Multivariate Verfahren 2

factor analysis

Helmut Waldl

April 30th 2012

factor analysis

principal components

With factor analysis we try to explain the total covariance matrix of the observable variables as good as possible with a minimum number of factors.

We may achieve this objective with a linear transformation: the observable variables are mapped to uncorrelated new variables, the so-called **principal components** which are sorted in descending order by their variances.

The first $k < p$ principal components are the new variables (factors) that jointly explain the biggest proportion of the total variance amongst all k-tuples of new variables.

Transforming to principal components is pure data manipulation which contrary to factor analysis is not based on an explicit model.

factor analysis

principal components

With factor analysis we try to explain the total covariance matrix of the observable variables as good as possible with a minimum number of factors.

We may achieve this objective with a linear transformation: the observable variables are mapped to uncorrelated new variables, the so-called **principal components** which are sorted in descending order by their variances.

The first $k < p$ principal components are the new variables (factors) that jointly explain the biggest proportion of the total variance amongst all k-tuples of new variables.

Transforming to principal components is pure data manipulation which contrary to factor analysis is not based on an explicit model.

factor analysis

principal components

With factor analysis we try to explain the total covariance matrix of the observable variables as good as possible with a minimum number of factors.

We may achieve this objective with a linear transformation: the observable variables are mapped to uncorrelated new variables, the so-called **principal components** which are sorted in descending order by their variances.

The first $k < p$ principal components are the new variables (factors) that jointly explain the biggest proportion of the total variance amongst all k-tuples of new variables.

Transforming to principal components is pure data manipulation which contrary to factor analysis is not based on an explicit model.

factor analysis

principal components

With factor analysis we try to explain the total covariance matrix of the observable variables as good as possible with a minimum number of factors.

We may achieve this objective with a linear transformation: the observable variables are mapped to uncorrelated new variables, the so-called **principal components** which are sorted in descending order by their variances.

The first $k < p$ principal components are the new variables (factors) that jointly explain the biggest proportion of the total variance amongst all k-tuples of new variables.

Transforming to principal components is pure data manipulation which contrary to factor analysis is not based on an explicit model.

factor analysis

principal components

Principal components focus on the variances of the observable variables, whereas in factor analysis the covariances take center stage.

Principal components are the basis for two important approximative methods for parameter estimation in factor analysis models:

- principal component analysis
- principal factor analysis (principal-axis factoring, common factor analysis)

Both methods are based on the so-called **principal axis transformation**

Attention! Here we always analyze standardized data. The standardization is particularly important with principal component analysis because the results are not scale-invariant as in ML-factor analysis there.

factor analysis

principal components

Principal components focus on the variances of the observable variables, whereas in factor analysis the covariances take center stage.

Principal components are the basis for two important approximative methods for parameter estimation in factor analysis models:

- principal component analysis
- principal factor analysis (principal-axis factoring, common factor analysis)

Both methods are based on the so-called **principal axis transformation**

Attention! Here we always analyze standardized data. The standardization is particularly important with principal component analysis because the results are not scale-invariant as in ML-factor analysis there.

factor analysis

principal components

Principal components focus on the variances of the observable variables, whereas in factor analysis the covariances take center stage.

Principal components are the basis for two important approximative methods for parameter estimation in factor analysis models:

- principal component analysis
- principal factor analysis (principal-axis factoring, common factor analysis)

Both methods are based on the so-called **principal axis transformation**

Attention! Here we always analyze standardized data. The standardization is particularly important with principal component analysis because the results are not scale-invariant as in ML-factor analysis there.

factor analysis

principal components

Principal components focus on the variances of the observable variables, whereas in factor analysis the covariances take center stage.

Principal components are the basis for two important approximative methods for parameter estimation in factor analysis models:

- principal component analysis
- principal factor analysis (principal-axis factoring, common factor analysis)

Both methods are based on the so-called **principal axis transformation**

Attention! Here we always analyze standardized data. The standardization is particularly important with principal component analysis because the results are not scale-invariant as in ML-factor analysis there.

factor analysis

principal components

Principal components focus on the variances of the observable variables, whereas in factor analysis the covariances take center stage.

Principal components are the basis for two important approximative methods for parameter estimation in factor analysis models:

- principal component analysis
- principal factor analysis (principal-axis factoring, common factor analysis)

Both methods are based on the so-called **principal axis transformation**

Attention! Here we always analyze standardized data. The standardization is particularly important with principal component analysis because the results are not scale-invariant as in ML-factor analysis there.

factor analysis

principal components

Principal components focus on the variances of the observable variables, whereas in factor analysis the covariances take center stage.

Principal components are the basis for two important approximative methods for parameter estimation in factor analysis models:

- principal component analysis
- principal factor analysis (principal-axis factoring, common factor analysis)

Both methods are based on the so-called **principal axis transformation**

Attention! Here we always analyze standardized data. The standardization is particularly important with principal component analysis because the results are not scale-invariant as in ML-factor analysis there.

factor analysis

principal components

Principal components focus on the variances of the observable variables, whereas in factor analysis the covariances take center stage.

Principal components are the basis for two important approximative methods for parameter estimation in factor analysis models:

- principal component analysis
- principal factor analysis (principal-axis factoring, common factor analysis)

Both methods are based on the so-called **principal axis transformation**

Attention! Here we always analyze standardized data. The standardization is particularly important with principal component analysis because the results are not scale-invariant as in ML-factor analysis there.

factor analysis

principal axis transformation

With principal axis transformation we decompose the standardized data matrix according to

$$Z = F \cdot L^T \qquad \text{with } (n \times p) - \text{matrix } F = (F_1 \dots F_p)$$

where the columns $F_1 \dots F_p$ are orthonormal and L is a $(p \times p)$-matrix.

$$R = Z^T Z = L \cdot F^T F \cdot L^T = L \cdot L^T$$

Principal component analysis takes the first k normalized principal axes $F_1, \dots, F_k$ as "most important factors" and yields the following decomposition of Z

$$Z = F^{(k)} L^{(k)\,T} + E \qquad \text{with } (n \times p) - \text{matrices } Z, E$$

$$(n \times k) - \text{matrix } F^{(k)} = (F_1 \dots F_k) \qquad \text{and } (p \times k) - \text{matrix } L^{(k)}$$

factor analysis

principal axis transformation

With principal axis transformation we decompose the standardized data matrix according to

$$Z = F \cdot L^T \qquad \text{with } (n \times p) - \text{matrix } F = (F_1 \dots F_p)$$

where the columns $F_1 \dots F_p$ are orthonormal and L is a $(p \times p)$-matrix.

$$R = Z^T Z = L \cdot F^T F \cdot L^T = L \cdot L^T$$

Principal component analysis takes the first k normalized principal axes $F_1, \dots, F_k$ as "most important factors" and yields the following decomposition of Z

$$Z = F^{(k)} L^{(k)\,T} + E \qquad \text{with } (n \times p) - \text{matrices } Z, E$$

$$(n \times k) - \text{matrix } F^{(k)} = (F_1 \dots F_k) \qquad \text{and } (p \times k) - \text{matrix } L^{(k)}$$

factor analysis

principal axis transformation

With principal axis transformation we decompose the standardized data matrix according to

$$Z = F \cdot L^T \qquad \text{with } (n \times p) - \text{matrix } F = (F_1 \dots F_p)$$

where the columns $F_1 \dots F_p$ are orthonormal and L is a $(p \times p)$-matrix.

$$R = Z^T Z = L \cdot F^T F \cdot L^T = L \cdot L^T$$

Principal component analysis takes the first k normalized principal axes $F_1, \dots, F_k$ as "most important factors" and yields the following decomposition of Z

$$Z = F^{(k)} L^{(k)\,T} + E \qquad \text{with } (n \times p) - \text{matrices } Z, E$$

$$(n \times k) - \text{matrix } F^{(k)} = (F_1 \dots F_k) \qquad \text{and } (p \times k) - \text{matrix } L^{(k)}$$

factor analysis

principal axis transformation

Principal factor analysis starts with the fundamental theorem for the empirical correlation matrix: $R = L \cdot L^T + V$

First we estimate the communalities $h_i^2 = \sum_{j=1}^{k} l_{ij}^2$ and in this way get V.

Afterwards we perform a principal axis transformation of the reduced correlation matrix $R - V = L \cdot L^T := R_h$.

factor analysis

principal axis transformation

Principal factor analysis starts with the fundamental theorem for the empirical correlation matrix: $R = L \cdot L^T + V$

First we estimate the communalities $h_i^2 = \sum_{j=1}^{k} l_{ij}^2$ and in this way get V.

Afterwards we perform a principal axis transformation of the reduced correlation matrix $R - V = L \cdot L^T := R_h$.

factor analysis

principal axis transformation

Principal factor analysis starts with the fundamental theorem for the empirical correlation matrix: $R = L \cdot L^T + V$

First we estimate the communalities $h_i^2 = \sum_{j=1}^{k} l_{ij}^2$ and in this way get V.

Afterwards we perform a principal axis transformation of the reduced correlation matrix $R - V = L \cdot L^T := R_h$.

factor analysis

principal axis transformation

Given: standardized data $Z = (Z_1, \ldots, Z_p)$

Find: $t_i \ldots p$-vectors with $\|t_i\| = 1 \quad i = 1, \ldots, p$ such that $H_i = Z \cdot t_i$ are pairwise orthogonal with $H_1^T H_1 = \max_i\{H_i^T H_i\}$, i.e. H_1 has maximal variance. H_2 varies maximal among all vectors orthogonal to H_1 etc.

Interpretation: H_1 has the biggest variability among all linear combinations of $Z_1, \ldots, Z_p$ and for this reason bears most information and is the most important factor, H_2 is the second most important factor etc.

Theorem: Let Z be a $(n \times p)$-matrix with $\mathrm{rk}(Z) = p < n$. Let further be $H = (H_1, \ldots, H_p)$ the matrix of the principal axes of Z.

Then we compute H as follows: $\boxed{H = Z \cdot T}$, in the $(p \times p)$-matrix T there are the normalized eigenvectors affiliated with the sorted eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$ of $R = Z^T Z$ (T is orthogonal anyway because R is symmetric).

factor analysis

principal axis transformation

Given: standardized data $Z = (Z_1, \ldots, Z_p)$

Find: $t_i \ldots p$-vectors with $\|t_i\| = 1 \quad i = 1, \ldots, p$ such that $H_i = Z \cdot t_i$ are pairwise orthogonal with $H_1^T H_1 = \max_i\{H_i^T H_i\}$, i.e. H_1 has maximal variance. H_2 varies maximal among all vectors orthogonal to H_1 etc.

Interpretation: H_1 has the biggest variability among all linear combinations of $Z_1, \ldots, Z_p$ and for this reason bears most information and is the most important factor, H_2 is the second most important factor etc.

Theorem: Let Z be a $(n \times p)$-matrix with $\text{rk}(Z) = p < n$. Let further be $H = (H_1, \ldots, H_p)$ the matrix of the principal axes of Z.

Then we compute H as follows: $\boxed{H = Z \cdot T}$, in the $(p \times p)$-matrix T there are the normalized eigenvectors affiliated with the sorted eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$ of $R = Z^T Z$ (T is orthogonal anyway because R is symmetric).

factor analysis

principal axis transformation

Given: standardized data $Z = (Z_1, \ldots, Z_p)$

Find: $t_i \ldots p$-vectors with $\|t_i\| = 1 \quad i = 1, \ldots, p$ such that $H_i = Z \cdot t_i$ are pairwise orthogonal with $H_1^T H_1 = \max_i\{H_i^T H_i\}$, i.e. H_1 has maximal variance. H_2 varies maximal among all vectors orthogonal to H_1 etc.

Interpretation: H_1 has the biggest variability among all linear combinations of $Z_1, \ldots, Z_p$ and for this reason bears most information and is the most important factor, H_2 is the second most important factor etc.

Theorem: Let Z be a $(n \times p)$-matrix with $\text{rk}(Z) = p < n$. Let further be $H = (H_1, \ldots, H_p)$ the matrix of the principal axes of Z.

Then we compute H as follows: $\boxed{H = Z \cdot T}$, in the $(p \times p)$-matrix T there are the normalized eigenvectors affiliated with the sorted eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$ of $R = Z^T Z$ (T is orthogonal anyway because R is symmetric).

factor analysis

principal axis transformation

Given: standardized data $Z = (Z_1, \ldots, Z_p)$

Find: $t_i \ldots p$-vectors with $\|t_i\| = 1 \quad i = 1, \ldots, p$ such that $H_i = Z \cdot t_i$ are pairwise orthogonal with $H_1^T H_1 = \max_i\{H_i^T H_i\}$, i.e. H_1 has maximal variance. H_2 varies maximal among all vectors orthogonal to H_1 etc.

Interpretation: H_1 has the biggest variability among all linear combinations of $Z_1, \ldots, Z_p$ and for this reason bears most information and is the most important factor, H_2 is the second most important factor etc.

Theorem: Let Z be a $(n \times p)$-matrix with $\mathrm{rk}(Z) = p < n$. Let further be $H = (H_1, \ldots, H_p)$ the matrix of the principal axes of Z.

Then we compute H as follows: $\boxed{H = Z \cdot T}$, in the $(p \times p)$-matrix T there are the normalized eigenvectors affiliated with the sorted eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$ of $R = Z^T Z$ (T is orthogonal anyway because R is symmetric).

factor analysis

principal axis transformation

<u>Given:</u> standardized data $Z = (Z_1, \ldots, Z_p)$

<u>Find:</u> $t_i \ldots p$-vectors with $\|t_i\| = 1 \quad i = 1, \ldots, p$ such that $H_i = Z \cdot t_i$ are pairwise orthogonal with $H_1^T H_1 = \max_i\{H_i^T H_i\}$, i.e. H_1 has maximal variance. H_2 varies maximal among all vectors orthogonal to H_1 etc.

<u>Interpretation:</u> H_1 has the biggest variability among all linear combinations of $Z_1, \ldots, Z_p$ and for this reason bears most information and is the most important factor, H_2 is the second most important factor etc.

Theorem: Let Z be a $(n \times p)$-matrix with $\mathrm{rk}(Z) = p < n$. Let further be $H = (H_1, \ldots, H_p)$ the matrix of the principal axes of Z.

Then we compute H as follows: $\boxed{H = Z \cdot T}$, in the $(p \times p)$-matrix T there are the normalized eigenvectors affiliated with the sorted eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$ of $R = Z^T Z$ (T is orthogonal anyway because R is symmetric).

factor analysis

properties of principal axes and principal components

- the principal axes $H_1, \ldots, H_p$ are pairwise orthogonal and have the empirical variances $\lambda_1, \ldots, \lambda_p$:

$$H^T H = T^T Z^T Z \cdot T = T^T R \cdot T = \Lambda = \mathrm{diag}(\lambda_1, \ldots \lambda_p)$$

  because $R = T \cdot \Lambda \cdot T^T$

- If R has no multiple eigenvalues the principal axes and the vectors t_i are unique except for the sign.

factor analysis

properties of principal axes and principal components

- the principal axes $H_1, \ldots, H_p$ are pairwise orthogonal and have the empirical variances $\lambda_1, \ldots, \lambda_p$:

$$H^T H = T^T Z^T Z \cdot T = T^T R \cdot T = \Lambda = \mathrm{diag}(\lambda_1, \ldots \lambda_p)$$

  because $R = T \cdot \Lambda \cdot T^T$

- If R has no multiple eigenvalues the principal axes and the vectors t_i are unique except for the sign.

factor analysis

properties of principal axes and principal components

- $Z = H \cdot T^T$. With **normalized principal axes** $F = H \cdot \Lambda^{-\frac{1}{2}}$ the so-called **principal components** we get a representation of Z in orthonormal factors $F = (F_1, \ldots, F_p)$:

$$Z = F \cdot L^T \qquad F = H \cdot \Lambda^{-\frac{1}{2}} \quad \Rightarrow \quad L = T \cdot \Lambda^{\frac{1}{2}}$$

$$F^T F = \Lambda^{-\frac{1}{2}} H^T H \cdot \Lambda^{-\frac{1}{2}} = \Lambda^{-\frac{1}{2}} \Lambda \cdot \Lambda^{-\frac{1}{2}} = I$$

$$R = Z^T Z = L \cdot F^T F \cdot L^T = L \cdot L^T = T \cdot \Lambda \cdot T^T$$

- the decomposition $Z = F \cdot \Lambda^{\frac{1}{2}} T^T$ is called the **singular value decomposition** of Z (the singular values of Z are the eigenvalues of $Z^T Z$)

factor analysis

properties of principal axes and principal components

- $Z = H \cdot T^T$. With **normalized principal axes** $F = H \cdot \Lambda^{-\frac{1}{2}}$ the so-called **principal components** we get a representation of Z in orthonormal factors $F = (F_1, \ldots, F_p)$:

$$Z = F \cdot L^T \qquad F = H \cdot \Lambda^{-\frac{1}{2}} \quad \Rightarrow \quad L = T \cdot \Lambda^{\frac{1}{2}}$$

$$F^T F = \Lambda^{-\frac{1}{2}} H^T H \cdot \Lambda^{-\frac{1}{2}} = \Lambda^{-\frac{1}{2}} \Lambda \cdot \Lambda^{-\frac{1}{2}} = I$$

$$R = Z^T Z = L \cdot F^T F \cdot L^T = L \cdot L^T = T \cdot \Lambda \cdot T^T$$

- the decomposition $Z = F \cdot \Lambda^{\frac{1}{2}} T^T$ is called the **singular value decomposition** of Z (the singular values of Z are the eigenvalues of $Z^T Z$)

factor analysis

properties of principal axes and principal components

- $Z = H \cdot T^T$. With **normalized principal axes** $F = H \cdot \Lambda^{-\frac{1}{2}}$ the so-called **principal components** we get a representation of Z in orthonormal factors $F = (F_1, \dots, F_p)$:

$$Z = F \cdot L^T \qquad F = H \cdot \Lambda^{-\frac{1}{2}} \quad \Rightarrow \quad L = T \cdot \Lambda^{\frac{1}{2}}$$

$$F^T F = \Lambda^{-\frac{1}{2}} H^T H \cdot \Lambda^{-\frac{1}{2}} = \Lambda^{-\frac{1}{2}} \Lambda \cdot \Lambda^{-\frac{1}{2}} = I$$

$$R = Z^T Z = L \cdot F^T F \cdot L^T = L \cdot L^T = T \cdot \Lambda \cdot T^T$$

- the decomposition $Z = F \cdot \Lambda^{\frac{1}{2}} T^T$ is called the **singular value decomposition** of Z (the singular values of Z are the eigenvalues of $Z^T Z$)

factor analysis

properties of principal axes and principal components

- $Z = H \cdot T^T$. With **normalized principal axes** $F = H \cdot \Lambda^{-\frac{1}{2}}$ the so-called **principal components** we get a representation of Z in orthonormal factors $F = (F_1, \ldots, F_p)$:

$$Z = F \cdot L^T \qquad F = H \cdot \Lambda^{-\frac{1}{2}} \quad \Rightarrow \quad L = T \cdot \Lambda^{\frac{1}{2}}$$

$$F^T F = \Lambda^{-\frac{1}{2}} H^T H \cdot \Lambda^{-\frac{1}{2}} = \Lambda^{-\frac{1}{2}} \Lambda \cdot \Lambda^{-\frac{1}{2}} = I$$

$$R = Z^T Z = L \cdot F^T F \cdot L^T = L \cdot L^T = T \cdot \Lambda \cdot T^T$$

- the decomposition $Z = F \cdot \Lambda^{\frac{1}{2}} T^T$ is called the **singular value decomposition** of Z (the singular values of Z are the eigenvalues of $Z^T Z$)

factor analysis

properties of principal axes and principal components

- $Z = H \cdot T^T$. With **normalized principal axes** $F = H \cdot \Lambda^{-\frac{1}{2}}$ the so-called **principal components** we get a representation of Z in orthonormal factors $F = (F_1, \ldots, F_p)$:

$$Z = F \cdot L^T \qquad F = H \cdot \Lambda^{-\frac{1}{2}} \quad \Rightarrow \quad L = T \cdot \Lambda^{\frac{1}{2}}$$

$$F^T F = \Lambda^{-\frac{1}{2}} H^T H \cdot \Lambda^{-\frac{1}{2}} = \Lambda^{-\frac{1}{2}} \Lambda \cdot \Lambda^{-\frac{1}{2}} = I$$

$$R = Z^T Z = L \cdot F^T F \cdot L^T = L \cdot L^T = T \cdot \Lambda \cdot T^T$$

- the decomposition $Z = F \cdot \Lambda^{\frac{1}{2}} T^T$ is called the **singular value decomposition** of Z (the singular values of Z are the eigenvalues of $Z^T Z$)

factor analysis

properties of principal axes and principal components

- $Z = H \cdot T^T$. With **normalized principal axes** $F = H \cdot \Lambda^{-\frac{1}{2}}$ the so-called **principal components** we get a representation of Z in orthonormal factors $F = (F_1, \ldots, F_p)$:

$$Z = F \cdot L^T \qquad F = H \cdot \Lambda^{-\frac{1}{2}} \quad \Rightarrow \quad L = T \cdot \Lambda^{\frac{1}{2}}$$

$$F^T F = \Lambda^{-\frac{1}{2}} H^T H \cdot \Lambda^{-\frac{1}{2}} = \Lambda^{-\frac{1}{2}} \Lambda \cdot \Lambda^{-\frac{1}{2}} = I$$

$$R = Z^T Z = L \cdot F^T F \cdot L^T = L \cdot L^T = T \cdot \Lambda \cdot T^T$$

- the decomposition $Z = F \cdot \Lambda^{\frac{1}{2}} T^T$ is called the **singular value decomposition** of Z (the singular values of Z are the eigenvalues of $Z^T Z$)

factor analysis

stepwise extraction of principal components

We start with the representation of Z:

$$Z = F \cdot L^T = (F_1 \dots F_p) \cdot (l_1 \dots l_p)^T = \sum_{i=1}^{p} F_i l_i^T$$

After the computation of the first principal component $F_1 = \frac{1}{\sqrt{\lambda_1}} H_1$ we may assess Z as a first approximation

$$Z \approx F_1 \cdot l_1^T = \frac{1}{\sqrt{\lambda_1}} H_1 \cdot t_1^T \sqrt{\lambda_1} = H_1 \cdot t_1^T$$

the residuals have the empirical covariance matrix

$$(Z - F_1 \cdot l_1^T)^T (Z - F_1 \cdot l_1 T) = Z^T Z - l_1 \cdot l_1^T$$

the sum of the residual variances is $p - \lambda_1$

factor analysis

stepwise extraction of principal components

We start with the representation of Z:

$$Z = F \cdot L^T = (F_1 \dots F_p) \cdot (l_1 \dots l_p)^T = \sum_{i=1}^{p} F_i l_i^T$$

After the computation of the first principal component $F_1 = \frac{1}{\sqrt{\lambda_1}} H_1$ we may assess Z as a first approximation

$$Z \approx F_1 \cdot l_1^T = \frac{1}{\sqrt{\lambda_1}} H_1 \cdot t_1^T \sqrt{\lambda_1} = H_1 \cdot t_1^T$$

the residuals have the empirical covariance matrix

$$(Z - F_1 \cdot l_1^T)^T (Z - F_1 \cdot l_1 T) = Z^T Z - l_1 \cdot l_1^T$$

the sum of the residual variances is $p - \lambda_1$

factor analysis

stepwise extraction of principal components

We start with the representation of Z:

$$Z = F \cdot L^T = (F_1 \ldots F_p) \cdot (l_1 \ldots l_p)^T = \sum_{i=1}^{p} F_i l_i^T$$

After the computation of the first principal component $F_1 = \frac{1}{\sqrt{\lambda_1}} H_1$ we may assess Z as a first approximation

$$Z \approx F_1 \cdot l_1^T = \frac{1}{\sqrt{\lambda_1}} H_1 \cdot t_1^T \sqrt{\lambda_1} = H_1 \cdot t_1^T$$

the residuals have the empirical covariance matrix

$$(Z - F_1 \cdot l_1^T)^T (Z - F_1 \cdot l_1 T) = Z^T Z - l_1 \cdot l_1^T$$

the sum of the residual variances is $p - \lambda_1$

factor analysis

stepwise extraction of principal components

We start with the representation of Z:

$$Z = F \cdot L^T = (F_1 \dots F_p) \cdot (l_1 \dots l_p)^T = \sum_{i=1}^{p} F_i l_i^T$$

After the computation of the first principal component $F_1 = \frac{1}{\sqrt{\lambda_1}} H_1$ we may assess Z as a first approximation

$$Z \approx F_1 \cdot l_1^T = \frac{1}{\sqrt{\lambda_1}} H_1 \cdot t_1^T \sqrt{\lambda_1} = H_1 \cdot t_1^T$$

the residuals have the empirical covariance matrix

$$(Z - F_1 \cdot l_1^T)^T (Z - F_1 \cdot l_1 T) = Z^T Z - l_1 \cdot l_1^T$$

the sum of the residual variances is $p - \lambda_1$

factor analysis

stepwise extraction of principal components

Now we extract the second principal component $F_2 = \frac{1}{\lambda_2} H_2$. That yields the second approximation of Z

$$Z \approx F_1 \cdot l_1^T + F_2 \cdot l_2^T = H_1 \cdot t_1^T + H_2 \cdot t_2^T$$

the residuals have the empirical covariance matrix

$$(Z - F_1 \cdot l_1^T - F_2 \cdot l_2^T)^T (Z - F_1 \cdot l_1^T - F_2 \cdot l_2^T) = Z^T Z - l_1 \cdot l_1^T - l_2 \cdot l_2^T$$

the sum of the residual variances is $p - \lambda_1 - \lambda_2$
etc.

factor analysis

stepwise extraction of principal components

Now we extract the second principal component $F_2 = \frac{1}{\lambda_2} H_2$. That yields the second approximation of Z

$$Z \approx F_1 \cdot l_1^T + F_2 \cdot l_2^T = H_1 \cdot t_1^T + H_2 \cdot t_2^T$$

the residuals have the empirical covariance matrix

$$(Z - F_1 \cdot l_1^T - F_2 \cdot l_2^T)^T (Z - F_1 \cdot l_1^T - F_2 \cdot l_2^T) = Z^T Z - l_1 \cdot l_1^T - l_2 \cdot l_2^T$$

the sum of the residual variances is $p - \lambda_1 - \lambda_2$
etc.

factor analysis

stepwise extraction of principal components

Now we extract the second principal component $F_2 = \frac{1}{\lambda_2} H_2$. That yields the second approximation of Z

$$Z \approx F_1 \cdot l_1^T + F_2 \cdot l_2^T = H_1 \cdot t_1^T + H_2 \cdot t_2^T$$

the residuals have the empirical covariance matrix

$$(Z - F_1 \cdot l_1^T - F_2 \cdot l_2^T)^T (Z - F_1 \cdot l_1^T - F_2 \cdot l_2^T) = Z^T Z - l_1 \cdot l_1^T - l_2 \cdot l_2^T$$

the sum of the residual variances is $p - \lambda_1 - \lambda_2$
etc.

factor analysis

stepwise extraction of principal components

Finally (after p extracted principal components) we get $Z = \sum_{i=1}^{p} F_i \cdot l_i^T$ without residuum, i.e.

$$\mathbf{0} = (Z - \sum_{i=1}^{p} F_i \cdot l_i^T)^T (Z - \sum_{i=1}^{p} F_i \cdot l_i^T) \quad \Longleftrightarrow \quad Z^T Z = \sum_{i=1}^{p} l_i \cdot l_i^T = L \cdot L^T$$

and the sum of the variances p is partitioned to the principal components as follows: $p = \lambda_1 + \lambda_2 + \ldots + \lambda_p$

i.e. we may use the principal axis transformation as a method of factor analysis if we want to express a maximal portion of the total variance p with a minimal number of factors.

factor analysis

stepwise extraction of principal components

Finally (after p extracted principal components) we get $Z = \sum_{i=1}^{p} F_i \cdot l_i^T$ without residuum, i.e.

$$\mathbf{0} = (Z - \sum_{i=1}^{p} F_i \cdot l_i^T)^T (Z - \sum_{i=1}^{p} F_i \cdot l_i^T) \quad \iff \quad Z^T Z = \sum_{i=1}^{p} l_i \cdot l_i^T = L \cdot L^T$$

and the sum of the variances p is partitioned to the principal components as follows: $p = \lambda_1 + \lambda_2 + \ldots + \lambda_p$

i.e. we may use the principal axis transformation as a method of factor analysis if we want to express a maximal portion of the total variance p with a minimal number of factors.

factor analysis

stepwise extraction of principal components

Finally (after p extracted principal components) we get $Z = \sum_{i=1}^{p} F_i \cdot l_i^T$ without residuum, i.e.

$$\mathbf{0} = (Z - \sum_{i=1}^{p} F_i \cdot l_i^T)^T (Z - \sum_{i=1}^{p} F_i \cdot l_i^T) \quad \Longleftrightarrow \quad Z^T Z = \sum_{i=1}^{p} l_i \cdot l_i^T = L \cdot L^T$$

and the sum of the variances p is partitioned to the principal components as follows: $p = \lambda_1 + \lambda_2 + \ldots + \lambda_p$

i.e. we may use the principal axis transformation as a method of factor analysis if we want to express a maximal portion of the total variance p with a minimal number of factors.