Lecture 7: Analysis of Factors and Canonical Correlations

Måns Thulin

Department of Mathematics, Uppsala University
thulin@math.uu.se

Multivariate Methods • 6/5 2011

Outline

- Factor analysis
  - Basic idea: factor the covariance matrix
  - Methods for factoring
- Canonical correlation analysis
  - Basic idea: correlations between sets
  - Finding the canonical correlations

Factor analysis

Assume that we have a data set with many variables and that it is reasonable to believe that all these, to some extent, depend on a few underlying but unobservable factors.

The purpose of *factor analysis* is to find dependencies on such factors and to use this to reduce the dimensionality of the data set. In particular, the covariance matrix is described by the factors.

Factor analysis: an early example

C. Spearman (1904), *General Intelligence, Objectively Determined and Measured*, The American Journal of Psychology.

Children's performance in mathematics (X_1), French (X_2) and English (X_3) was measured. Correlation matrix:

$$\mathbf{R} = \begin{bmatrix} 1 & 0.67 & 0.64 \\ & 1 & 0.67 \\ & & 1 \end{bmatrix}$$

Assume the following model:

$$X_1 = \lambda_1 f + \epsilon_1, \quad X_2 = \lambda_2 f + \epsilon_2, \quad X_3 = \lambda_3 f + \epsilon_3$$

where f is an underlying "common factor" ("general ability"),

λ_1, λ_2, λ_3 are "factor loadings" and

ϵ_1, ϵ_2, ϵ_3 are random disturbance terms.

Factor analysis: an early example

Model:

$$X_i = \lambda_i f + \epsilon_i, \quad i = 1, 2, 3$$

with the unobservable factor

$$f = \text{“General ability”}$$

The variation of ϵ_i consists of two parts:

- a part that represents the extent to which an individual's matehematics ability, say, differs from her general ability
- a "measurement error" due to the experimental setup, since examination is only an approximate measure of her ability in the subject

The relative sizes of $\lambda_i f$ and ϵ_i tell us to which extent variation between individuals can be described by the factor.

Factor analysis: linear model

In factor analysis, a linear model is assumed:

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{LF} + \boldsymbol{\epsilon}$$

$\mathbf{X}$ $(p \times 1)$ = observable random vector
$\mathbf{L}$ $(p \times m)$ = matrix of *factor loadings*
$\mathbf{F}$ $(m \times 1)$ = *common factors*; unobserved values of factors which describe major features of members of the population
$\boldsymbol{\epsilon}$ $(p \times 1)$ = *errors/specific factors*; measurement error and variation not accounted for by the common factors

μ_i = mean of variable i
ϵ_i = ith specific factor
F_j = jth common factor
ℓ_{ij} = loading of the ith variable on the jth factor

The above model differs from ordinary linear models in that the independent variables **LF** are unobservable and random.

Factor analysis: linear model

We assume that the unobservable random vectors $\mathbf{F}$ and ϵ satisfy the following conditions:

- $\mathbf{F}$ and ϵ are independent.
- $\mathrm{E}(\epsilon) = \mathbf{0}$ and $\mathrm{Cov}(\epsilon) = \mathbf{\Psi}$, where $\mathbf{\Psi}$ is a diagonal matrix.
- $\mathrm{E}(\mathbf{F}) = \mathbf{0}$ and $\mathrm{Cov}(\mathbf{F}) = \mathbf{I}$.

Thus, the factors are assumed to be uncorrelated. This is called the *orthogonal factor model*.

Factor analysis: factoring $\boldsymbol{\Sigma}$

Given the model $\mathbf{X} = \boldsymbol{\mu} + \mathbf{LF} + \epsilon$, what is $\boldsymbol{\Sigma}$?

See blackboard!

We find that $\text{Cov}(\mathbf{X}) = \mathbf{LL}' + \boldsymbol{\Psi}$ where

$$\begin{aligned} \text{V}(X_i) &= \ell_{i1}^2 + \cdots + \ell_{im}^2 + \psi_i \quad \text{and} \\ \text{Cov}(X_i, X_k) &= \ell_{i1}\ell_{k1} + \cdots + \ell_{im}\ell_{km}. \end{aligned}$$

Furthermore, $\text{Cov}(\mathbf{X}, \mathbf{F}) = \mathbf{L}$, so that $\text{Cov}(X_i, F_j) = \ell_{ij}$.

If $m =$ the number of factors is much smaller than $p =$the number of measured attributes, the covariance of $\mathbf{X}$ can be described by the pm elements of $\mathbf{LL'}$ and the p nonzero elements of $\boldsymbol{\Psi}$, rather than the $(p^2 + p)/2$ elements of $\boldsymbol{\Sigma}$.

Factor analysis: factoring $\boldsymbol{\Sigma}$

The marginal variance σ_{ii} can also be partitioned into two parts:

The *i*th communality: The proportion of the variance at the ith measurement X_i contributed by the factors $F_1, F_2, \ldots, F_m$.

The uniqueness or **specific variance:** The remaining proportion of the variance of the ith measurement, associated with ϵ_i.

From $\text{Cov}(\mathbf{X}) = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$ we have that

$$\sigma_{ii} = \underbrace{\ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{im}^2}_{Communality, h_i^2} + \underbrace{\psi_i}_{Specific \ \ variance}$$

so

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

Orthogonal transformations

Consider an orthogonal matrix $\mathbf{T}$.

See blackboard!

We find that factor loadings $\mathbf{L}$ are determined only up to an orthogonal matrix $\mathbf{T}$. With $\mathbf{L}^* = \mathbf{LT}$ and $\mathbf{F}^* = \mathbf{T'F}$, the pairs

$$(\mathbf{L}^*, \mathbf{F}^*) \quad \text{and} \quad (\mathbf{L}, \mathbf{F})$$

give equally valid decompositions.

The communalities given by the diagonal elements of $\mathbf{LL}' = (\mathbf{L}^*)(\mathbf{L}^*)'$ are also unaffected by the choice of $\mathbf{T}$.

Factor analysis: estimation

- The factor model is determined uniquely up to an orthogonal transformation of the factors. How then can $\mathbf{L}$ and $\mathbf{\Psi}$ be estimated? Is some $\mathbf{L}^* = \mathbf{LT}$ better than others?
- **Idea:** Find some $\mathbf{L}$ and $\mathbf{\Psi}$ and then consider various $\mathbf{L}^*$.
- Methods for estimation:
  - Principal component method
  - Principal factor method
  - Maximum likelihood method

Factor analysis: principal component method

Let $(\lambda_i, \mathbf{e_i})$ be the eigenvalue-eigenvector pairs of $\mathbf{\Sigma}$, with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$.

From the spectral theorem (and the principal components lecture) we know that

$$\mathbf{\Sigma} = \lambda_1 \mathbf{e_1} \mathbf{e_1'} + \lambda_2 \mathbf{e_2} \mathbf{e_2'} + \ldots + \lambda_p \mathbf{e_p} \mathbf{e_p'}.$$

Let $\mathbf{L} = (\sqrt{\lambda_1}\mathbf{e_1}, \sqrt{\lambda_2}\mathbf{e_2}, \ldots, \sqrt{\lambda_p}\mathbf{e_p})$. Then

$$\mathbf{\Sigma} = \lambda_1 \mathbf{e_1} \mathbf{e_1'} + \ldots + \lambda_p \mathbf{e_p} \mathbf{e_p'} = \mathbf{L}\mathbf{L}' = \mathbf{L}\mathbf{L}' + \mathbf{0}.$$

Thus $\mathbf{L}$ is given by $\sqrt{\lambda_i}$ times the coefficients of the principal components, and $\mathbf{\Psi} = \mathbf{0}$.

Factor analysis: principal component method

Now, if $\lambda_{m+1}, \lambda_{m+2}, \ldots, \lambda_p$ are small, then the first m principal components explain most of $\mathbf{\Sigma}$.

Thus, with $\mathbf{L_m} = (\sqrt{\lambda_1}\mathbf{e_1}, \ldots, \sqrt{\lambda_m}\mathbf{e_m})$,

$$\mathbf{\Sigma} \approx \mathbf{L_m}\mathbf{L_m}'.$$

With specific factors, this becomes

$$\mathbf{\Sigma} \approx \mathbf{L_m}\mathbf{L_m}' + \mathbf{\Psi}$$

where $\Psi_i = \sigma_{ii} - \sum_{i=1}^{m} l_{ij}^2$.

As estimators for the factor loadings and specific variances, we take

$$\widetilde{\mathbf{L}} = \widetilde{\mathbf{L}}_m = \left(\sqrt{\hat{\lambda}_1}\hat{\mathbf{e}}_\mathbf{1}, \ldots, \sqrt{\hat{\lambda}_m}\hat{\mathbf{e}}_\mathbf{m}\right)$$

where $(\hat{\lambda}_i, \hat{\mathbf{e}}_\mathbf{i})$ are the eigenvalue-eigenvector pairs of the sample covariance matrix $\mathbf{S}$, and

$$\widetilde{\mathbf{\Psi}} = s_{ii} - \sum_{i=1}^{m} \widetilde{l}_{ij}^2.$$

Factor analysis: principal component method

In many cases, the correlation matrix **R** (which is also the covariance matrix of the standardized data) is used instead of **S**, to avoid problems related to measurements being in different scales.

Example: consumer-preference data – p. 491 in J&W.

See R code!

- Is this a good example? Can factor analysis be applied to ordinal data?

Example: stock price data – p. 493 in J&W.

See R code!

Factor analysis: other estimation methods

- Principal factor method:
  - Modification of principal component approach.
  - Uses initial estimates of the specific variances.

- Maximum likelihood method:
  - Assuming normal data, the maximum likelihood estimators of $\mathbf{L}$ and $\boldsymbol{\Psi}$ are derived.
  - In general the estimators must be calculated by numerical maximization of a function of matrices.

Factor analysis: factor rotation

Given an orthogonal matrix $\mathbf{T}$, a different but equally valid possible estimator of the factor loadings is $\widetilde{\mathbf{L}}^* = \widetilde{\mathbf{L}}\mathbf{T}$.

We would like to find a $\widetilde{\mathbf{L}}^*$ that gives a nice and simple interpretation of the corresponding factors.
Ideally: a pattern of loadings such that each variable loads highly on a single factor and has small to moderate loadings on the remaining factors.

The varimax criterion:

Define $\widetilde{\ell}^*_{ij} = \hat{\ell}^*_{ij}/\hat{h}_i$. This scaling gives variables with smaller communalities more influence.

Select the orthogonal transformation $\mathbf{T}$ that makes

$$V = \frac{1}{p}\sum_{i=1}^{m}\left[\sum_{i=1}^{p}\widetilde{\ell}^{*4}_{ij} - \frac{1}{p}\left(\sum_{i=1}^{p}\widetilde{\ell}^{*2}_{ij}\right)^2\right]$$

as large as possible.

Factor analysis: factor rotation

Example: consumer-preference data – p. 508 in J&W.

See R code!

Example: stock price data – p. 510 in J&W.

See R code!

Factor analysis: limitations of orthogonal factor model

- The factor model is determined uniquely up to an orthogonal transformation of the factors.
- Linearity: the linear covariance approximation $\mathbf{LL}' + \mathbf{\Psi}$ may not be appropriate.
- The factor model is most useful when m is small, but in many cases $mp + p$ parameters are not adequate and $\mathbf{\Sigma}$ is not close to $\mathbf{LL}' + \mathbf{\Psi}$.

Factor analysis: further topics

Some further topics of interest are:

- Tests for the number of common factors.
- Factor scores.
- Oblique factor model
  - In which $Cov(\mathbf{F})$ is not diagonal; the factors are correlated.

Canonical correlations

Canonical correlation analysis – CCA – is a means of assessing the relationship *between two sets* of variables.

The idea is to study the correlation between a linear combination of the variables in one set and a linear combination of the variables in another set.

Canonical correlations: the exams problem

In a classical CCA problem exams in five topics are considered.

Exams are closed-book (C) or open-book (O):

Mechanics (C), Vectors (C),
Algebra (O), Analysis (O), Statistics (O).

Question: how highly correlated is a student's performance in closed-book exams with her performance in open-book exams?

Canonical correlations: basic idea

Given a data set $\mathbf{X}$, partition the collection of variables into two sets:

$$\mathbf{X}^{(1)} \quad (p \times 1) \quad \text{and} \quad \mathbf{X}^{(2)} \quad (q \times 1).$$

(Assume $p \leq q$.)

For these random vectors:

$$\begin{aligned} E(\mathbf{X}^{(1)}) &= \mu^{(1)}, \quad \text{Cov}(\mathbf{X}^{(1)}) = \mathbf{\Sigma}_{11} \\ E(\mathbf{X}^{(2)}) &= \mu^{(2)}, \quad \text{Cov}(\mathbf{X}^{(2)}) = \mathbf{\Sigma}_{22} \end{aligned}$$

and

$$\text{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \mathbf{\Sigma}_{12} = \mathbf{\Sigma}'_{21}.$$

Canonical correlations: basic idea

Introduce the linear combinations

$$U = \mathbf{a}'\mathbf{X}^{(1)}, \qquad V = \mathbf{b}'\mathbf{X}^{(2)}.$$

For these, $Var(U) = \mathbf{a}'\mathbf{\Sigma}_{11}\mathbf{a}$, $Var(V) = \mathbf{b}'\mathbf{\Sigma}_{22}\mathbf{b}$ and $Cov(U, V) = \mathbf{a}'\mathbf{\Sigma}_{12}\mathbf{b}$.

Goal: Seek $\mathbf{a}$ and $\mathbf{b}$ such that

$$\text{Corr}(U, V) = \frac{\mathbf{a}'\mathbf{\Sigma}_{12}\mathbf{b}}{\sqrt{\mathbf{a}'\mathbf{\Sigma}_{11}\mathbf{a}}\sqrt{\mathbf{b}'\mathbf{\Sigma}_{22}\mathbf{b}}}$$

is as large as possible.

Canonical correlations: some definitions

The **first pair of canonical variables** is the pair of linear combinations U_1 and V_1 having unit variances, which maximize the correlation

$$\text{Corr}(U, V) = \frac{\mathbf{a}'\mathbf{\Sigma}_{12}\mathbf{b}}{\sqrt{\mathbf{a}'\mathbf{\Sigma}_{11}\mathbf{a}}\sqrt{\mathbf{b}'\mathbf{\Sigma}_{22}\mathbf{b}}}$$

The **second pair of canonical variables** is the pair of linear combinations U_2 and V_2 having unit variances, which maximize the correlation among all choices that are uncorrelated with the first pair of canonical variables.

The k**th pair of canonical variables** is the pair of linear combinations U_k and V_k having unit variances, which maximize the correlation among all choices that are uncorrelated with the previous $k-1$ canonical variable pairs.

The correlation between the kth pair of canonical variables is called the k**th canonical correlation**.

Canonical correlations: solution

Result 10.1. The kth pair of canonical variables is given by

$$U_k = \mathbf{e}_k' \mathbf{\Sigma}_{11}^{-1/2} \mathbf{X}^{(1)} \quad \text{and} \quad V_k = \mathbf{f}_k' \mathbf{\Sigma}_{22}^{-1/2} \mathbf{X}^{(2)}.$$

We have that

$$Var(U_k) = Var(V_k) = 1$$

and

$$Cov(U_k, V_k) = \rho_k^*,$$

where $\rho_1^{*2} \geq \rho_2^{*2} \geq \ldots \geq \rho_p^{*2}$ and $(\rho_k^{*2}, \mathbf{e_k})$ and $(\rho_k^{*2}, \mathbf{f_k})$ are the eigenvalue-eigenvectors pairs of

$\mathbf{\Sigma}_{11}^{-1/2} \mathbf{\Sigma}_{12} \mathbf{\Sigma}_{22}^{-1/2} \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{11}^{-1/2}$ and

$\mathbf{\Sigma}_{22}^{-1/2} \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{11}^{-1/2} \mathbf{\Sigma}_{12} \mathbf{\Sigma}_{22}^{-1/2}$, respectively.

Canonical correlations: solution

Alternatively, the canonical variables coefficient vectors **a** and **b** and the corresponding correlations can be found by solving the eigenvalue equations

$$
\begin{aligned}
\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\,\mathbf{a} &= \rho^{*2}\mathbf{a} \\
\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\,\mathbf{b} &= \rho^{*2}\mathbf{b}
\end{aligned}
$$

This is often more practical to use when computing the coefficients and the correlations.

Canonical correlations: head-lengths of sons

Example: Consider the two sons born first in n families. For these, consider the following measurements:
X_1: head length of first son; X_2: head breadth of first son; X_3: head length of second son; X_4: head breadth of second son.

Observations:

$$\bar{\mathbf{x}} = \begin{bmatrix} 185.72 \\ 151.12 \\ 183.84 \\ 149.24 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 91.481 & 50.753 & 66.875 & 44.267 \\ & 52.186 & 49.259 & 33.651 \\ & & 96.775 & 54.278 \\ & & & 43.222 \end{bmatrix}$$

See R code!

Correlation matrices:

$$\mathbf{R}_{11} = \begin{bmatrix} 1 & 0.7346 \\ 0.7346 & 1 \end{bmatrix}, \quad \mathbf{R}_{22} = \begin{bmatrix} 1 & 0.8392 \\ 0.8392 & 1 \end{bmatrix}$$

$$\mathbf{R}_{12} = \mathbf{R}_{21}' = \begin{bmatrix} 0.7107 & 0.7040 \\ 0.6931 & 0.7085 \end{bmatrix}$$

Canonical correlations: head-lengths of sons

The eigenvalues of $\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21}$ are 0.6218 and 0.0029.

Thus the canonical correlations are: $\sqrt{0.6218} = 0.7886$ and $\sqrt{0.0029} = 0.0539$ (or are they? Remember to check signs!).

From the eigenvectors we obtain the canonical correlation vectors :

$$\mathbf{a}_1 = \begin{bmatrix} 0.727 \\ 0.687 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 0.704 \\ -0.710 \end{bmatrix}$$

and

$$\mathbf{b}_1 = \begin{bmatrix} 0.684 \\ 0.730 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} 0.709 \\ -0.705 \end{bmatrix}$$

Note: The canonical correlation $\hat{\rho}_1^* = -0.7886$ exceeds any of the individual correlations between a variable of the first set and a variable of the second set.

Canonical correlations: head-lengths of sons

The first canonical correlation variables are

$$\begin{cases} U = 0.727x_1^{(1)} + 0.687x_2^{(1)} \\ V = 0.684x_1^{(2)} + 0.730x_2^{(2)} \end{cases}$$

with

$$Corr(U, V) = -0.7886.$$

Interpretation: The sum of length and breadth of head size of each brother. These variables are highly negatively correlated between brothers.

Canonical correlations: head-lengths of sons

The second canonical correlation variables are

$$\begin{cases} U = 0.704x_1^{(1)} - 0.710x_2^{(1)} \\ V = 0.709x_1^{(2)} - 0.705x_2^{(2)} \end{cases}$$

with

$$Corr(U, V) = 0.0539.$$

Interpretation: Seem to measure the difference between length and breadth. Head shape of first and second brothers appear to have little correlation.

Canonical correlations: poor summary of variability

Example: Consider the following covariance matrix ($p = q = 2$):

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix} = \begin{bmatrix} 100 & 0 & 0 & 0 \\ 0 & 1 & .95 & 0 \\ 0 & .95 & 1 & 0 \\ 0 & 0 & 0 & 100 \end{bmatrix}$$

It can be shown that the first pair of canonical variates is

$$U_1 = X_2^{(1)}, \quad V_1 = X_1^{(2)}$$

with correlation

$$\rho_1^* = \text{Corr}(U_1, V_1) = 0.95.$$

However, $U_1 = X_2^{(1)}$ provides a very poor summary of the variability in the first set.

Most of the variability in the first set is in $X_1^{(1)}$, which is uncorrelated with U_1. (The same situation is true for $V_1 = X_1^{(2)}$ in the second set.)

Canonical correlations: further topics

Some further topics of interest are:

- Tests for $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.
- Sample descriptive measures.
- Connection to MANOVA.

Summary

- Factor analysis
  - Basic idea: factor the covariance matrix
  - Methods for factoring
    - Principal components
    - Maximum likelihood
- Canonical correlation analysis
  - Basic idea: correlations between sets
  - Finding the canonical correlations
    - Eigenvalues and eigenvectors