a correlation matrix $\boldsymbol{\rho}$, and

$$\begin{pmatrix} \text{Proportion of Total sample} \\ \text{variance due to } j\text{th factor} \end{pmatrix} = \frac{\tilde{l}_{1j}^{\;2} + \tilde{l}_{2j}^{\;2} + \cdots + \tilde{l}_{pj}^{\;2}}{p}$$

```
##
## Call:
## factanal(x = food, factors = 2)
##
## Uniquenesses:
##       Oil  Density   Crispy Fracture Hardness
##     0.334    0.156    0.042    0.256    0.407
##
## Loadings:
##          Factor1 Factor2
## Oil      -0.816
## Density   0.919
## Crispy   -0.745   0.635
## Fracture  0.645  -0.573
## Hardness          0.764
##
##                Factor1 Factor2
## SS loadings      2.490   1.316
## Proportion Var   0.498   0.263
## Cumulative Var   0.498   0.761
##
## Test of the hypothesis that 2 factors are sufficient.
## The chi square statistic is 0.27 on 1 degree of freedom.
## The p-value is 0.603
```

$$\text{SS loadings} = {\tilde{l}_{1j}}^2 + {\tilde{l}_{2j}}^2 + \cdots + {\tilde{l}_{pj}}^2$$

$$\begin{pmatrix}\text{Proportion of Total sample} \\ \text{variance due to } j\text{th factor}\end{pmatrix} = \frac{{\tilde{l}_{1j}}^2 + {\tilde{l}_{2j}}^2 + \cdots + {\tilde{l}_{pj}}^2}{p}$$

Factor Rotation

- Goal is simple structure
- Make factors more easily interpretable
  - While keeping the number of factors and communalities of **X**s fixed!!!
- **Rotation does NOT improve fit!**

When number of factors $m > 1$, there is always some inherent ambiguity associated with the factor model. To see this, let $\boldsymbol{T}$ be any $m \times m$ orthogonal matrix, so that $\boldsymbol{TT'} = \boldsymbol{T'T} = \boldsymbol{I}$. Then the expression in $\boldsymbol{X} - \boldsymbol{\mu} = \boldsymbol{LF} + \boldsymbol{\varepsilon}$ can be written

$$\boldsymbol{X} - \boldsymbol{\mu} = \boldsymbol{LTT'F} + \boldsymbol{\varepsilon} = \boldsymbol{L}^*\boldsymbol{F}^* + \boldsymbol{\varepsilon}$$

Where $\boldsymbol{L}^* = \boldsymbol{LT}$ and $\boldsymbol{F}^* = \boldsymbol{T'F}$

$$E(\boldsymbol{F}^*) = E(\boldsymbol{T'F}) = \boldsymbol{T'}E(\boldsymbol{F}) = \boldsymbol{0}$$

$$Cov(\boldsymbol{F}^*) = Cov(\boldsymbol{T'F}) = \boldsymbol{T'}Cov(\boldsymbol{F})\boldsymbol{T} = \boldsymbol{T'IT} = \boldsymbol{T'T} = \boldsymbol{I}$$

- That is, the factors $\boldsymbol{F}$ and $\boldsymbol{F}^* = \boldsymbol{T'F}$ have the same statistical properties, and even though the loadings $\boldsymbol{L}^*$ are, in general, different from the loadings $\boldsymbol{L}$.
- So the loadings $\boldsymbol{L}$ and $\boldsymbol{L}^*$ give the same representation.
- The communalities are also unaffected by the choice of $\boldsymbol{T}$

Types of Rotations:

- Orthogonal Rotation (The uncorrelated common factors are regarded as unit vectors along perpendicular coordinate axes):
  - Quartimax
  - Varimax

    Varimax procedure selects the orthogonal transformation $\boldsymbol{T}$ that makes,

$$V = 1/p \sum_{j=1}^{m} \left[\sum_{i=1}^{p} \tilde{l}_{ij}^{*4} - \left(\sum_{i=1}^{p} \tilde{l}_{ij}^{*2} \right)^{2} /p \right]$$

    as large as possible.

$$V \propto \sum_{j=1}^{m} (\text{variance of squares of (scaled) loadings for jth factor})$$

    So, maximizing V corresponds to "spreading out" the squares of the loadings on each factor as much as possible.
  - Equimax
- Oblique Rotation (Oblique rotations are so named because they correspond to a nonrigid rotation of coordinate axes leading to new axes that are not perpendicular)

Factor Scores

- The estimated values of the common factors, called factor scores
- Factor scores are not estimates of unknown parameters in the usual sense.
- The factors **F** themselves are variables
- "Object's" score is weighted combination of scores on input variables

$$\hat{\boldsymbol{f}} = \widehat{\boldsymbol{W}}\boldsymbol{X}$$

- These weights are **NOT** the factor loadings!
- Different approaches exist for estimating $\widehat{\boldsymbol{W}}$:
  - Weighted least squares method
  - Ordinary least squares method
  - Regression method
- Factor scores are not unique
- Using factors scores instead of factor indicators can reduce measurement error, but does **NOT** remove it.
- Therefore, using factor scores as predictors in conventional regressions leads to inconsistent coefficient estimators!

The Weighted Least Squares Method

Suppose first that the mean vector $\boldsymbol{\mu}$, the factor loadings $\boldsymbol{L}$, and the specific variance $\boldsymbol{\Psi}$ are known for the factor model

$$\boldsymbol{X}-\boldsymbol{\mu}=\boldsymbol{L}\boldsymbol{F}+\boldsymbol{\varepsilon}$$

Further, regard the specific factors $\boldsymbol{\varepsilon}=\left[\varepsilon_1,\varepsilon_2,\dots,\varepsilon_p\right]$ as errors. Since $Var(\varepsilon_i)=\psi_i,\quad i=1,2,\dots,p$ need not be equal, Bartlett has suggested that weighted least squares be used to estimate the common factor values.

Bartlett proposed choosing the estimates $\hat{\mathbf{f}}$ of $\mathbf{f}$ to minimize the sum of the squares of the errors, weighted by the reciprocal of their variances, is

$$\sum_{i=1}^{p}\frac{{\varepsilon_i}^2}{\psi_i}=\boldsymbol{\varepsilon}'\boldsymbol{\Psi}^{-1}\boldsymbol{\varepsilon}=(\mathbf{x}-\boldsymbol{\mu}-\boldsymbol{L}\mathbf{f})'\boldsymbol{\Psi}^{-1}(\mathbf{x}-\boldsymbol{\mu}-\boldsymbol{L}\mathbf{f})$$

The solution is

$$\hat{\mathbf{f}}=\left(\boldsymbol{L}'\boldsymbol{\Psi}^{-1}\boldsymbol{L}\right)^{-1}\boldsymbol{L}'\boldsymbol{\Psi}^{-1}(\mathbf{x}-\boldsymbol{\mu})$$

By maximum likelihood method, we have the estimates $\hat{\boldsymbol{L}}$, $\widehat{\boldsymbol{\Psi}}$, and $\hat{\boldsymbol{\mu}}=\bar{\mathbf{x}}$ and we take this as the true values and obtain the factor scores for the jth case as

$$\hat{\mathbf{f}}_j=\left(\hat{\boldsymbol{L}}'\widehat{\boldsymbol{\Psi}}^{-1}\hat{\boldsymbol{L}}\right)^{-1}\hat{\boldsymbol{L}}'\widehat{\boldsymbol{\Psi}}^{-1}\left(\mathbf{x}_j-\bar{\mathbf{x}}\right),\qquad j=1,2,\dots,n$$

If rotated loading $\boldsymbol{L}^*=\boldsymbol{L}\boldsymbol{T}$ are used, then $\hat{\mathbf{f}}_j{}^*=\boldsymbol{T}'\hat{\mathbf{f}}_j$

The Unweighted (Ordinary) Least Squares Method

Suppose first that the mean vector $\boldsymbol{\mu}$, the factor loadings $\boldsymbol{L}$, and the specific variance $\boldsymbol{\Psi}$ are known for the factor model

$$\boldsymbol{X} - \boldsymbol{\mu} = \boldsymbol{L}\boldsymbol{F} + \boldsymbol{\varepsilon}$$

the specific factors $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p]$as errors.

- If the factor loadings $\boldsymbol{L}$ are estimated by the principal component method, it is customary to generate factor scores using an unweighted (ordinary) least squares procedure.
- Implicitly, this amounts to assuming that the $Var(\varepsilon_i) = \psi_i, \quad i = 1,2, \dots, p$ are equal or nearly equal.

The factor scores are then

$$\hat{\mathbf{f}}_j = \left(\hat{\boldsymbol{L}}'\hat{\boldsymbol{L}}\right)^{-1}\hat{\boldsymbol{L}}'(\mathbf{x}_j - \bar{\mathbf{x}}), \qquad j = 1,2, \dots, n$$

Since $\hat{\boldsymbol{L}} = \left[\sqrt{\hat{\lambda}_1}, \hat{\boldsymbol{e}}_1 \quad \dots \quad \sqrt{\hat{\lambda}_m}, \hat{\boldsymbol{e}}_m\right]$ we have then

$$\hat{\mathbf{f}}_j = \begin{bmatrix} \frac{1}{\sqrt{\hat{\lambda}_1}}\hat{\mathbf{e}}_1'(\mathbf{x}_j - \bar{\mathbf{x}}) \\ \frac{1}{\sqrt{\hat{\lambda}_2}}\hat{\mathbf{e}}_2'(\mathbf{x}_j - \bar{\mathbf{x}}) \\ \vdots \\ \frac{1}{\sqrt{\hat{\lambda}_m}}\hat{\mathbf{e}}_m'(\mathbf{x}_j - \bar{\mathbf{x}}) \end{bmatrix}$$

we see that the $\hat{\mathbf{f}}_j$ are nothing more than the first m (scaled) principal components, evaluated at $\mathbf{x}_j$.

The Regression Method

Suppose first that the mean vector $\boldsymbol{\mu}$, the factor loadings $\boldsymbol{L}$, and the specific variance $\boldsymbol{\Psi}$ are known for the factor model

$$\boldsymbol{X} - \boldsymbol{\mu} = \boldsymbol{LF} + \boldsymbol{\varepsilon}$$

the specific factors $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p]$as errors.

- When the common factors $\boldsymbol{F}$ and the specific factors (or errors) $\boldsymbol{\varepsilon}$ are jointly normally distributed the linear combination $\boldsymbol{X} - \boldsymbol{\mu} = \boldsymbol{LF} + \boldsymbol{\varepsilon}$ has an $N_{m+p}(\mathbf{0}, \boldsymbol{\Sigma} = \boldsymbol{LL}' + \boldsymbol{\Psi})$ distribution.
- Given any vector of observations $\mathbf{x}_j$, and taking the maximum likelihood estimates $\hat{\boldsymbol{L}}$ and $\widehat{\boldsymbol{\Psi}}$, and $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ as the true values, the jth factor score vector is given by

$$\begin{aligned}\hat{\mathbf{f}}_j &= \hat{\boldsymbol{L}}'\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}), \qquad j = 1,2,\dots,n \\ &= \hat{\boldsymbol{L}}'\boldsymbol{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})\end{aligned}$$

Or if correlation matrix is factored,

$$\hat{\mathbf{f}}_j = \hat{\boldsymbol{L}}_Z{}'\boldsymbol{R}^{-1}\boldsymbol{z}_j, \qquad j = 1,2,\dots,n$$

Where $\boldsymbol{z}_j = \boldsymbol{D}^{-1/2}(\mathbf{x}_j - \bar{\mathbf{x}})$, $\quad \boldsymbol{D}^{1/2}$ is a standard deviation diagonal matrix, and

$$\hat{\boldsymbol{\rho}} = \hat{\boldsymbol{L}}_Z\hat{\boldsymbol{L}}_Z{}' + \widehat{\boldsymbol{\Psi}}_Z$$

Major Steps in EFA

1. Data collection and preparation (including evaluate data set, e.g KMO, Bartlett sphericity, normality if used MLE)
2. Choose number of factors to extract
3. Model fitting (Extracting initial factors)
4. Rotation to a final solution
5. Model diagnosis/refinement
6. Interpret & write-up results
7. Derivation of factor scores to be used in further analysis (e.g. SEM analysis)

Perspectives and A Strategy for Factor Analysis

- There are many decisions that must be made in any factor analytic study. Probably the most important decision is the choice of m, the number of common factors.
  - Although a large sample test of the adequacy of a model is available for a given m, it is suitable only for data that are approximately normally distributed. Moreover, the test will most assuredly reject the model for small m if the number of variables and observations is large.
  - The final choice of m is based on some combination of: (1) the proportion of the sample variance explained, (2) subject-matter knowledge, and (3) the "reasonableness" of the results.
- The choice of the solution method and type of rotation is a less crucial decision.
- At the present time, factor analysis still maintains the flavor of an art, and no single strategy.

Jhonson et al. (2002) suggest and illustrate one reasonable option:

1. Perform a principal component factor analysis. This method is particularly appropriate for a first pass through the data. (It is not required that $\boldsymbol{R}$ or $\boldsymbol{S}$ be nonsingular.)
   a) Look for suspicious observations by plotting the factor scores. Also, calculate standardized scores for each observation and squared distances
   b) Try a varimax rotation.
2. Perform a maximum likelihood factor analysis, including a varimax rotation.
3. Compare the solutions obtained from the two factor analyses.
   a) Do the loadings group in the same manner?
   b) Plot factor scores obtained for principal components against scores from the maximum likelihood analysis.
4. Repeat the first three steps for other numbers of common factors m. Do extra factors necessarily contribute to the understanding and interpretation of the data?
5. For large data sets, split them in half and perform a factor analysis on each part.
6. Compare the two results with each other and with that obtained from the complete data set to check the stability of the solution. (The data might be divided at random or by placing the first half of the cases in one group and the second half of the cases in the other group.)

See Example 9.14 (Jhonson, 2002)

TABLE 9.10 FACTOR ANALYSIS OF CHICKEN-BONE DATA

Principal Component

Variable	Estimated factor loadings F_1	F_2	F_3	Rotated estimated loadings F_1^*	F_2^*	F_3^*	$\tilde{\psi}_i$
1. Skull length	.741	.350	.573	.355	.244	.902	.00
2. Skull breadth	.604	.720	−.340	.235	.949	.211	.00
3. Femur length	.929	−.233	−.075	.921	.164	.218	.08
4. Tibia length	.943	−.175	−.067	.904	.212	.252	.08
5. Humerus length	.948	−.143	−.045	.888	.228	.283	.08
6. Ulna length	.945	−.189	−.047	.908	.192	.264	.07
Cumulative proportion of total (standardized) sample variance explained	.743	.873	.950	.576	.763	.950	

Maximum Likelihood

Variable	Estimated factor loadings F_1	F_2	F_3	Rotated estimated loadings F_1^*	F_2^*	F_3^*	$\hat{\psi}_i$
1. Skull length	.602	.214	.286	.467	.506	.128	.51
2. Skull breadth	.467	.177	.652	.211	.792	.050	.33
3. Femur length	.926	.145	−.057	.890	.289	.084	.12
4. Tibia length	1.000	.000	−.000	.936	.345	−.073	.00
5. Humerus length	.874	.463	−.012	.831	.362	.396	.02
6. Ulna length	.894	.336	−.039	.857	.325	.272	.09
Cumulative proportion of total (standardized) sample variance explained	.667	.738	.823	.559	.779	.823	

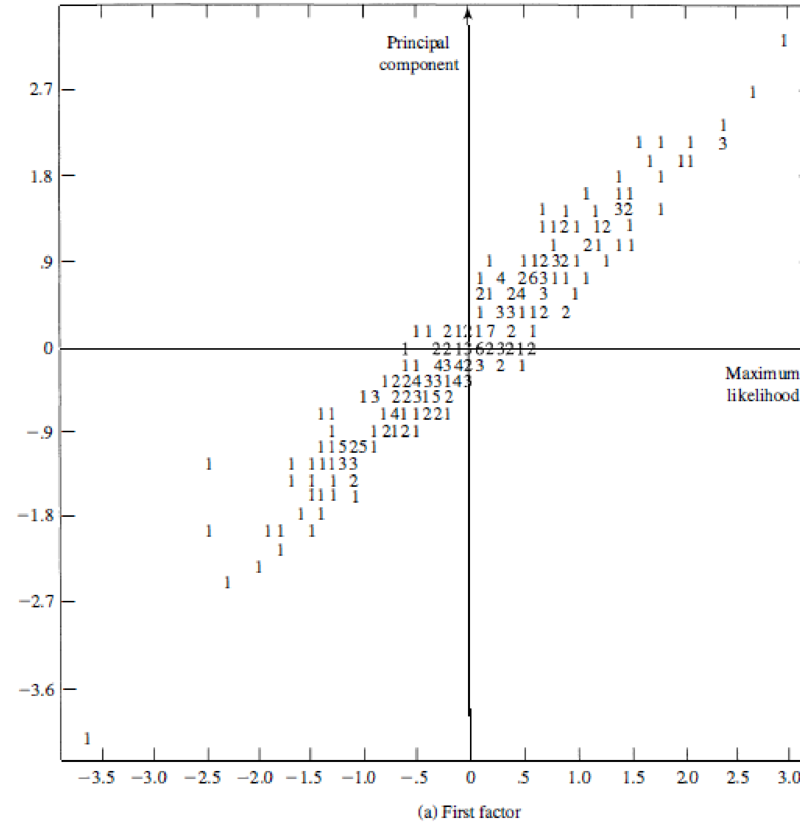
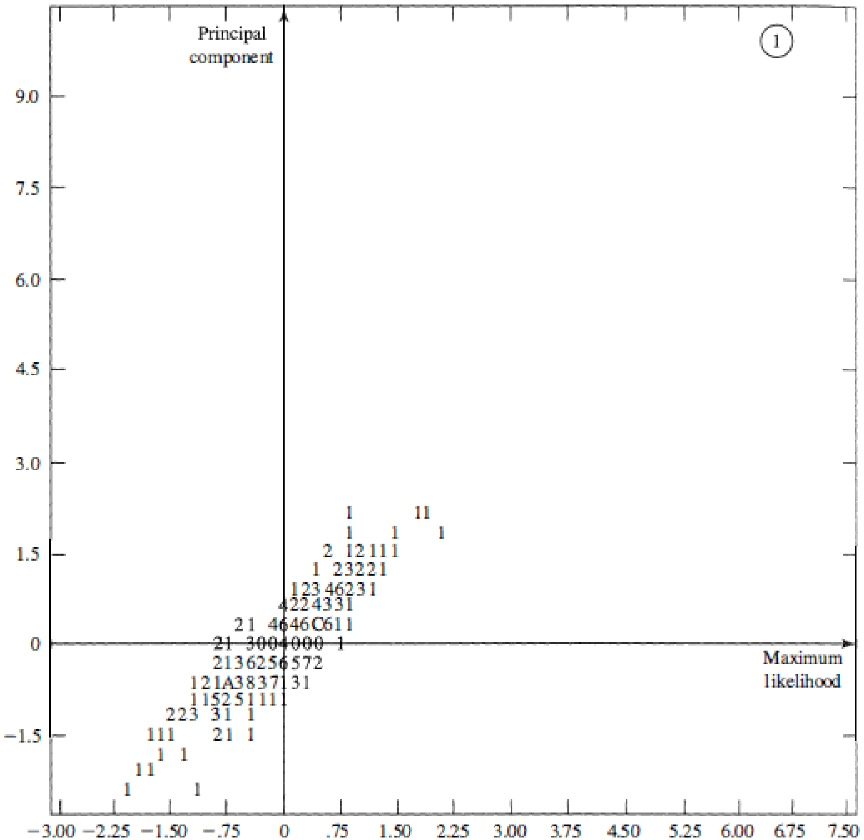
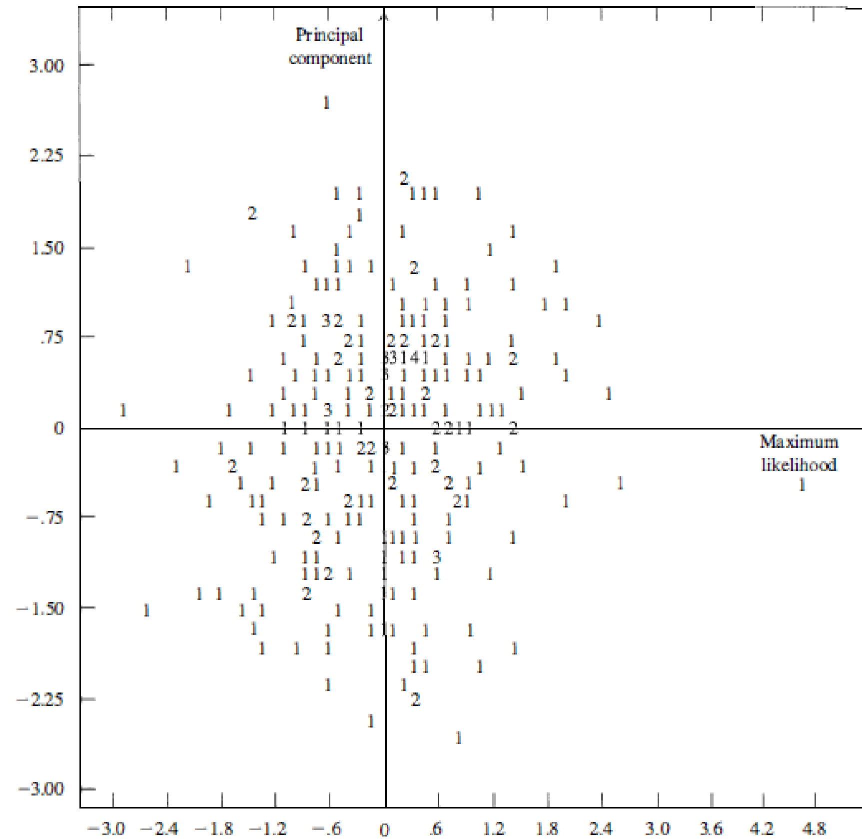


Figure 9.6 Pairs of factor scores for the chicken-bone data. (Loadings are estimated by principal component and maximum likelihood methods.)



(b) Second factor

Figure 9.6 *(continued)*



(c) Third factor

Figure 9.6 *(continued)*

- If the loadings on a particular factor agree, the pairs of scores should cluster tightly about the 45° line through the origin.
- Sets of loadings that do not agree will produce factor scores that deviate from this pattern.
- If the latter occurs, it is usually associated with the last factors and may suggest that the number of factors is too large. That is, the last factors are not meaningful. This seems to be the case with the third factor in the chicken-bone data, as indicated by Plot (c) in Figure 9.6.

Plots of pairs of factor scores using estimated loadings from two solution methods are also good tools for detecting outliers.

- If the sets of loadings for a factor tend to agree, outliers will appear as points in the neighborhood of the 45 ° line, but far from the origin and the cluster of the remaining points.
- It is clear from Plot (b) in Figure 9.6 that one of the 276 observations is not consistent with the others. It has an unusually large F2-score.
- When this point, [39.1, 39.3, 75.7, 115, 73.4, 69.1], was removed and the analysis repeated, the loadings were not altered appreciably.

Relationship between factor analysis and principal component analysis

- Both factor analysis and principal component analysis have the goal of reducing dimensionality.
- The differences are:
  - In factor analysis, the variables are expressed as linear combinations of the factors, whereas the principal components are linear functions of the variables,
  - in principal component analysis, the emphasis is on explaining the total variance $\sum_i^p s_{ii}$, as contrasted with the attempt to explain the covariances in factor analysis,
  - principal component analysis requires essentially no assumptions, while factor analysis makes several key assumptions,
  - the principal components are unique (assuming distinct eigenvalues of $\boldsymbol{S}$), whereas the factors are not unique, subject to an arbitrary rotation,
  - if we change the number of factors, the (estimated) factors change. This does not happen in principal components
  - the calculation of factor scores is not as straightforward as the calculation of principal component scores.

Application with R

Maximum Likelihood method:

```
factanal(x, factors =n, data = NULL, covmat = NULL, n.obs = NA, subset, na.action, start =
NULL,
scores = c("none", "regression", "Bartlett"), rotation = "varimax", control = NULL, ...)
```

Keterangan:

x	: berupa formula (tanpa variable respon) atau matriks numerik dari objek
factors	: banyak faktor yang akan diestimasi.
data	: data frame yang digunakan apabila x berupa formula.
Covmat	: matriks varians-kovarians dalam hal ini matriks korelasi juga merupakan matriks varians-kovarians
n.obs	: banyaknya pengamatan dari data, opsi ini digunakan apabila opsi 'covmat' adalah matriks kovarians.
subset	: Spesifikasi pengamatan yang digunakan. Digunakan apabila opsi 'x' digunakan sebagai matriks data atau formula.
na.action	: opsi untuk data hilang, digunakan apabila opsi 'x' berupa formula
start	: dengan nilai default 'NULL' adalah matriks yang berisi nilai awal dengan tiap kolom merupakan set awal uniquenesses.
Scores	: metode menghitung skor factor. Ada dua tipe yaitu "regression" bila menggunakan metode Thompson, dan "Bartlett"' bila menggunakan meto de Bartlett's weighted least-squares
Rotation	: tipe rotasi yang digunakan, secara default bernilai "none"

```
#normality test
library(MVN)
mvn(dataku, mvnTest = "mardia")

#Bartlett test dan KMO
library(REdaS)
bart_spher(dataku)
KMOS(dataku)

#Analisis Faktor Eksploratori dg MLE
R<-cov(dataku)
eigen<-eigen(R)
fa1<-factanal(factors=2,covmat=R)
```

See Beverit & Hothorn (2011) p.148 – example of exploratory factor analysis: expectations of life

Other Package

```
library(psych)
fa(r, nfactors=1,n.obs = NA,n.iter=1, rotate="oblimin", scores="regression",
residuals=FALSE, SMC=TRUE, covar=FALSE,missing=FALSE,impute="median", min.err =
0.001, max.iter = 50,symmetric=TRUE, warnings=TRUE, fm=“minres",
alpha=.1,p=.05,oblique.scores=FALSE,np.obs=NULL,use="pairwise",cor="cor",
correct=.5,weight=NULL,...)
```

r	: a correlation or covariance matrix or a raw data matrix. If raw data, the correlation matrix will be found using pairwise deletion. If covariances are supplied, they will be converted to correlations unless the covar option is TRUE.
nfactors	: number of factors to extract, default is 1
rotate	:"none", "varimax", "quartimax", "bentlerT", "equamax", "varimin", "geominT" and "bifactor" are orthogonal rotations. "Promax", "promax", "oblimin", "simplimax", "bentlerQ, "geominQ" and "biquartimin" and "cluster" are possible oblique transformations of the solution. The default is to do a oblimin transformation, although versions prior to 2009 defaulted to varimax. SPSS seems to do a Kaiser normalization before doing Promax, this is done here by the call to "promax" which does the normalization before calling Promax in GPArotation

scores	the default="regression" finds factor scores using regression. Alternatives for estimating factor scores include simple regression ("Thurstone"), correlaton preserving ("tenBerge") as well as "Anderson" and "Bartlett" using the appropriate algorithms (factor.scores). Although scores="tenBerge" is probably preferred for most solutions, it will lead to problems with some improper correlation matrices.
fm	Factoring method fm="minres" will do a minimum residual as will fm="uls". Both of these use a first derivative. fm="ols" differs very slightly from "minres" in that it minimizes the entire residual matrix using an OLS procedure but uses the empirical first derivative. This will be slower. fm="wls" will do a weighted least squares (WLS) solution, fm="gls" does a generalized weighted least squares (GLS), fm="pa" will do the principal factor solution, fm="ml" will do a maximum likelihood factor analysis. fm="minchi" will minimize the sample size weighted chi square when treating pairwise correlations with different number of subjects per pair. fm ="minrank" will do a minimum rank factor analysis. "old.min" will do minimal residual the way it was done prior to April, 2017 (see discussion below). fm="alpha" will do alpha factor analysis as described in Kaiser and Coffey (1965)

Examples:

1. https://towardsdatascience.com/exploratory-factor-analysis-in-r-e31b0015f224
2. https://rpubs.com/pjmurphy/758265
3. https://www.promptcloud.com/exploratory-factor-analysis-in-r/#:~:text=What%20is%20exploratory%20factor%20analysis,a%20smaller%20number%20of%20variables.