Multivariate Verfahren 2

factor analysis

Helmut Waldl

May 7th and 8th 2012

factor analysis

principal component analysis

According to the factor analysis model with k factors we try to decompose the data matrix of the standardized data:

$$Z = \underbrace{F}_{(n \times k)} \cdot \underbrace{L^T}_{(k \times p)} + \underbrace{E}_{(n \times p)}$$

We proceed merely descriptive: Z is approximated by a linear combination of k orthonormal vectors of factor values such that E is "minimized"

Contrary to ML factor analysis we do not need a stochastic model with various assumptions for the data, we describe the data directly.

factor analysis

principal component analysis

According to the factor analysis model with k factors we try to decompose the data matrix of the standardized data:

$$Z = \underbrace{F}_{(n \times k)} \cdot \underbrace{L^T}_{(k \times p)} + \underbrace{E}_{(n \times p)}$$

We proceed merely descriptive: Z is approximated by a linear combination of k orthonormal vectors of factor values such that E is "minimized"

Contrary to ML factor analysis we do not need a stochastic model with various assumptions for the data, we describe the data directly.

factor analysis

principal component analysis

According to the factor analysis model with k factors we try to decompose the data matrix of the standardized data:

$$Z = \underbrace{F}_{(n\times k)} \cdot \underbrace{L^T}_{(k\times p)} + \underbrace{E}_{(n\times p)}$$

We proceed merely descriptive: Z is approximated by a linear combination of k orthonormal vectors of factor values such that E is "minimized"

Contrary to ML factor analysis we do not need a stochastic model with various assumptions for the data, we describe the data directly.

factor analysis

principal component analysis

With principal component analysis we use the first k principal components $F_1, \ldots, F_k$ of the principal axis transformation as factors to describe the standardized data Z, the last $p-k$ principal components are neglected.

We only need a criterion for the determination of the number of factors k.

The complete principal component partition yields

$$Z = F \cdot L^T \qquad F = (F_1 \ldots F_p) \qquad L = (l_1 \ldots l_p) \qquad L = T \cdot \Lambda^{\frac{1}{2}}$$

with $R = L \cdot L^T$.

The part of the complete variance $p = \mathrm{tr}(R)$ of the standardized data explained by the jth factor is

$$\mathrm{tr}((F_j l_j^T)^T (F_j l_j^T)) = \mathrm{tr}(l_j \cdot l_j^T) = \sum_{i=1}^{p} l_{ij}^2 = \lambda_j$$

factor analysis

principal component analysis

With principal component analysis we use the first k principal components $F_1, \ldots, F_k$ of the principal axis transformation as factors to describe the standardized data Z, the last $p - k$ principal components are neglected.

We only need a criterion for the determination of the number of factors k.

The complete principal component partition yields

$$Z = F \cdot L^T \qquad F = (F_1 \ldots F_p) \qquad L = (l_1 \ldots l_p) \qquad L = T \cdot \Lambda^{\frac{1}{2}}$$

with $R = L \cdot L^T$.

The part of the complete variance $p = \mathrm{tr}(R)$ of the standardized data explained by the jth factor is

$$\mathrm{tr}((F_j l_j^T)^T (F_j l_j^T)) = \mathrm{tr}(l_j \cdot l_j^T) = \sum_{i=1}^{p} l_{ij}^2 = \lambda_j$$

factor analysis

principal component analysis

With principal component analysis we use the first k principal components $F_1, \ldots, F_k$ of the principal axis transformation as factors to describe the standardized data Z, the last $p-k$ principal components are neglected.

We only need a criterion for the determination of the number of factors k.

The complete principal component partition yields

$$Z = F \cdot L^T \qquad F = (F_1 \ldots F_p) \qquad L = (l_1 \ldots l_p) \qquad L = T \cdot \Lambda^{\frac{1}{2}}$$

with $R = L \cdot L^T$.

The part of the complete variance $p = \mathrm{tr}(R)$ of the standardized data explained by the jth factor is

$$\mathrm{tr}((F_j l_j^T)^T (F_j l_j^T)) = \mathrm{tr}(l_j \cdot l_j^T) = \sum_{i=1}^{p} l_{ij}^2 = \lambda_j$$

factor analysis

principal component analysis

With principal component analysis we use the first k principal components $F_1, \ldots, F_k$ of the principal axis transformation as factors to describe the standardized data Z, the last $p - k$ principal components are neglected.

We only need a criterion for the determination of the number of factors k.

The complete principal component partition yields

$$Z = F \cdot L^T \qquad F = (F_1 \ldots F_p) \qquad L = (l_1 \ldots l_p) \qquad L = T \cdot \Lambda^{\frac{1}{2}}$$

with $R = L \cdot L^T$.

The part of the complete variance $p = \mathrm{tr}(R)$ of the standardized data explained by the jth factor is

$$\mathrm{tr}((F_j l_j^T)^T (F_j l_j^T)) = \mathrm{tr}(l_j \cdot l_j^T) = \sum_{i=1}^{p} l_{ij}^2 = \lambda_j$$

factor analysis

principal component analysis

If we choose k factors then the principal component analysis estimates the model as follows:

$$Z = F^{(k)} L^{(k)T} + E^{(k)} \qquad \text{with } F^{(k)} = (F_1 \ldots F_k),$$

$$L^{(k)} = (l_1 \ldots l_k) \text{ and } E^{(k)} = \sum_{i=k+1}^{p} F_i \cdot l_i^T$$

further

$$R = L^{(k)} L^{(k)T} + U^{(k)} \qquad \text{with } U^{(k)} = E^{(k)T} E^{(k)}$$

Unlike the fundamental theorem of factor analysis $U^{(k)}$ in general is no diagonal matrix. Therefore from the factor analytical point of view principal component analysis is only an approximative method.

factor analysis

principal component analysis

If we choose k factors then the principal component analysis estimates the model as follows:

$$Z = F^{(k)} L^{(k)T} + E^{(k)} \qquad \text{with } F^{(k)} = (F_1 \ldots F_k),$$

$$L^{(k)} = (l_1 \ldots l_k) \text{ and } E^{(k)} = \sum_{i=k+1}^{p} F_i \cdot l_i^T$$

further

$$R = L^{(k)} L^{(k)T} + U^{(k)} \qquad \text{with } U^{(k)} = E^{(k)T} E^{(k)}$$

Unlike the fundamental theorem of factor analysis $U^{(k)}$ in general is no diagonal matrix. Therefore from the factor analytical point of view principal component analysis is only an approximative method.

factor analysis

principal component analysis

If we choose k factors then the principal component analysis estimates the model as follows:

$$Z = F^{(k)} L^{(k)T} + E^{(k)} \qquad \text{with } F^{(k)} = (F_1 \dots F_k),$$

$$L^{(k)} = (l_1 \dots l_k) \text{ and } E^{(k)} = \sum_{i=k+1}^{p} F_i \cdot l_i^T$$

further

$$R = L^{(k)} L^{(k)T} + U^{(k)} \qquad \text{with } U^{(k)} = E^{(k)T} E^{(k)}$$

Unlike the fundamental theorem of factor analysis $U^{(k)}$ in general is no diagonal matrix. Therefore from the factor analytical point of view principal component analysis is only an approximative method.

factor analysis

principal component analysis

With k given the principal component analysis yields least square solutions for the k-factor model in the following sense:

$$\mathrm{tr}((Z - F^{(k)} L^{(k)T})^T (Z - F^{(k)} L^{(k)T})) = \mathrm{tr}(R - L^{(k)} L^{(k)T}) =$$

$$= \sum_{i=k+1}^{p} \lambda_i = \min_{\underset{(n\times k)}{F},\underset{(p\times k)}{L}} \{\mathrm{tr}((Z - F \cdot L^T)^T (Z - F \cdot L^T))\}$$

with orthonormal F

factor analysis

principal component analysis

With k given the principal component analysis yields least square solutions for the k-factor model in the following sense:

$$\mathrm{tr}((Z - F^{(k)} L^{(k)T})^T (Z - F^{(k)} L^{(k)T})) = \mathrm{tr}(R - L^{(k)} L^{(k)T}) =$$

$$= \sum_{i=k+1}^{p} \lambda_i = \min_{\underset{(n\times k)}{F}, \underset{(p\times k)}{L}} \{\mathrm{tr}((Z - F \cdot L^T)^T (Z - F \cdot L^T))\}$$

with orthonormal F

factor analysis

determination of k

1. All principal components F_j with an associated eigenvalue $\lambda_j \geq 1$ included, i.e. all components that explain more than the average variance of one variable ("eigenvalue-criterion"): $k = \max\{j \mid \lambda_j \geq 1\}$
2. We include principal components until a predetermined portion c of the total variance p is reached: $k = \min\{r \mid \lambda_1 + \cdots + \lambda_r \geq c \cdot p\}$
3. With independently normal distributed data the eigenvalues of the empirical covariance matrix have a smooth and uniformly descending graph if we sort and plot them. However with dependent data we mostly find a clear break. Before the break we have relatively big eigenvalues, afterwards the eigenvalues are small and about the same size.

factor analysis

determination of k

1. All principal components F_j with an associated eigenvalue $\lambda_j \geq 1$ included, i.e. all components that explain more than the average variance of one variable ("eigenvalue-criterion"): $k = \max\{j \mid \lambda_j \geq 1\}$
2. We include principal components until a predetermined portion c of the total variance p is reached: $k = \min\{r \mid \lambda_1 + \cdots + \lambda_r \geq c \cdot p\}$
3. With independently normal distributed data the eigenvalues of the empirical covariance matrix have a smooth and uniformly descending graph if we sort and plot them. However with dependent data we mostly find a clear break. Before the break we have relatively big eigenvalues, afterwards the eigenvalues are small and about the same size.

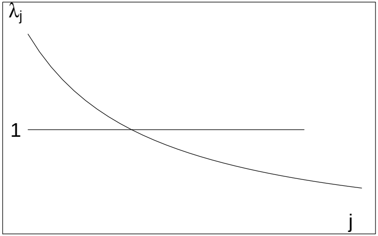
factor analysis

determination of k

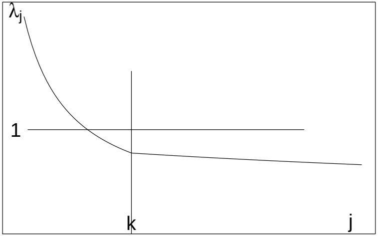
1. All principal components F_j with an associated eigenvalue $\lambda_j \geq 1$ included, i.e. all components that explain more than the average variance of one variable ("eigenvalue-criterion"): $k = \max\{j \mid \lambda_j \geq 1\}$
2. We include principal components until a predetermined portion c of the total variance p is reached: $k = \min\{r \mid \lambda_1 + \cdots + \lambda_r \geq c \cdot p\}$
3. With independently normal distributed data the eigenvalues of the empirical covariance matrix have a smooth and uniformly descending graph if we sort and plot them. However with dependent data we mostly find a clear break. Before the break we have relatively big eigenvalues, afterwards the eigenvalues are small and about the same size.

factor analysis

determination of k



independent variables



dependent variables

We assume the principal components affiliated with eigenvalues after the break as random and choose the position of the break as number of factors k (*Scree-test*, *scree-plot*).

factor analysis

determination of k

4. Significance inspection of the principal components according to Bartlett.
   Assumption: normal distribution of the data.

   We test whether the $p - k$ smallest eigenvalues of R differ from each other or not:

   $$H_0(k) : \lambda_{k+1} = \lambda_{k+2} = \cdots = \lambda_p$$

   We start with a small k_0 (stepwise extraction) and include principal components until H_0 is **not** rejected for the first time.

factor analysis

determination of k

4. Significance inspection of the principal components according to Bartlett.
   Assumption: normal distribution of the data.

   We test whether the $p - k$ smallest eigenvalues of R differ from each other or not:

   $$H_0(k): \lambda_{k+1} = \lambda_{k+2} = \cdots = \lambda_p$$

   We start with a small k_0 (stepwise extraction) and include principal components until H_0 is **not** rejected for the first time.

factor analysis

determination of k

4. Significance inspection of the principal components according to Bartlett.
   Assumption: normal distribution of the data.

   We test whether the $p-k$ smallest eigenvalues of R differ from each other or not:

   $$H_0(k): \lambda_{k+1} = \lambda_{k+2} = \cdots = \lambda_p$$

   We start with a small k_0 (stepwise extraction) and include principal components until H_0 is **not** rejected for the first time.

factor analysis

determination of k

4. Significance inspection of the principal components according to Bartlett.

   Assumption: normal distribution of the data.

   We test whether the $p - k$ smallest eigenvalues of R differ from each other or not:

   $$H_0(k): \lambda_{k+1} = \lambda_{k+2} = \cdots = \lambda_p$$

   We start with a small k_0 (stepwise extraction) and include principal components until H_0 is **not** rejected for the first time.

factor analysis

determination of k

The test statistic

$$U_k = (n-1)\cdot \underbrace{\left(-\ln|R| + \ln(\lambda_1 \cdot \ldots \cdot \lambda_k)\right) + (p-k)\cdot \ln \frac{p-\lambda_1-\cdots-\lambda_k}{p-k}}_{\ln \frac{\prod_{i=1}^{k}\lambda_i \cdot \left(\frac{p-\sum_{i=1}^{k}\lambda_i}{p-k}\right)^{p-k}}{\prod_{i=1}^{p}\lambda_i}}$$

is approximatively $\chi^2_{d_k}$-distributed (caution: not asymptotical) under H_0 with degrees of freedom $d_f = \frac{1}{2}(p-k+2)(p-k-1)$

Remark: under H_0 we have $|R| = \prod_{i=1}^{k}\lambda_i \cdot \left(\frac{p-\sum_{i=1}^{k}\lambda_i}{p-k}\right)^{p-k}$

factor analysis

determination of k

The test statistic

$$U_k = (n-1)\cdot \underbrace{\left(-\ln|R| + \ln(\lambda_1\cdot\ldots\cdot\lambda_k)\right) + (p-k)\cdot\ln\frac{p-\lambda_1-\cdots-\lambda_k}{p-k}}_{\ln\frac{\prod_{i=1}^k \lambda_i\cdot\left(\frac{p-\sum_{i=1}^k\lambda_i}{p-k}\right)^{p-k}}{\prod_{i=1}^p\lambda_i}}$$

is approximatively $\chi^2_{d_k}$-distributed (caution: not asymptotical) under H_0 with degrees of freedom $d_f = \frac{1}{2}(p-k+2)(p-k-1)$

<u>Remark:</u> under H_0 we have $|R| = \prod_{i=1}^k \lambda_i \cdot \left(\frac{p-\sum_{i=1}^k\lambda_i}{p-k}\right)^{p-k}$

factor analysis

principal factor analysis

We start with the factor analysis model and the fundamental theorem $\Sigma = L \cdot L^T + V$

We try to partition the empirical correlation matrix according to the fundamental theorem: $R = L \cdot L^T + V$ with $V = \mathrm{diag}\{v_1^2, \ldots, v_p^2\}$. The communalities $h_i^2 = 1 - v_i^2$ are also combined to a diagonal matrix $K = I - V$

The principal factor analysis is made up of two steps:

1. The specific variances v_i^2 are estimated, in fact via the communalities
2. We perform a principal axis transformation with the reduced empirical correlation matrix $R_h = R - V$:
   $\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_p\}$ with $\lambda_1 \geq \ldots \geq \lambda_p$ eigenvalues of R_h and $t_1, \ldots, t_p$ the corresponding normalized eigenvectors.

factor analysis

principal factor analysis

We start with the factor analysis model and the fundamental theorem $\Sigma = L \cdot L^T + V$

We try to partition the empirical correlation matrix according to the fundamental theorem: $R = L \cdot L^T + V$ with $V = \mathrm{diag}\{v_1^2, \ldots, v_p^2\}$. The communalities $h_i^2 = 1 - v_i^2$ are also combined to a diagonal matrix $K = I - V$

The principal factor analysis is made up of two steps:

1. The specific variances v_i^2 are estimated, in fact via the communalities
2. We perform a principal axis transformation with the reduced empirical correlation matrix $R_h = R - V$:
   $\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_p\}$ with $\lambda_1 \geq \ldots \geq \lambda_p$ eigenvalues of R_h and $t_1, \ldots, t_p$ the corresponding normalized eigenvectors.

factor analysis

principal factor analysis

We start with the factor analysis model and the fundamental theorem $\Sigma = L \cdot L^T + V$

We try to partition the empirical correlation matrix according to the fundamental theorem: $R = L \cdot L^T + V$ with $V = \text{diag}\{v_1^2, \ldots, v_p^2\}$. The communalities $h_i^2 = 1 - v_i^2$ are also combined to a diagonal matrix $K = I - V$

The principal factor analysis is made up of two steps:

1. The specific variances v_i^2 are estimated, in fact via the communalities
2. We perform a principal axis transformation with the reduced empirical correlation matrix $R_h = R - V$: $\Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_p\}$ with $\lambda_1 \geq \ldots \geq \lambda_p$ eigenvalues of R_h and $t_1, \ldots, t_p$ the corresponding normalized eigenvectors.

factor analysis

principal factor analysis

We start with the factor analysis model and the fundamental theorem $\Sigma = L \cdot L^T + V$

We try to partition the empirical correlation matrix according to the fundamental theorem: $R = L \cdot L^T + V$ with $V = \text{diag}\{v_1^2, \ldots, v_p^2\}$. The communalities $h_i^2 = 1 - v_i^2$ are also combined to a diagonal matrix $K = I - V$

The principal factor analysis is made up of two steps:

1. The specific variances v_i^2 are estimated, in fact via the communalities
2. We perform a principal axis transformation with the reduced empirical correlation matrix $R_h = R - V$:
   $\Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_p\}$ with $\lambda_1 \geq \ldots \geq \lambda_p$ eigenvalues of R_h and $t_1, \ldots, t_p$ the corresponding normalized eigenvectors.

factor analysis

principal factor analysis

With $T = (t_1 \cdots t_p)$ and $L = T \cdot \Lambda^{\frac{1}{2}}$ we have

$$L \cdot L^T = T \cdot \Lambda \cdot T^T = R_h$$

We get the factor values with

$$Z = F \cdot \Lambda^{\frac{1}{2}} T^T \qquad \Longleftrightarrow \qquad F = Z \cdot T \cdot \Lambda^{-\frac{1}{2}}$$

and therefore

$$F = Z \cdot \underbrace{T \cdot \Lambda^{-\frac{1}{2}} \overbrace{\Lambda^{-\frac{1}{2}} \overbrace{T^T T}^{I} \cdot \Lambda^{\frac{1}{2}}}^{I}}_{} = Z \cdot \underbrace{T \cdot \Lambda^{-\frac{1}{2}}\Lambda^{-\frac{1}{2}} T^T}_{R_h^{-1}} \underbrace{T \cdot \Lambda^{\frac{1}{2}}}_{L} = Z \cdot R_h^{-1} L$$

factor analysis

principal factor analysis

With $T = (t_1 \cdots t_p)$ and $L = T \cdot \Lambda^{\frac{1}{2}}$ we have

$$L \cdot L^T = T \cdot \Lambda \cdot T^T = R_h$$

We get the factor values with

$$Z = F \cdot \Lambda^{\frac{1}{2}} T^T \qquad \Longleftrightarrow \qquad F = Z \cdot T \cdot \Lambda^{-\frac{1}{2}}$$

and therefore

$$F = Z \cdot \underbrace{T \cdot \Lambda^{-\frac{1}{2}} \Lambda^{-\frac{1}{2}} T^T}_{R_h^{-1}} \underbrace{T \cdot \Lambda^{\frac{1}{2}}}_{L} = Z \cdot R_h^{-1} L$$

factor analysis

principal factor analysis

With $T = (t_1 \cdots t_p)$ and $L = T \cdot \Lambda^{\frac{1}{2}}$ we have

$$L \cdot L^T = T \cdot \Lambda \cdot T^T = R_h$$

We get the factor values with

$$Z = F \cdot \Lambda^{\frac{1}{2}} T^T \qquad \Longleftrightarrow \qquad F = Z \cdot T \cdot \Lambda^{-\frac{1}{2}}$$

and therefore

$$F = Z \cdot \underbrace{T \cdot \Lambda^{-\frac{1}{2}} \Lambda^{-\frac{1}{2}} T^T}_{R_h^{-1}} \underbrace{T \cdot \Lambda^{\frac{1}{2}}}_{L} = Z \cdot R_h^{-1} L$$

factor analysis

principal factor analysis

$F = (F_1 \cdots F_p)$. $F_1, \ldots, F_p$ are called **principal factors**, they explain the variances $\lambda_1, \ldots, \lambda_p$. Here we have: $\sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} h_i^2$

Caution: here we may have eigenvalues ≤ 0. Such eigenvalues yield meaningless factors. That´s why we limit ourself to the first k principal factors $F_1, \ldots, F_k$ as with principal component analysis.

$$Z = F^{(k)} L^{(k)T} + E^{(k)}$$

$$R = \underbrace{L^{(k)} L^{(k)T} + \overbrace{\sum_{i=k+1}^{p} l_i l_i^T}^{U^{(k)}}}_{R_h} + V$$

As with principal component analysis $U^{(k)}$ in general is not diagonal, i.e. also principal factor analysis is only an approximative method from the factor analysis point of view.

factor analysis

principal factor analysis

$F = (F_1 \cdots F_p)$. $F_1, \ldots, F_p$ are called **principal factors**, they explain the variances $\lambda_1, \ldots, \lambda_p$. Here we have: $\sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} h_i^2$

Caution: here we may have eigenvalues ≤ 0. Such eigenvalues yield meaningless factors. That´s why we limit ourself to the first k principal factors $F_1, \ldots, F_k$ as with principal component analysis.

$$Z = F^{(k)} L^{(k)T} + E^{(k)}$$

$$R = \underbrace{L^{(k)} L^{(k)T} + \overbrace{\sum_{i=k+1}^{p} l_i l_i^T + V}^{U^{(k)}}}_{R_h}$$

As with principal component analysis $U^{(k)}$ in general is not diagonal, i.e. also principal factor analysis is only an approximative method from the factor analysis point of view.

factor analysis

principal factor analysis

$F = (F_1 \cdots F_p)$. $F_1, \ldots, F_p$ are called **principal factors**, they explain the variances $\lambda_1, \ldots, \lambda_p$. Here we have: $\sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} h_i^2$

Caution: here we may have eigenvalues ≤ 0. Such eigenvalues yield meaningless factors. That´s why we limit ourself to the first k principal factors $F_1, \ldots, F_k$ as with principal component analysis.

$$Z = F^{(k)} L^{(k)T} + E^{(k)}$$

$$R = \underbrace{L^{(k)} L^{(k)T} + \overbrace{\sum_{i=k+1}^{p} l_i l_i^T}^{U^{(k)}}}_{R_h} \overbrace{+ V}$$

As with principal component analysis $U^{(k)}$ in general is not diagonal, i.e. also principal factor analysis is only an approximative method from the factor analysis point of view.

factor analysis

principal factor analysis

$F = (F_1 \cdots F_p)$. $F_1, \ldots, F_p$ are called **principal factors**, they explain the variances $\lambda_1, \ldots, \lambda_p$. Here we have: $\sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} h_i^2$

<u>Caution:</u> here we may have eigenvalues ≤ 0. Such eigenvalues yield meaningless factors. That´s why we limit ourself to the first k principal factors $F_1, \ldots, F_k$ as with principal component analysis.

$$Z = F^{(k)} L^{(k)T} + E^{(k)}$$

$$R = \underbrace{L^{(k)} L^{(k)T} + \overbrace{\sum_{i=k+1}^{p} l_i l_i^T + V}^{U^{(k)}}}_{R_h}$$

As with principal component analysis $U^{(k)}$ in general is not diagonal, i.e. also principal factor analysis is only an approximative method from the factor analysis point of view.

factor analysis

principal factor analysis

The two steps of principal factor analysis may be repeated iteratively:

We start e.g. with $V_0 = 0$ or $V_0 = I - K_0$ where in K_0 there are first estimates for the communalities. With this we get:

$$R - V_0 = L_1^{(k)} L_1^{(k)T} + \sum_{i=k+1}^{p} l_{1i} l_{1i}^T$$

$$\text{and } R = L_1^{(k)} L_1^{(k)T} + U_1^{(k)} \text{ respectively}$$

where $U_1^{(k)} = V_0 + \sum_{i=k+1}^{p} l_{1i} l_{1i}^T$

$K_1 = \mathrm{diag}(L_1^{(k)} L_1^{(k)T})$ is the new estimator for the diagonal matrix of communalities, $V_1 = \mathrm{diag}(U_1^{(k)})$ is the new estimator for the specific variances.

factor analysis

principal factor analysis

The two steps of principal factor analysis may be repeated iteratively:

We start e.g. with $V_0 = 0$ or $V_0 = I - K_0$ where in K_0 there are first estimates for the communalities. With this we get:

$$R - V_0 = L_1^{(k)} L_1^{(k)T} + \sum_{i=k+1}^{p} l_{1i} l_{1i}^T$$

$$\text{and } R = L_1^{(k)} L_1^{(k)T} + U_1^{(k)} \text{ respectively}$$

where $U_1^{(k)} = V_0 + \sum_{i=k+1}^{p} l_{1i} l_{1i}^T$

$K_1 = \text{diag}(L_1^{(k)} L_1^{(k)T})$ is the new estimator for the diagonal matrix of communalities, $V_1 = \text{diag}(U_1^{(k)})$ is the new estimator for the specific variances.

factor analysis

principal factor analysis

The two steps of principal factor analysis may be repeated iteratively:

We start e.g. with $V_0 = 0$ or $V_0 = I - K_0$ where in K_0 there are first estimates for the communalities. With this we get:

$$R - V_0 = L_1^{(k)} L_1^{(k)T} + \sum_{i=k+1}^{p} l_{1i} l_{1i}^T$$

$$\text{and } R = L_1^{(k)} L_1^{(k)T} + U_1^{(k)} \text{ respectively}$$

where $U_1^{(k)} = V_0 + \sum_{i=k+1}^{p} l_{1i} l_{1i}^T$

$K_1 = \text{diag}(L_1^{(k)} L_1^{(k)T})$ is the new estimator for the diagonal matrix of communalities, $V_1 = \text{diag}(U_1^{(k)})$ is the new estimator for the specific variances.

factor analysis

principal factor analysis

Now we perform a principal axis transformation of $R - V_1 = L_2^{(k)} L_2^{(k)T} + \sum_{i=k+1}^{p} l_{2i} l_{2i}^T$ and $R = L_2^{(k)} L_2^{(k)T} + U_2^{(k)}$ respectively etc.

After an appropriate termination condition is met, e.g. $\frac{\|V_n - V_{n-1}\|}{\|V_n\|} < \varepsilon$ or after a predetermined number of iteration steps, the iterations are stopped.

Then we have $R = L_n^{(k)} L_n^{(k)T} + U_n^{(k)},\ V_n = \text{diag}(U_n^{(k)})$ and the estimated factor values are $F_n^{(k)} = Z \cdot \underbrace{(R - V_n)^{-1}}_{R_h^{-1}} L_n^{(k)}$

Usually the number of factors remains constant during the iterations, however k may be redetermined after each iteration step.

factor analysis

principal factor analysis

Now we perform a principal axis transformation of $R - V_1 = L_2^{(k)} L_2^{(k)T} + \sum_{i=k+1}^{p} l_{2i} l_{2i}^T$ and $R = L_2^{(k)} L_2^{(k)T} + U_2^{(k)}$ respectively etc.

After an appropriate termination condition is met, e.g. $\frac{\|V_n - V_{n-1}\|}{\|V_n\|} < \varepsilon$ or after a predetermined number of iteration steps, the iterations are stopped.

Then we have $R = L_n^{(k)} L_n^{(k)T} + U_n^{(k)},\ V_n = \mathrm{diag}(U_n^{(k)})$ and the estimated factor values are $F_n^{(k)} = Z \cdot \underbrace{(R - V_n)^{-1}}_{R_h^{-1}} L_n^{(k)}$

Usually the number of factors remains constant during the iterations, however k may be redetermined after each iteration step.

factor analysis

principal factor analysis

Now we perform a principal axis transformation of $R - V_1 = L_2^{(k)} L_2^{(k)T} + \sum_{i=k+1}^{p} l_{2i} l_{2i}^T$ and $R = L_2^{(k)} L_2^{(k)T} + U_2^{(k)}$ respectively etc.

After an appropriate termination condition is met, e.g. $\frac{\|V_n - V_{n-1}\|}{\|V_n\|} < \varepsilon$ or after a predetermined number of iteration steps, the iterations are stopped.

Then we have $R = L_n^{(k)} L_n^{(k)T} + U_n^{(k)},\ V_n = \mathrm{diag}(U_n^{(k)})$ and the estimated factor values are $F_n^{(k)} = Z \cdot \underbrace{(R - V_n)^{-1}}_{R_h^{-1}} L_n^{(k)}$

Usually the number of factors remains constant during the iterations, however k may be redetermined after each iteration step.

factor analysis

principal factor analysis

Now we perform a principal axis transformation of $R - V_1 = L_2^{(k)} L_2^{(k)T} + \sum_{i=k+1}^{p} l_{2i} l_{2i}^T$ and $R = L_2^{(k)} L_2^{(k)T} + U_2^{(k)}$ respectively etc.

After an appropriate termination condition is met, e.g. $\frac{\|V_n - V_{n-1}\|}{\|V_n\|} < \varepsilon$ or after a predetermined number of iteration steps, the iterations are stopped.

Then we have $R = L_n^{(k)} L_n^{(k)T} + U_n^{(k)},\ V_n = \mathrm{diag}(U_n^{(k)})$ and the estimated factor values are $F_n^{(k)} = Z \cdot \underbrace{(R - V_n)^{-1}}_{R_h^{-1}} L_n^{(k)}$

Usually the number of factors remains constant during the iterations, however k may be redetermined after each iteration step.

factor analysis

principal factor analysis

Caution:

- The described iteration need not always converge.
- If the iterations converge the limit may be a meaningless solution, i.e. if $K_n \stackrel{n\to\infty}{\longrightarrow} K = \mathrm{diag}\{h_1^2, \ldots, h_p^2\}$ and $V_n \stackrel{n\to\infty}{\longrightarrow} V = \mathrm{diag}\{v_1^2, \ldots, v_p^2\}$ then possibly $h_i^2 > 1$ and $v_i^2 < 0$

If in the iteration process specific variances $v_i^2 < 0$ appear for the first time most statistics software packages stop the iterations with the last appropriate solution (**Heywood-case**).

Caution:

- The described iteration need not always converge.
- If the iterations converge the limit may be a meaningless solution, i.e. if $K_n \stackrel{n\to\infty}{\longrightarrow} K = \mathrm{diag}\{h_1^2, \ldots, h_p^2\}$ and $V_n \stackrel{n\to\infty}{\longrightarrow} V = \mathrm{diag}\{v_1^2, \ldots, v_p^2\}$ then possibly $h_i^2 > 1$ and $v_i^2 < 0$

If in the iteration process specific variances $v_i^2 < 0$ appear for the first time most statistics software packages stop the iterations with the last appropriate solution (**Heywood-case**).

factor analysis

principal factor analysis

Caution:

- The described iteration need not always converge.
- If the iterations converge the limit may be a meaningless solution, i.e. if $K_n \overset{n\to\infty}{\longrightarrow} K = \mathrm{diag}\{h_1^2, \ldots, h_p^2\}$ and $V_n \overset{n\to\infty}{\longrightarrow} V = \mathrm{diag}\{v_1^2, \ldots, v_p^2\}$ then possibly $h_i^2 > 1$ and $v_i^2 < 0$

If in the iteration process specific variances $v_i^2 < 0$ appear for the first time most statistics software packages stop the iterations with the last appropriate solution (**Heywood-case**).

Caution:

- The described iteration need not always converge.
- If the iterations converge the limit may be a meaningless solution, i.e. if $K_n \overset{n\to\infty}{\longrightarrow} K = \text{diag}\{h_1^2, \ldots, h_p^2\}$ and $V_n \overset{n\to\infty}{\longrightarrow} V = \text{diag}\{v_1^2, \ldots, v_p^2\}$ then possibly $h_i^2 > 1$ and $v_i^2 < 0$

If in the iteration process specific variances $v_i^2 < 0$ appear for the first time most statistics software packages stop the iterations with the last appropriate solution (**Heywood-case**).

factor analysis

estimation of communalities

With the above iterative procedure of the principal factor analysis the communalities are estimated with the model parameters k, L and V.

We may also estimate the communalities independently of k and L, and that with the **multiple correlation coefficient** ρ_j of the standardized data $Z = (z_1, \ldots, z_p)$

$$\rho_j = \frac{\sqrt{\sigma_{(j)} \Sigma_{22}^{-1} \sigma_{(j)}^T}}{\sigma_j} \qquad \text{where}$$

$$Cov(z_j, z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_p) = \begin{pmatrix} \sigma_j^2 & \sigma_{(j)} \\ \sigma_{(j)}^T & \Sigma_{22} \end{pmatrix}$$

Consider the order of the variates z!

factor analysis

estimation of communalities

With the above iterative procedure of the principal factor analysis the communalities are estimated with the model parameters k, L and V.

We may also estimate the communalities independently of k and L, and that with the **multiple correlation coefficient** ρ_j of the standardized data $Z = (z_1, \ldots, z_p)$

$$\rho_j = \frac{\sqrt{\sigma_{(j)} \Sigma_{22}^{-1} \sigma_{(j)}^T}}{\sigma_j} \qquad \text{where}$$

$$Cov(z_j, z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_p) = \begin{pmatrix} \sigma_j^2 & \sigma_{(j)} \\ \sigma_{(j)}^T & \Sigma_{22} \end{pmatrix}$$

Consider the order of the variates z!

factor analysis

estimation of communalities

With the above iterative procedure of the principal factor analysis the communalities are estimated with the model parameters k, L and V.

We may also estimate the communalities independently of k and L, and that with the **multiple correlation coefficient** ρ_j of the standardized data $Z = (z_1, \ldots, z_p)$

$$\rho_j = \frac{\sqrt{\sigma_{(j)} \Sigma_{22}^{-1} \sigma_{(j)}^T}}{\sigma_j} \qquad \text{where}$$

$$Cov(z_j, z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_p) = \begin{pmatrix} \sigma_j^2 & \sigma_{(j)} \\ \sigma_{(j)}^T & \Sigma_{22} \end{pmatrix}$$

Consider the order of the variates z!

factor analysis

estimation of communalities

ρ_j specifies the maximal correlation between z_j and all possible linear combinations of the remaining z_i $i \neq j$.

$\Longrightarrow$ the variance z_j has in common with the remaining z_i $i \neq j$ is at least ρ_j^2, i.e. $h_j^2 \geq \rho_j^2$

As estimator for ρ_j^2 we use

$$\widehat{\rho_j^2} = r_j^2 = 1 - \frac{1}{r^{jj}}$$

where r^{jj} is the jth diagonal element of R^{-1}. We have

$$r_j^2 \xrightarrow[]{\substack{p \to \infty \\ \frac{k}{p} \to 0}} h_j^2$$

factor analysis

estimation of communalities

ρ_j specifies the maximal correlation between z_j and all possible linear combinations of the remaining z_i $i \neq j$.

$\Longrightarrow$ the variance z_j has in common with the remaining z_i $i \neq j$ is at least ρ_j^2, i.e. $h_j^2 \geq \rho_j^2$

As estimator for ρ_j^2 we use

$$\widehat{\rho_j^2} = r_j^2 = 1 - \frac{1}{r^{jj}}$$

where r^{jj} is the jth diagonal element of R^{-1}. We have

$$r_j^2 \quad \xrightarrow[]{\substack{p \to \infty \\ \frac{k}{p} \to 0}} \quad h_j^2$$

factor analysis

estimation of communalities

ρ_j specifies the maximal correlation between z_j and all possible linear combinations of the remaining z_i $i \neq j$.

$\Longrightarrow$ the variance z_j has in common with the remaining z_i $i \neq j$ is at least ρ_j^2, i.e. $h_j^2 \geq \rho_j^2$

As estimator for ρ_j^2 we use

$$\widehat{\rho_j^2} = r_j^2 = 1 - \frac{1}{r^{jj}}$$

where r^{jj} is the jth diagonal element of R^{-1}. We have

$$r_j^2 \quad \overset{\substack{p \to \infty \\ \frac{k}{p} \to 0}}{\longrightarrow} \quad h_j^2$$