Multivariate Verfahren 2

factor analysis

Helmut Waldl

April 23th and 24th 2012

factor analysis

ML factor analysis

Assumption: $\mathbf{y} \sim \mathbf{N}(\mu, \Sigma)$

We need to assume a probability distribution to establish the likelihood function. The estimates $\hat{L}$ and $\hat{V}$ can also be computed without assuming a normal distribution. In this case we get so-called **quasi-ML-estimates** which are also consistent and asymptotically normal distributed (if $\mathbf{y}$ has first and second moments μ and Σ and if Σ is positive definite).

Given: $k < p$ and the empirical covariance matrix S

$$S = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T \qquad \bar{\mathbf{y}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{y}_i$$

Find: estimates for L and V for $\Sigma = L \cdot L^T + V$ where L is a $(p \times k)$-matrix and $V = \text{diag}(v_1^2, \ldots, v_p^2)$.
side condition: $L^T \cdot V^{-1} \cdot L$ diagonal

factor analysis

ML factor analysis

Assumption: $\mathbf{y} \sim \mathbf{N}(\boldsymbol{\mu}, \Sigma)$

We need to assume a probability distribution to establish the likelihood function. The estimates $\hat{L}$ and $\hat{V}$ can also be computed without assuming a normal distribution. In this case we get so-called **quasi-ML-estimates** which are also consistent and asymptotically normal distributed (if $\mathbf{y}$ has first and second moments $\boldsymbol{\mu}$ and Σ and if Σ is positive definite).

Given: $k < p$ and the empirical covariance matrix S

$$S = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T \qquad \bar{\mathbf{y}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{y}_i$$

Find: estimates for L and V for $\Sigma = L \cdot L^T + V$ where L is a $(p \times k)$-matrix and $V = \text{diag}(v_1^2, \ldots, v_p^2)$.
side condition: $L^T \cdot V^{-1} \cdot L$ diagonal

factor analysis

ML factor analysis

Assumption: $\mathbf{y} \sim \mathbf{N}(\boldsymbol{\mu}, \Sigma)$

We need to assume a probability distribution to establish the likelihood function. The estimates $\hat{L}$ and $\hat{V}$ can also be computed without assuming a normal distribution. In this case we get so-called **quasi-ML-estimates** which are also consistent and asymptotically normal distributed (if $\mathbf{y}$ has first and second moments $\boldsymbol{\mu}$ and Σ and if Σ is positive definite).

Given: $k < p$ and the empirical covariance matrix S

$$S = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T \qquad \bar{\mathbf{y}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{y}_i$$

Find: estimates for L and V for $\Sigma = L \cdot L^T + V$ where L is a $(p \times k)$-matrix and $V = \text{diag}(v_1^2, \ldots, v_p^2)$.
side condition: $L^T \cdot V^{-1} \cdot L$ diagonal

factor analysis

ML factor analysis

Assumption: $\mathbf{y} \sim \mathbf{N}(\mu, \Sigma)$

We need to assume a probability distribution to establish the likelihood function. The estimates $\hat{L}$ and $\hat{V}$ can also be computed without assuming a normal distribution. In this case we get so-called **quasi-ML-estimates** which are also consistent and asymptotically normal distributed (if $\mathbf{y}$ has first and second moments μ and Σ and if Σ is positive definite).

Given: $k < p$ and the empirical covariance matrix S

$$S = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T \qquad \bar{\mathbf{y}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{y}_i$$

Find: estimates for L and V for $\Sigma = L \cdot L^T + V$ where L is a $(p \times k)$-matrix and $V = \text{diag}(v_1^2, \ldots, v_p^2)$.
side condition: $L^T \cdot V^{-1} \cdot L$ diagonal

factor analysis

ML factor analysis

We determine the likelihood function of the empirical covariance matrix S and not of our data $\mathbf{y}$.

The reason for this approach is that we are basically interested in the explanation of the covariances Σ and not in the explanation of the data $\mathbf{y}$ or the expectations $\boldsymbol{\mu}$.

We take the unbiased estimate for Σ, S and maximize its likelihood varying Σ such that S is most likely for Σ.

$$(n-1) \cdot \underset{(p \times p)}{S} \sim \mathbf{W}(\Sigma, n-1) \qquad \ldots \text{Wishart distribution}$$

factor analysis

ML factor analysis

We determine the likelihood function of the empirical covariance matrix S and not of our data $\mathbf{y}$.

The reason for this approach is that we are basically interested in the explanation of the covariances Σ and not in the explanation of the data $\mathbf{y}$ or the expectations $\boldsymbol{\mu}$.

We take the unbiased estimate for Σ, S and maximize its likelihood varying Σ such that S is most likely for Σ.

$$(n-1) \cdot \underset{(p \times p)}{S} \sim \mathbf{W}(\Sigma, n-1) \qquad \ldots \text{Wishart distribution}$$

factor analysis

ML factor analysis

We determine the likelihood function of the empirical covariance matrix S and not of our data $\mathbf{y}$.

The reason for this approach is that we are basically interested in the explanation of the covariances Σ and not in the explanation of the data $\mathbf{y}$ or the expectations $\boldsymbol{\mu}$.

We take the unbiased estimate for Σ, S and maximize its likelihood varying Σ such that S is most likely for Σ.

$$(n-1) \cdot \underset{(p \times p)}{S} \sim \mathbf{W}(\Sigma, n-1) \qquad \ldots \text{Wishart distribution}$$

factor analysis

ML factor analysis

excursion: the Wishart-distribution is the multi-dimensional generalization of the χ^2-distribution:

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{iid}{\sim} \mathbf{N}(\mathbf{o}, \Sigma)$ with $(p \times p)$ covariance matrix Σ.
Let further $X = (\mathbf{x}_1 \ldots \mathbf{x}_n)^T$.
Then the $(p \times p)$-matrix W has a Wishart distribution:

$$W = X^T X = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T \sim \mathbf{W}(\Sigma, n)$$

The likelihood function of S is

$$L(\Sigma|S) = \text{const} \cdot |S|^{\frac{n-p-2}{2}} \cdot |\Sigma|^{-\frac{n-1}{2}} \cdot \exp\left(-\frac{n-1}{2} \cdot \text{tr}(S \cdot \Sigma^{-1})\right)$$

factor analysis

ML factor analysis

excursion: the Wishart-distribution is the multi-dimensional generalization of the χ^2-distribution:

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{iid}{\sim} \mathbf{N}(\mathbf{o}, \Sigma)$ with $(p \times p)$ covariance matrix Σ.
Let further $X = (\mathbf{x}_1 \ldots \mathbf{x}_n)^T$.
Then the $(p \times p)$-matrix W has a Wishart distribution:

$$W = X^T X = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T \sim \mathbf{W}(\Sigma, n)$$

The likelihood function of S is

$$L(\Sigma|S) = \text{const} \cdot |S|^{\frac{n-p-2}{2}} \cdot |\Sigma|^{-\frac{n-1}{2}} \cdot \exp\left(-\frac{n-1}{2} \cdot \text{tr}(S \cdot \Sigma^{-1})\right)$$

factor analysis

ML factor analysis

excursion: the Wishart-distribution is the multi-dimensional generalization of the χ^2-distribution:

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{iid}{\sim} \mathbf{N}(\mathbf{o}, \Sigma)$ with $(p \times p)$ covariance matrix Σ.
Let further $X = (\mathbf{x}_1 \ldots \mathbf{x}_n)^T$.
Then the $(p \times p)$-matrix W has a Wishart distribution:

$$W = X^T X = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T \sim \mathbf{W}(\Sigma, n)$$

The likelihood function of S is

$$L(\Sigma|S) = \text{const} \cdot |S|^{\frac{n-p-2}{2}} \cdot |\Sigma|^{-\frac{n-1}{2}} \cdot \exp\left(-\frac{n-1}{2} \cdot \text{tr}(S \cdot \Sigma^{-1})\right)$$

factor analysis

ML factor analysis

excursion: the Wishart-distribution is the multi-dimensional generalization of the χ^2-distribution:

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{iid}{\sim} \mathbf{N}(\mathbf{o}, \Sigma)$ with $(p \times p)$ covariance matrix Σ.
Let further $X = (\mathbf{x}_1 \ldots \mathbf{x}_n)^T$.
Then the $(p \times p)$-matrix W has a Wishart distribution:

$$W = X^T X = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T \sim \mathbf{W}(\Sigma, n)$$

The likelihood function of S is

$$L(\Sigma|S) = \text{const} \cdot |S|^{\frac{n-p-2}{2}} \cdot |\Sigma|^{-\frac{n-1}{2}} \cdot \exp\left(-\frac{n-1}{2} \cdot \text{tr}(S \cdot \Sigma^{-1})\right)$$

factor analysis

ML factor analysis

Maximizing the likelihood L varying Σ is equivalent to minimizing

$$f(L, V|S) = \ln|\Sigma| + \mathrm{tr}(S \cdot \Sigma^{-1}) \qquad \Sigma = L \cdot L^T + V$$

varying L and V. At the same time we have to consider the side condition: $L^T \cdot V^{-1} \cdot L$ diagonal.

i.e. $\frac{\partial f}{\partial L} = 0$, $\frac{\partial f}{\partial V} = 0$ plus side conditions $\longrightarrow$

conditional equations for $\hat{L}$ and $\hat{V}$:

$$\hat{L} = S \cdot \hat{\Sigma}^{-1} \cdot \hat{L} \qquad \hat{\Sigma} = \hat{L} \cdot \hat{L}^T + \hat{V}$$

$$\mathrm{diag}(\hat{\Sigma}) = \mathrm{diag}(S) \qquad \hat{L}^T \cdot \hat{V}^{-1} \cdot \hat{L} \quad \text{diagonal}$$

factor analysis

ML factor analysis

Maximizing the likelihood L varying Σ is equivalent to minimizing

$$f(L, V|S) = \ln|\Sigma| + \mathrm{tr}(S \cdot \Sigma^{-1}) \qquad \Sigma = L \cdot L^T + V$$

varying L and V. At the same time we have to consider the side condition: $L^T \cdot V^{-1} \cdot L$ diagonal.

i.e. $\frac{\partial f}{\partial L} = 0$, $\frac{\partial f}{\partial V} = 0$ plus side conditions $\longrightarrow$

conditional equations for $\hat{L}$ and $\hat{V}$:

$$\hat{L} = S \cdot \hat{\Sigma}^{-1} \cdot \hat{L} \qquad \hat{\Sigma} = \hat{L} \cdot \hat{L}^T + \hat{V}$$

$$\mathrm{diag}(\hat{\Sigma}) = \mathrm{diag}(S) \qquad \hat{L}^T \cdot \hat{V}^{-1} \cdot \hat{L} \quad \text{diagonal}$$

factor analysis

ML factor analysis

Maximizing the likelihood L varying Σ is equivalent to minimizing

$$f(L, V|S) = \ln|\Sigma| + \mathrm{tr}(S \cdot \Sigma^{-1}) \qquad \Sigma = L \cdot L^T + V$$

varying L and V. At the same time we have to consider the side condition: $L^T \cdot V^{-1} \cdot L$ diagonal.

i.e. $\frac{\partial f}{\partial L} = 0$, $\frac{\partial f}{\partial V} = 0$ plus side conditions $\longrightarrow$

conditional equations for $\hat{L}$ and $\hat{V}$:

$$\hat{L} = S \cdot \hat{\Sigma}^{-1} \cdot \hat{L} \qquad \hat{\Sigma} = \hat{L} \cdot \hat{L}^T + \hat{V}$$

$$\mathrm{diag}(\hat{\Sigma}) = \mathrm{diag}(S) \qquad \hat{L}^T \cdot \hat{V}^{-1} \cdot \hat{L} \quad \text{diagonal}$$

factor analysis

ML factor analysis

Maximizing the likelihood L varying Σ is equivalent to minimizing

$$f(L, V|S) = \ln|\Sigma| + \mathrm{tr}(S \cdot \Sigma^{-1}) \qquad \Sigma = L \cdot L^T + V$$

varying L and V. At the same time we have to consider the side condition: $L^T \cdot V^{-1} \cdot L$ diagonal.

i.e. $\frac{\partial f}{\partial L} = 0$, $\frac{\partial f}{\partial V} = 0$ plus side conditions $\longrightarrow$

conditional equations for $\hat{L}$ and $\hat{V}$:

$$\hat{L} = S \cdot \hat{\Sigma}^{-1} \cdot \hat{L} \qquad \hat{\Sigma} = \hat{L} \cdot \hat{L}^T + \hat{V}$$

$$\mathrm{diag}(\hat{\Sigma}) = \mathrm{diag}(S) \qquad \hat{L}^T \cdot \hat{V}^{-1} \cdot \hat{L} \quad \text{diagonal}$$

factor analysis

ML factor analysis

In practice: $f(L, V|S)$ is minimized iterative in two steps:

1. We fix $V > 0$ and determine $\min_L f(L, V|S) = f_0(V)$
2. We minimize $f_0(V)$ for V

We cannot discus the numerical problems minimizing $f(L, V|S)$. Optimization is done most efficiently with quasi-Newton methods.

Frequently only the solution $\hat{L}$ is specified. $\hat{V}$ can then simply be computed: $\hat{V} = \text{diag}(S - \hat{L} \cdot \hat{L}^T)$

factor analysis

ML factor analysis

In practice: $f(L, V|S)$ is minimized iterative in two steps:

1. We fix $V > 0$ and determine $\min_L f(L, V|S) = f_0(V)$
2. We minimize $f_0(V)$ for V

We cannot discus the numerical problems minimizing $f(L, V|S)$. Optimization is done most efficiently with quasi-Newton methods.

Frequently only the solution $\hat{L}$ is specified. $\hat{V}$ can then simply be computed: $\hat{V} = \text{diag}(S - \hat{L} \cdot \hat{L}^T)$

factor analysis

ML factor analysis

In practice: $f(L, V|S)$ is minimized iterative in two steps:

1. We fix $V > 0$ and determine $\min_L f(L, V|S) = f_0(V)$
2. We minimize $f_0(V)$ for V

We cannot discus the numerical problems minimizing $f(L, V|S)$. Optimization is done most efficiently with quasi-Newton methods.

Frequently only the solution $\hat{L}$ is specified. $\hat{V}$ can then simply be computed: $\hat{V} = \text{diag}(S - \hat{L} \cdot \hat{L}^T)$

factor analysis

ML factor analysis

In practice: $f(L, V|S)$ is minimized iterative in two steps:

1. We fix $V > 0$ and determine $\min_L f(L, V|S) = f_0(V)$
2. We minimize $f_0(V)$ for V

We cannot discus the numerical problems minimizing $f(L, V|S)$. Optimization is done most efficiently with quasi-Newton methods.

Frequently only the solution $\hat{L}$ is specified. $\hat{V}$ can then simply be computed: $\hat{V} = \text{diag}(S - \hat{L} \cdot \hat{L}^T)$

factor analysis

ML factor analysis

In practice: $f(L, V|S)$ is minimized iterative in two steps:

1. We fix $V > 0$ and determine $\min_L f(L, V|S) = f_0(V)$
2. We minimize $f_0(V)$ for V

We cannot discus the numerical problems minimizing $f(L, V|S)$.
Optimization is done most efficiently with quasi-Newton methods.

Frequently only the solution $\hat{L}$ is specified. $\hat{V}$ can then simply be computed: $\hat{V} = \text{diag}(S - \hat{L} \cdot \hat{L}^T)$

factor analysis

ML factor analysis

In practice: $f(L, V|S)$ is minimized iterative in two steps:

1. We fix $V > 0$ and determine $\min_L f(L, V|S) = f_0(V)$
2. We minimize $f_0(V)$ for V

We cannot discus the numerical problems minimizing $f(L, V|S)$. Optimization is done most efficiently with quasi-Newton methods.

Frequently only the solution $\hat{L}$ is specified. $\hat{V}$ can then simply be computed: $\hat{V} = \text{diag}(S - \hat{L} \cdot \hat{L}^T)$

factor analysis

ML factor analysis

Uniqueness of $\hat{V}$:

If $f_0(V)$ has several relative minima in the admissible region $0 < v_i^2 \leq s_{ii}$ then the numerical computation possibly yields different estimates for different starting values of V.

This indicates that V is not unique in the decomposition $\Sigma = L \cdot L^T + V$ with k factors.

From theory we know:
If k is the true number of factors and V is unique then the probability that $f_0(V)$ has just one minimum converges to 1 as $n \to \infty$.

factor analysis

ML factor analysis

Uniqueness of $\hat{V}$:

If $f_0(V)$ has several relative minima in the admissible region $0 < v_i^2 \leq s_{ii}$ then the numerical computation possibly yields different estimates for different starting values of V.

This indicates that V is not unique in the decomposition $\Sigma = L \cdot L^T + V$ with k factors.

From theory we know:

If k is the true number of factors and V is unique then the probability that $f_0(V)$ has just one minimum converges to 1 as $n \to \infty$.

factor analysis

ML factor analysis

Uniqueness of $\hat{V}$:

If $f_0(V)$ has several relative minima in the admissible region $0 < v_i^2 \leq s_{ii}$ then the numerical computation possibly yields different estimates for different starting values of V.

This indicates that V is not unique in the decomposition $\Sigma = L \cdot L^T + V$ with k factors.

From theory we know:
If k is the true number of factors and V is unique then the probability that $f_0(V)$ has just one minimum converges to 1 as $n \to \infty$.

factor analysis

determination of k, testing the model fitting

Assumption: normal distribution of $\mathbf{y}$

Given: k and the corresponding ML-estimates for Σ ($\hat{\Sigma} = \hat{L} \cdot \hat{L}^T + \hat{V}$)

We test $H_0 : \Sigma = L \cdot L^T + V$, $L \ldots (p \times k)$, $V > 0$ against

$H_1 : \Sigma \ldots (p \times p)$ arbitrary, positive definite.

We may test these hypotheses with a likelihood ratio test.

Under H_0 the likelihood function $L(\Sigma|S)$ is maximized by $\hat{\Sigma}$, i.e.

$$L_0 := \max_{\Sigma \in H_0} L(\Sigma|S) = L(\hat{\Sigma}|S)$$

Under H_1 we take the likelihood function $L_1 := L(S|S)$.

factor analysis

determination of k, testing the model fitting

Assumption: normal distribution of $\mathbf{y}$

Given: k and the corresponding ML-estimates for Σ ($\hat{\Sigma} = \hat{L} \cdot \hat{L}^T + \hat{V}$)

We test $H_0 : \Sigma = L \cdot L^T + V$, $L \ldots (p \times k)$, $V > 0$ against

$H_1 : \Sigma \ldots (p \times p)$ arbitrary, positive definite.

We may test these hypotheses with a likelihood ratio test.

Under H_0 the likelihood function $L(\Sigma|S)$ is maximized by $\hat{\Sigma}$, i.e.

$$L_0 := \max_{\Sigma \in H_0} L(\Sigma|S) = L(\hat{\Sigma}|S)$$

Under H_1 we take the likelihood function $L_1 := L(S|S)$.

factor analysis

determination of k, testing the model fitting

Assumption: normal distribution of $\mathbf{y}$

Given: k and the corresponding ML-estimates for Σ ($\hat{\Sigma} = \hat{L} \cdot \hat{L}^T + \hat{V}$)

We test $H_0 : \Sigma = L \cdot L^T + V$, $L \ldots (p \times k)$, $V > 0$ against

$H_1 : \Sigma \ldots (p \times p)$ arbitrary, positive definite.

We may test these hypotheses with a likelihood ratio test.

Under H_0 the likelihood function $L(\Sigma|S)$ is maximized by $\hat{\Sigma}$, i.e.

$$L_0 := \max_{\Sigma \in H_0} L(\Sigma|S) = L(\hat{\Sigma}|S)$$

Under H_1 we take the likelihood function $L_1 := L(S|S)$.

factor analysis

determination of k, testing the model fitting

Assumption: normal distribution of $\mathbf{y}$

Given: k and the corresponding ML-estimates for Σ ($\hat{\Sigma} = \hat{L} \cdot \hat{L}^T + \hat{V}$)

We test $H_0 : \Sigma = L \cdot L^T + V$, $L \ldots (p \times k)$, $V > 0$ against

$H_1 : \Sigma \ldots (p \times p)$ arbitrary, positive definite.

We may test these hypotheses with a likelihood ratio test.

Under H_0 the likelihood function $L(\Sigma|S)$ is maximized by $\hat{\Sigma}$, i.e.

$$L_0 := \max_{\Sigma \in H_0} L(\Sigma|S) = L(\hat{\Sigma}|S)$$

Under H_1 we take the likelihood function $L_1 := L(S|S)$.

factor analysis

determination of k, testing the model fitting

Assumption: normal distribution of $\mathbf{y}$

Given: k and the corresponding ML-estimates for Σ ($\hat{\Sigma} = \hat{L} \cdot \hat{L}^T + \hat{V}$)

We test $H_0 : \Sigma = L \cdot L^T + V$, $L \ldots (p \times k)$, $V > 0$ against

$H_1 : \Sigma \ldots (p \times p)$ arbitrary, positive definite.

We may test these hypotheses with a likelihood ratio test.

Under H_0 the likelihood function $L(\Sigma|S)$ is maximized by $\hat{\Sigma}$, i.e.

$$L_0 := \max_{\Sigma \in H_0} L(\Sigma|S) = L(\hat{\Sigma}|S)$$

Under H_1 we take the likelihood function $L_1 := L(S|S)$.

factor analysis

determination of k, testing the model fitting

Assumption: normal distribution of $\mathbf{y}$

Given: k and the corresponding ML-estimates for Σ ($\hat{\Sigma} = \hat{L} \cdot \hat{L}^T + \hat{V}$)

We test $H_0 : \Sigma = L \cdot L^T + V$, $L \ldots (p \times k)$, $V > 0$ against

$H_1 : \Sigma \ldots (p \times p)$ arbitrary, positive definite.

We may test these hypotheses with a likelihood ratio test.

Under H_0 the likelihood function $L(\Sigma|S)$ is maximized by $\hat{\Sigma}$, i.e.

$$L_0 := \max_{\Sigma \in H_0} L(\Sigma|S) = L(\hat{\Sigma}|S)$$

Under H_1 we take the likelihood function $L_1 := L(S|S)$.

factor analysis

determination of k, testing the model fitting

We use the test statistic

$$U_k = -2 \cdot \ln \frac{L_0}{L_1} = (n-1) \cdot (\ln |\hat{\Sigma}| - \ln |S|)$$

The exact distribution of U_k is unknown, for big n we know: $U_k \overset{a}{\sim} \chi^2_{d_k}$ with $d_k = \frac{1}{2}((p-k)^2 - (p+k))$ degrees of freedom.

Rejecting H_0 we may imply that we need at least $k+1$ factors.

factor analysis

determination of k, testing the model fitting

We use the test statistic

$$U_k = -2 \cdot \ln \frac{L_0}{L_1} = (n-1) \cdot (\ln |\hat{\Sigma}| - \ln |S|)$$

The exact distribution of U_k is unknown, for big n we know: $U_k \overset{a}{\sim} \chi^2_{d_k}$ with $d_k = \frac{1}{2}((p-k)^2 - (p+k))$ degrees of freedom.

Rejecting H_0 we may imply that we need at least $k+1$ factors.

factor analysis

determination of k, testing the model fitting

We use the test statistic

$$U_k = -2 \cdot \ln \frac{L_0}{L_1} = (n-1) \cdot (\ln |\hat{\Sigma}| - \ln |S|)$$

The exact distribution of U_k is unknown, for big n we know: $U_k \overset{a}{\sim} \chi^2_{d_k}$ with $d_k = \frac{1}{2}((p-k)^2 - (p+k))$ degrees of freedom.

Rejecting H_0 we may imply that we need at least $k+1$ factors.

factor analysis

determination of k, testing the model fitting

We use the test statistic

$$U_k = -2 \cdot \ln \frac{L_0}{L_1} = (n-1) \cdot (\ln |\hat{\Sigma}| - \ln |S|)$$

The exact distribution of U_k is unknown, for big n we know: $U_k \overset{a}{\sim} \chi^2_{d_k}$ with $d_k = \frac{1}{2}((p-k)^2 - (p+k))$ degrees of freedom.

Rejecting H_0 we may imply that we need at least $k+1$ factors.

factor analysis

determination of k, testing the model fitting

According to **Bartlett** the approximation to χ^2 distribution can be improved with $n' = n - \frac{2p+5}{6} - \frac{2}{3}k$ substituting n in U_k.

For $k = 0$ we get **Bartlett's test**: here we test whether there are significant dependencies between the y_i at all.

$$H_0 : \Sigma = V = \mathrm{diag}(v_1^2, \ldots, v_p^2)$$

test statistic: $U_0 = -(n - 1 - \frac{2p+5}{6}) \ln |R|$ where R is the empirical correlation matrix.

factor analysis

determination of k, testing the model fitting

According to **Bartlett** the approximation to χ^2 distribution can be improved with $n' = n - \frac{2p+5}{6} - \frac{2}{3}k$ substituting n in U_k.

For $k = 0$ we get **Bartlett's test**: here we test whether there are significant dependencies between the y_i at all.

$$H_0 : \Sigma = V = \mathrm{diag}(v_1^2, \ldots, v_p^2)$$

test statistic: $U_0 = -(n - 1 - \frac{2p+5}{6}) \ln |R|$ where R is the empirical correlation matrix.

factor analysis

determination of k, testing the model fitting

According to **Bartlett** the approximation to χ^2 distribution can be improved with $n' = n - \frac{2p+5}{6} - \frac{2}{3}k$ substituting n in U_k.

For $k = 0$ we get **Bartlett's test**: here we test whether there are significant dependencies between the y_i at all.

$$H_0 : \Sigma = V = \text{diag}(v_1^2, \ldots, v_p^2)$$

test statistic: $U_0 = -(n - 1 - \frac{2p+5}{6}) \ln |R|$ where R is the empirical correlation matrix.

factor analysis

determination of k, testing the model fitting

According to **Bartlett** the approximation to χ^2 distribution can be improved with $n' = n - \frac{2p+5}{6} - \frac{2}{3}k$ substituting n in U_k.

For $k = 0$ we get **Bartlett's test**: here we test whether there are significant dependencies between the y_i at all.

$$H_0 : \Sigma = V = \text{diag}(v_1^2, \ldots, v_p^2)$$

test statistic: $U_0 = -(n - 1 - \frac{2p+5}{6}) \ln |R|$ where R is the empirical correlation matrix.

factor analysis

determination of the number of faktors k

We may express the above null hypothesis also as

$$H_0 = H(k) : \text{k factors suffice to explain the data}$$

If we want to determine k with the above test we have to start with a small k_0 (possibly $k_0 = 0$ and execute the test. If $H(k_0)$ is rejected we test $H(k_0 + 1)$ etc.
The procedure stops if the null hypothesis cannot be rejected for the first time. This null hypothesis is $H(\hat{k})$, the hypothesis associated with the estimated number of factors $\hat{k}$.

Remark: in this procedure we have to re-estimate the parameters L and V in each step.

Remark: the individual tests of this composite testing procedure are **not** independent. Hence the only thing we may say about error probabilities is

$$P(\text{number of factors} \geq \hat{k}) \geq 1 - \alpha$$

factor analysis

determination of the number of faktors k

We may express the above null hypothesis also as

$$H_0 = H(k) : \text{k factors suffice to explain the data}$$

If we want to determine k with the above test we have to start with a small k_0 (possibly $k_0 = 0$ and execute the test. If $H(k_0)$ is rejected we test $H(k_0 + 1)$ etc.
The procedure stops if the null hypothesis cannot be rejected for the first time. This null hypothesis is $H(\hat{k})$, the hypothesis associated with the estimated number of factors $\hat{k}$.

Remark: in this procedure we have to re-estimate the parameters L and V in each step.

Remark: the individual tests of this composite testing procedure are **not** independent. Hence the only thing we may say about error probabilities is

$$P(\text{number of factors} \geq \hat{k}) \geq 1 - \alpha$$

factor analysis

determination of the number of faktors k

We may express the above null hypothesis also as

$$H_0 = H(k) : \text{k factors suffice to explain the data}$$

If we want to determine k with the above test we have to start with a small k_0 (possibly $k_0 = 0$ and execute the test. If $H(k_0)$ is rejected we test $H(k_0 + 1)$ etc.
The procedure stops if the null hypothesis cannot be rejected for the first time. This null hypothesis is $H(\hat{k})$, the hypothesis associated with the estimated number of factors $\hat{k}$.

Remark: in this procedure we have to re-estimate the parameters L and V in each step.

Remark: the individual tests of this composite testing procedure are **not** independent. Hence the only thing we may say about error probabilities is

$$P(\text{number of factors} \geq \hat{k}) \geq 1 - \alpha$$

factor analysis

determination of the number of faktors k

We may express the above null hypothesis also as

$$H_0 = H(k) : \text{k factors suffice to explain the data}$$

If we want to determine k with the above test we have to start with a small k_0 (possibly $k_0 = 0$ and execute the test. If $H(k_0)$ is rejected we test $H(k_0 + 1)$ etc.
The procedure stops if the null hypothesis cannot be rejected for the first time. This null hypothesis is $H(\hat{k})$, the hypothesis associated with the estimated number of factors $\hat{k}$.

Remark: in this procedure we have to re-estimate the parameters L and V in each step.

Remark: the individual tests of this composite testing procedure are **not** independent. Hence the only thing we may say about error probabilities is

$$P(\text{number of factors} \geq \hat{k}) \geq 1 - \alpha$$

factor analysis

determination of the number of faktors k

We may express the above null hypothesis also as

$$H_0 = H(k) : \text{k factors suffice to explain the data}$$

If we want to determine k with the above test we have to start with a small k_0 (possibly $k_0 = 0$ and execute the test. If $H(k_0)$ is rejected we test $H(k_0 + 1)$ etc.
The procedure stops if the null hypothesis cannot be rejected for the first time. This null hypothesis is $H(\hat{k})$, the hypothesis associated with the estimated number of factors $\hat{k}$.

Remark: in this procedure we have to re-estimate the parameters L and V in each step.

Remark: the individual tests of this composite testing procedure are **not** independent. Hence the only thing we may say about error probabilities is

$$P(\text{number of factors} \geq \hat{k}) \geq 1 - \alpha$$

factor analysis

data standardization, factor analysis with R instead of S

The factor analysis model is scale-invariant in the following sense:
If we rescale our data: $\tilde{\mathbf{y}} = D \cdot \mathbf{y}$ with $D = \text{diag}(d_i, \ldots, d_p)$, $d_i > 0$
then the loading matrix is rescaled in same way and the specific variances are rescaled quadratic, the factors remain constant:

$$\tilde{\mathbf{y}} - \tilde{\mu} = D(\mathbf{y} - \mu) = \underbrace{D \cdot L}_{\tilde{L}} \cdot \mathbf{f} + \underbrace{D \cdot \mathbf{e}}_{\tilde{\mathbf{e}}} = \tilde{L} \cdot \mathbf{f} + \tilde{\mathbf{e}}$$

$$\tilde{\Sigma} = D \cdot \Sigma \cdot D = D \cdot L \cdot L^T \cdot D^T + \underbrace{D \cdot V \cdot D}_{\tilde{V}} = \tilde{L} \cdot \tilde{L}^T + \tilde{V}$$

factor analysis

data standardization, factor analysis with R instead of S

The factor analysis model is scale-invariant in the following sense:
If we rescale our data: $\tilde{\mathbf{y}} = D \cdot \mathbf{y}$ with $D = \text{diag}(d_i, \ldots, d_p)$, $d_i > 0$
then the loading matrix is rescaled in same way and the specific variances are rescaled quadratic, the factors remain constant:

$$\tilde{\mathbf{y}} - \tilde{\mu} = D(\mathbf{y} - \mu) = \underbrace{D \cdot L}_{\tilde{L}} \cdot \mathbf{f} + \underbrace{D \cdot \mathbf{e}}_{\tilde{\mathbf{e}}} = \tilde{L} \cdot \mathbf{f} + \tilde{\mathbf{e}}$$

$$\tilde{\Sigma} = D \cdot \Sigma \cdot D = D \cdot L \cdot L^T \cdot D^T + \underbrace{D \cdot V \cdot D}_{\tilde{V}} = \tilde{L} \cdot \tilde{L}^T + \tilde{V}$$

factor analysis

data standardization, factor analysis with R instead of S

The factor analysis model is scale-invariant in the following sense:
If we rescale our data: $\tilde{\mathbf{y}} = D \cdot \mathbf{y}$ with $D = \text{diag}(d_i, \ldots, d_p)$, $d_i > 0$
then the loading matrix is rescaled in same way and the specific variances are rescaled quadratic, the factors remain constant:

$$\tilde{\mathbf{y}} - \tilde{\boldsymbol{\mu}} = D(\mathbf{y} - \boldsymbol{\mu}) = \underbrace{D \cdot L}_{\tilde{L}} \cdot \mathbf{f} + \underbrace{D \cdot \mathbf{e}}_{\tilde{\mathbf{e}}} = \tilde{L} \cdot \mathbf{f} + \tilde{\mathbf{e}}$$

$$\tilde{\Sigma} = D \cdot \Sigma \cdot D = D \cdot L \cdot L^T \cdot D^T + \underbrace{D \cdot V \cdot D}_{\tilde{V}} = \tilde{L} \cdot \tilde{L}^T + \tilde{V}$$

factor analysis

data standardization, factor analysis with R instead of S

The factor analysis model is scale-invariant in the following sense:
If we rescale our data: $\tilde{\mathbf{y}} = D \cdot \mathbf{y}$ with $D = \text{diag}(d_i, \ldots, d_p)$, $d_i > 0$
then the loading matrix is rescaled in same way and the specific variances are rescaled quadratic, the factors remain constant:

$$\tilde{\mathbf{y}} - \tilde{\boldsymbol{\mu}} = D(\mathbf{y} - \boldsymbol{\mu}) = \underbrace{D \cdot L}_{\tilde{L}} \cdot \mathbf{f} + \underbrace{D \cdot \mathbf{e}}_{\tilde{\mathbf{e}}} = \tilde{L} \cdot \mathbf{f} + \tilde{\mathbf{e}}$$

$$\tilde{\Sigma} = D \cdot \Sigma \cdot D = D \cdot L \cdot L^T \cdot D^T + \underbrace{D \cdot V \cdot D}_{\tilde{V}} = \tilde{L} \cdot \tilde{L}^T + \tilde{V}$$

factor analysis

data standardization, factor analysis with R instead of S

The ML estimates of the factor analysis model are scale-invariant in the following sense:

$$f(L, V|S) = \ln(\Sigma) + \mathrm{tr}(S \cdot \Sigma^{-1}) \to \text{MIN!} \quad \Longleftrightarrow$$

$$\Longleftrightarrow \quad f(\tilde{L}, \tilde{V}|\tilde{S}) = \ln(\tilde{\Sigma}) + \mathrm{tr}(\tilde{S} \cdot \tilde{\Sigma}^{-1}) \to \text{MIN!}$$

Because of this scale-invariance a factor analysis of the rescaled empirical covariance matrix $\tilde{S} = D \cdot S \cdot D$ yield essentially the same estimates as a factor analysis of S (just convert the results).

factor analysis

data standardization, factor analysis with R instead of S

The ML estimates of the factor analysis model are scale-invariant in the following sense:

$$f(L, V|S) = \ln(\Sigma) + \mathrm{tr}(S \cdot \Sigma^{-1}) \to \text{MIN!} \quad \Longleftrightarrow$$

$$\Longleftrightarrow \quad f(\tilde{L}, \tilde{V}|\tilde{S}) = \ln(\tilde{\Sigma}) + \mathrm{tr}(\tilde{S} \cdot \tilde{\Sigma}^{-1}) \to \text{MIN!}$$

Because of this scale-invariance a factor analysis of the rescaled empirical covariance matrix $\tilde{S} = D \cdot S \cdot D$ yield essentially the same estimates as a factor analysis of S (just convert the results).

factor analysis

data standardization, factor analysis with R instead of S

The ML estimates of the factor analysis model are scale-invariant in the following sense:

$$f(L, V|S) = \ln(\Sigma) + \mathrm{tr}(S \cdot \Sigma^{-1}) \to \mathrm{MIN!} \quad \Longleftrightarrow$$

$$\Longleftrightarrow \quad f(\tilde{L}, \tilde{V}|\tilde{S}) = \ln(\tilde{\Sigma}) + \mathrm{tr}(\tilde{S} \cdot \tilde{\Sigma}^{-1}) \to \mathrm{MIN!}$$

Because of this scale-invariance a factor analysis of the rescaled empirical covariance matrix $\tilde{S} = D \cdot S \cdot D$ yield essentially the same estimates as a factor analysis of S (just convert the results).

factor analysis

data standardization, factor analysis with R instead of S

The ML estimates of the factor analysis model are scale-invariant in the following sense:

$$f(L, V|S) = \ln(\Sigma) + \mathrm{tr}(S \cdot \Sigma^{-1}) \to \mathrm{MIN!} \quad \Longleftrightarrow$$

$$\Longleftrightarrow \quad f(\tilde{L}, \tilde{V}|\tilde{S}) = \ln(\tilde{\Sigma}) + \mathrm{tr}(\tilde{S} \cdot \tilde{\Sigma}^{-1}) \to \mathrm{MIN!}$$

Because of this scale-invariance a factor analysis of the rescaled empirical covariance matrix $\tilde{S} = D \cdot S \cdot D$ yield essentially the same estimates as a factor analysis of S (just convert the results).

factor analysis

data standardization, factor analysis with R instead of S

Thus we may as well analyze standardized data, the empirical correlation matrix R then supersedes the empirical covariance matrix S.

This results in loadings bounded on $[-1; 1]$ and communalities and specific variances bounded on $[0; 1]$.

In practice factor analysis is mostly done with standardized data, R then is not interpreted as correlation matrix but as covariance matrix of the standardized data

this causes just one problem with ML factor analysis: $(n-1) \cdot R$ is not Wishart distributed.

factor analysis

data standardization, factor analysis with R instead of S

Thus we may as well analyze standardized data, the empirical correlation matrix R then supersedes the empirical covariance matrix S.

This results in loadings bounded on $[-1; 1]$ and communalities and specific variances bounded on $[0; 1]$.

In practice factor analysis is mostly done with standardized data, R then is not interpreted as correlation matrix but as covariance matrix of the standardized data

this causes just one problem with ML factor analysis: $(n-1) \cdot R$ is not Wishart distributed.

factor analysis

data standardization, factor analysis with R instead of S

Thus we may as well analyze standardized data, the empirical correlation matrix R then supersedes the empirical covariance matrix S.

This results in loadings bounded on $[-1; 1]$ and communalities and specific variances bounded on $[0; 1]$.

In practice factor analysis is mostly done with standardized data, R then is not interpreted as correlation matrix but as covariance matrix of the standardized data

this causes just one problem with ML factor analysis: $(n-1) \cdot R$ is not Wishart distributed.

factor analysis

data standardization, factor analysis with R instead of S

Thus we may as well analyze standardized data, the empirical correlation matrix R then supersedes the empirical covariance matrix S.

This results in loadings bounded on $[-1; 1]$ and communalities and specific variances bounded on $[0; 1]$.

In practice factor analysis is mostly done with standardized data, R then is not interpreted as correlation matrix but as covariance matrix of the standardized data

this causes just one problem with ML factor analysis: $(n-1) \cdot R$ is not Wishart distributed.

factor analysis

data standardization, empirical correlation matrix

From the original data $\mathbf{Y} = (y_{ij})$, $i = 1, \ldots, n$, $j = 1, \ldots, p$ we get

$$z_{ij} = \frac{y_{ij} - \bar{y}_j}{\sqrt{n-1} \cdot s_j} \qquad \bar{y}_j = \frac{1}{n} \sum_{i=1}^{n} y_{ij} \qquad s_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_{ij} - \bar{y}_j)^2$$

z_{ij} are the "standardized data" (note: division by $\sqrt{n-1}$!), we combine them to the data matrix $\mathbf{Z} = (z_{ij})$.

Let $R = (r_{ij})$ be the empirical correlation matrix

$$r_{ij} = \frac{\sum_k (y_{ki} - \bar{y}_i)(y_{kj} - \bar{y}_j)}{\sqrt{\sum_k (y_{ki} - \bar{y}_i)^2 \sum_k (y_{kj} - \bar{y}_j)^2}}$$

then we have $\boxed{R = \mathbf{Z}^T \mathbf{Z}}$

factor analysis

data standardization, empirical correlation matrix

From the original data $\mathbf{Y} = (y_{ij})$, $i = 1, \ldots, n$, $j = 1, \ldots, p$ we get

$$z_{ij} = \frac{y_{ij} - \bar{y}_j}{\sqrt{n-1} \cdot s_j} \qquad \bar{y}_j = \frac{1}{n} \sum_{i=1}^{n} y_{ij} \qquad s_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_{ij} - \bar{y}_j)^2$$

z_{ij} are the "standardized data" (note: division by $\sqrt{n-1}$!), we combine them to the data matrix $\mathbf{Z} = (z_{ij})$.

Let $R = (r_{ij})$ be the empirical correlation matrix

$$r_{ij} = \frac{\sum_k (y_{ki} - \bar{y}_i)(y_{kj} - \bar{y}_j)}{\sqrt{\sum_k (y_{ki} - \bar{y}_i)^2 \sum_k (y_{kj} - \bar{y}_j)^2}}$$

then we have $\boxed{R = \mathbf{Z}^T \mathbf{Z}}$

factor analysis

data standardization, empirical correlation matrix

From the original data $\mathbf{Y} = (y_{ij})$, $i = 1, \ldots, n$, $j = 1, \ldots, p$ we get

$$z_{ij} = \frac{y_{ij} - \bar{y}_j}{\sqrt{n-1} \cdot s_j} \qquad \bar{y}_j = \frac{1}{n} \sum_{i=1}^{n} y_{ij} \qquad s_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_{ij} - \bar{y}_j)^2$$

z_{ij} are the "standardized data"(note: division by $\sqrt{n-1}$!), we combine them to the data matrix $\mathbf{Z} = (z_{ij})$.

Let $R = (r_{ij})$ be the empirical correlation matrix

$$r_{ij} = \frac{\sum_k (y_{ki} - \bar{y}_i)(y_{kj} - \bar{y}_j)}{\sqrt{\sum_k (y_{ki} - \bar{y}_i)^2 \sum_k (y_{kj} - \bar{y}_j)^2}}$$

then we have $\boxed{R = \mathbf{Z}^T \mathbf{Z}}$

factor analysis

data standardization, empirical correlation matrix

From the original data $\mathbf{Y} = (y_{ij})$, $i = 1, \ldots, n$, $j = 1, \ldots, p$ we get

$$z_{ij} = \frac{y_{ij} - \bar{y}_j}{\sqrt{n-1} \cdot s_j} \qquad \bar{y}_j = \frac{1}{n} \sum_{i=1}^{n} y_{ij} \qquad s_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_{ij} - \bar{y}_j)^2$$

z_{ij} are the "standardized data"(note: division by $\sqrt{n-1}$!), we combine them to the data matrix $\mathbf{Z} = (z_{ij})$.

Let $R = (r_{ij})$ be the empirical correlation matrix

$$r_{ij} = \frac{\sum_k (y_{ki} - \bar{y}_i)(y_{kj} - \bar{y}_j)}{\sqrt{\sum_k (y_{ki} - \bar{y}_i)^2 \sum_k (y_{kj} - \bar{y}_j)^2}}$$

then we have $\boxed{R = \mathbf{Z}^T \mathbf{Z}}$

factor analysis

data standardization, empirical correlation matrix

From the original data $\mathbf{Y} = (y_{ij})$, $i = 1, \ldots, n$, $j = 1, \ldots, p$ we get

$$z_{ij} = \frac{y_{ij} - \bar{y}_j}{\sqrt{n-1} \cdot s_j} \qquad \bar{y}_j = \frac{1}{n} \sum_{i=1}^{n} y_{ij} \qquad s_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_{ij} - \bar{y}_j)^2$$

z_{ij} are the "standardized data"(note: division by $\sqrt{n-1}$!), we combine them to the data matrix $\mathbf{Z} = (z_{ij})$.

Let $R = (r_{ij})$ be the empirical correlation matrix

$$r_{ij} = \frac{\sum_k (y_{ki} - \bar{y}_i)(y_{kj} - \bar{y}_j)}{\sqrt{\sum_k (y_{ki} - \bar{y}_i)^2 \sum_k (y_{kj} - \bar{y}_j)^2}}$$

then we have $\boxed{R = \mathbf{Z}^T \mathbf{Z}}$