Displays for Statistics 5401/8401

Lecture 28

November 11, 2005

Christopher Bingham, Instructor

612-625-1024, kb@umn.edu
372 Ford Hall

Class Web Page

http://www.stat.umn.edu/~kb/classes/5401

Review

The *factor analysis model* with m factors is $\underset{p\times 1}{\mathbf{x}} = \underset{p\times 1}{\boldsymbol{\mu}} + \underset{p\times m}{\mathbf{L}}\,\underset{m\times 1}{\mathbf{f}} + \underset{p\times 1}{\boldsymbol{\varepsilon}}$

$$V[\boldsymbol{\varepsilon}] = \boldsymbol{\Psi} = \text{diag}[\psi_1, \psi_2, \ldots, \psi_p]$$

- Elements f_j of $\mathbf{f}$ are *common factors*.
- Elements ε_k of $\boldsymbol{\varepsilon}$ are *unique factors*. and are uncorrelated with $f_1,\ldots,f_m$.
- Elements ℓ_{kj} of $\mathbf{L}$ are *loadings* of variable k on factor j.
- The diagonal elements $\psi_i = V[\varepsilon_i]$ of $\boldsymbol{\Psi}$ are called the *uniquenesses* or *specific variances*.
- $h_k^2 \equiv \sigma_{kk} - \psi_k = V[\sum_{1\leq j\leq m}\ell_{kj}f_j] = V[x_k-\mu_k-\varepsilon_k]$ are the *communalities*. You can show that $|\rho_{k\ell}| \leq (h_k/\sqrt{\sigma_{kk}})(h_\ell/\sqrt{\sigma_{\ell\ell}})$, so when h_k^2 is small relative to σ_{kk}, x_k can't be highly correlated with other variables.

Except by convention or subject matter considerations, nothing can be said about $\boldsymbol{\mu}_f = E[\mathbf{f}]$ and the m by m matrix $\boldsymbol{\Gamma} \equiv V[\mathbf{f}]$.

However, since factors are unobservable, you lose no generality by assuming $\boldsymbol{\mu}_f = \mathbf{0}$, and $\gamma_{jj} = V[f_j] = 1$

Often factors are assumed to be *uncorrelated* so that $\boldsymbol{\Gamma} = V[\mathbf{f}] = \mathbf{I}_m$.

Vocabulary

When $\boldsymbol{\Gamma} = V[\mathbf{f}] = \mathbf{I}_m$, the model is an *orthogonal factor model*.

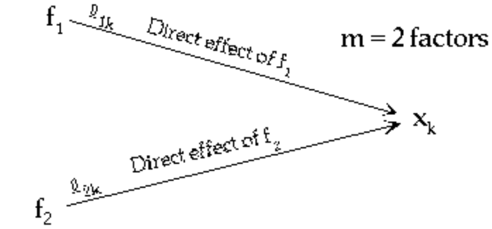
For factors $f_j = z_j/\sqrt{\lambda_j}$ defined in terms of PCs have $\boldsymbol{\Gamma} = V[\mathbf{f}] = \mathbf{I}_m$ and are therefore orthogonal factors.

This is because the principal components $z_j = \mathbf{v}_j(\mathbf{x} - \boldsymbol{\mu})$, are uncorrelated with $V[z_j] = \lambda_j$.

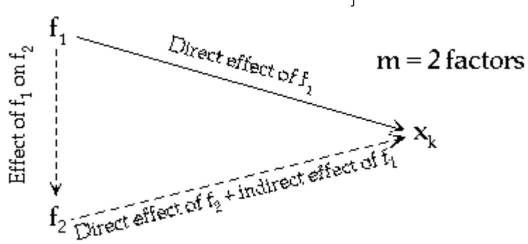
When $\boldsymbol{\Gamma} = V[\mathbf{f}] \neq \mathbf{I}_m$, the factor model is *oblique*.

- Orthogonal factors are attractive because you can unambiguously separate the effects of different factors.
- The attraction of oblique factor analysis is that you may be able to obtain a simpler $\mathbf{L}$.

When factors are uncorrelated, there is no ambiguity in defining the *effect* of factor j on variable k. It is simply ℓ_{kj}.



When factor f_j and f_k are *correlated*, there is also an *indirect* effect of f_j because the value of f_k may be changed by a change in the value of f_j.



The orthogonal factor model

When $\mathbf{\Gamma} = V[\mathbf{f}] = \mathbf{I}_m$, some formulas are simpler.

- $\mathrm{Cov}[x_k, f_j] =$
  $\mathrm{Cov}[\mu_k + \sum_{1\le i\le m}\ell_{ki}f_i + \varepsilon_k, f_j] = \ell_{kj}$
- $\mathrm{Corr}[x_k, f_j] = \ell_{kj}/\sqrt{\sigma_{kk}}$
- $h_k^2 = V[x_k - \varepsilon_k] = \sum_{1\le j\le m}\ell_{kj}^2$, sum of squares of *row* k of $\mathbf{L}$.
- $\psi_k = V[\varepsilon_k] = \sigma_{kk} - h_k^2 = \sigma_{kk} - \sum_{1\le j\le m}\ell_{kj}^2$
- $\sigma_{kk} = V[x_k] = \sum_{1\le j\le m}\ell_{kj}^2 + \psi_k.$
- $\sigma_{k\ell} = \mathrm{Cov}[x_k, x_\ell] = \sum_{1\le j\le m}\ell_{kj}\ell_{\ell j}$

Note: These are <u>wrong</u> when factors are *not* orthogonal. In general, when $V[\mathbf{f}] = \mathbf{\Gamma} = [\gamma_{j_1j_2}]$,

$$\sigma_{kk} = V[x_k] = \sum_{1\le j_1\le m}\sum_{1\le j_2\le m}\gamma_{j_1j_2}\ell_{kj_1}\ell_{kj_2} + \psi_k.$$
$$\sigma_{k\ell} = \mathrm{Cov}[x_k, x_\ell] = \sum_{1\le j_1\le m}\sum_{1\le j_2\le m}\gamma_{j_1j_2}\ell_{kj_1}\ell_{\ell j_2}$$

The factor analytic model $\mathbf{x} = \boldsymbol{\mu} + \mathbf{Lf} + \boldsymbol{\varepsilon}$ implies the following structure for $\mathbf{\Sigma}$:

$$V[\mathbf{x}] = \mathbf{\Sigma} = \mathbf{V} + \mathbf{\Psi} = \mathbf{L\Gamma L}' + \mathbf{\Psi},\ \mathbf{\Gamma} = V[\mathbf{f}]$$

where

- $\mathbf{V} = V[\mathbf{Lf}] = \mathbf{L\Gamma L}'$ has *rank m* < p
- $\mathbf{\Psi} = \mathbf{\Sigma} - \mathbf{V}$ is *diagonal* with $\psi_i \ge 0$

Vocabulary

A matrix $\mathbf{\Sigma}$ that can be represented as $\mathbf{\Sigma} = \mathbf{V} + \mathbf{\Psi}$, where

- $\mathbf{V}$ has <u>rank m</u> < p and is positive semi-definite (m eigenvalues > 0)
- $\mathbf{\Psi}$ is <u>diagonal</u> with $\psi_j \ge 0$

is said to have *factor analytic form*.

You can estimate $\mathbf{V}$ and $\mathbf{\Psi}$ without ambiguity, but not $\mathbf{L}$ or $\mathbf{\Gamma}$. When m > 1, there are infinitely many $\mathbf{L}$'s compatible with $\mathbf{V}$. When m = 1, there are two.

So far the focus has been on explaining the <u>covariances</u> $\sigma_{k\ell}$, $k \ne \ell$.

In practice, the emphasis is usually in explaining <u>correlations</u>.

When $\mathbf{\Delta} \equiv \mathrm{diag}[1/\sqrt{\sigma_{11}},\ldots,1/\sqrt{\sigma_{pp}}]$, since $\rho_{k\ell} = \sigma_{k\ell}/\{\sqrt{\sigma_{kk}}\sqrt{\sigma_{\ell\ell}}\}$, the population <u>correlation matrix</u> of $\mathbf{x}$ is

$$\boldsymbol{\rho} = \mathbf{\Delta\Sigma\Delta} = \mathbf{\Delta V\Delta} + \mathbf{\Delta\psi\Delta}$$
$$= \tilde{\mathbf{V}} + \tilde{\mathbf{\Psi}}$$

- $\tilde{\mathbf{V}} \equiv \mathbf{\Delta V\Delta}$, p by p *rank m*,
- $\tilde{\mathbf{\Psi}} \equiv \mathbf{\Delta\Psi\Delta}$, p by p $\mathrm{diag}[\tilde{\psi}_1,\ldots,\tilde{\psi}_p]$, with $\tilde{\psi}_k = \psi_k/\sigma_{kk}$

Thus $\boldsymbol{\rho}$ is also of *factor analytic form*. When $\mathbf{\Gamma} = \mathbf{I}_m$,

$$\tilde{\mathbf{V}} = \mathbf{\Delta V\Delta} = \mathbf{\Delta LL}'\mathbf{\Delta} = \tilde{\mathbf{L}}\tilde{\mathbf{L}}'$$

where

$$\tilde{\mathbf{L}} = \mathbf{\Delta L} = [\tilde{\boldsymbol{\ell}}_1,\ldots,\tilde{\boldsymbol{\ell}}_m],\ \tilde{\ell}_{kj} = \ell_{kj}/\sqrt{\sigma_{kk}}$$

Summary

- $\mathbf{\Sigma}$ has factor analytic structure $\Leftrightarrow$ $\boldsymbol{\rho}$ has factor analytic structure
- $\mathbf{x}$ follows a factor analytic model $\Leftrightarrow$ $\tilde{\mathbf{x}} = [\tilde{x}_1, \ldots, \tilde{x}_p]$ does, $\tilde{x}_k = (x_k - \mu_k)/\sqrt{\sigma_{kk}}$ = z-score computed from x_k.

There are *direct ways* to go between factor analytic representations for

- $\mathbf{\Sigma}$ in terms of $\mathbf{L}$ and $\mathbf{\Psi}$
- $\boldsymbol{\rho}$ in terms of $\tilde{\mathbf{L}}$ and $\tilde{\mathbf{\Psi}}$.

$\mathbf{\Sigma} \Rightarrow \boldsymbol{\rho}$	$\tilde{\mathbf{L}} = \mathbf{\Delta}\mathbf{L}$	$\tilde{\mathbf{\Psi}} = \mathbf{\Delta}\mathbf{\Psi}\mathbf{\Delta}$
$\boldsymbol{\rho} \Rightarrow \mathbf{\Sigma}$	$\mathbf{L} = \mathbf{\Delta}^{-1}\tilde{\mathbf{L}}$	$\mathbf{\Psi} = \mathbf{\Delta}^{-1}\tilde{\mathbf{\Psi}}\mathbf{\Delta}^{-1}$

This differs from the Principal Component model where there are no simple ways to go between covariance PCs and correlation PCs.

The quantities

- $\tilde{h}_k^2 = \sum_j \tilde{\ell}_{kj}^2 = h_k^2/\sigma_{kk}$
- $\tilde{\psi}_k = \psi_k/\sigma_{kk}$

based on the correlation matrix $\boldsymbol{\rho}$ are also called *communalities* and *uniquenesses*.

- $\tilde{h}_k^2 + \tilde{\psi}_k = 1 = V[\tilde{x}_k]$, $\tilde{x}_k = (x_k - \mu_k)/\sqrt{\sigma_{kk}}$
- $\tilde{h}_k^2 = h_k^2/\sigma_{kk}$ measures the influence of the common factors on $\tilde{x}_k$, the standardized version of x_k.

  Because $|\rho_{k\ell}| \leq \tilde{h}_k\tilde{h}_\ell$, low $\tilde{h}_k$ implies low $\rho_{k\ell}$, $\ell \neq k$ because x_k doesn't share much in common with x_ℓ.
- $\tilde{\psi}_k = \psi_k/\sigma_{kk}$ measures the influence of the unique factor ε_k on $\tilde{x}_k$.

- $\tilde{h}_k^2$ is analogous to multiple R^2 in regression.

  In fact, because the model says that *all* dependence of x_k on the other x_ℓ's comes through the f_j's, a first *guess* at $\tilde{h}_k^2$ might be R^2 from a regression of x_k on $x_1,\ldots,x_{k-1}, x_{k+1},\ldots,x_p$. They are <u>not</u> the same, however.
- $\tilde{\psi}_k = 1 - \tilde{h}_k^2$ is analogous to $1 - R^2$, so a first guess at $\tilde{\psi}_k$ might be $1 - R^2$ from that regression. This is often used to get starting values for iterative methods of factor extraction.
- $\tilde{\ell}_{kj}$ is the loading of standardized variable $\tilde{x}_k = (x_k - \mu_k)/\sqrt{\sigma_{kk}}$ on factor f_j and $\tilde{\ell}_{kj} = \text{corr}(x_k, f_j)$ (for orthogonal factor analysis).

Non-uniqueness of factors and factor loadings

A real problem with the factor analytic model is that loadings and factors are not <u>uniquely defined</u>.

Suppose the orthogonal factor analytic model

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{L}\mathbf{f} + \boldsymbol{\varepsilon}, \text{ with } \mathbf{\Gamma} = V[\mathbf{f}] = \mathbf{I}_m,$$
$$V[\boldsymbol{\varepsilon}] = \mathbf{\Psi} = \text{diag}[\psi_1,\ldots,\psi_p]$$

is a <u>correct</u> model for $\mathbf{x}$ in the sense that $E[\mathbf{x}] = \boldsymbol{\mu}$ and $V[\mathbf{x}] = \mathbf{L}\mathbf{L}' + \mathbf{\Psi}$.

The parameters are $\boldsymbol{\mu}$, $\mathbf{L}$ and $\mathbf{\Psi}$.

Q: What does it mean for parameters to not be unique?

A: There exist more than one set of parameter values which are consistent the <u>distribution</u> of your data.

In factor analysis , there is more than one $\mathbf{L}$ that is consistent with $V[\mathbf{x}]$.

- μ and Ψ are in fact unique.
- **L** and **f** are not unique.

You can always find (in many ways) a loading matrix $\mathbf{L}^* \neq \mathbf{L}$ and a vector $\mathbf{f}^* \neq \mathbf{f}$ of random factors f_j^* such that

- $\mathbf{L}^*\mathbf{f}^* = \mathbf{L}\mathbf{f}$
- $V[\mathbf{f}^*] = I_m$.

so that

$$\mathbf{x} = \mu + \mathbf{L}^*\mathbf{f}^* + \varepsilon,\ V[\mathbf{f}^*] = I_m,$$

is an orthogonal factor analytic model for **x** that is just as "correct" but *different* from the original one,

$$\mathbf{x} = \mu + \mathbf{L}\mathbf{f} + \varepsilon,\ V[\mathbf{f}] = I_m.$$

An expert in the field of application might prefer $\mathbf{L}^*$ to **L** but not on statistical grounds.

To be specific, choose *any* non-singular m×m **H** with *orthonormal* columns, that is, satisfying

$$\mathbf{H}'\mathbf{H} = \mathbf{H}\mathbf{H}' = I_m \quad (\mathbf{H}^{-1} = \mathbf{H}')$$

In other words, choose any *orthogonal* matrix **H**. Then define $\mathbf{L}^*$ and $\mathbf{f}^*$ as

$$\underset{p\times m}{\mathbf{L}^*} \equiv \underset{p\times m}{\mathbf{L}}\ \underset{m\times m}{\mathbf{H}} \quad \text{and} \quad \underset{m\times 1}{\mathbf{f}^*} \equiv \underset{p\times m}{\mathbf{H}'}\underset{m\times 1}{\mathbf{f}}$$

$\mathbf{L}^*$ is a new loading matrix and $\mathbf{f}^*$ is a new vector of factors which are linear combinations of the old factors in **f** .

Then

- $\mathbf{V}^* = \mathbf{L}^*\mathbf{L}^{*\prime} = \mathbf{L}\mathbf{H}\mathbf{H}'\mathbf{L}' = \mathbf{L}\mathbf{L}' = \mathbf{V}$
- $\mathbf{L}^*\mathbf{f}^* = \mathbf{L}\mathbf{H}\mathbf{H}'\mathbf{f} = \mathbf{L}\mathbf{f}$
- $V[\mathbf{f}^*] = \mathbf{H}'I_m\mathbf{H} = \mathbf{H}'\mathbf{H} = I_m = V[\mathbf{f}]$
- $\mathbf{x} = \mu + \mathbf{L}\mathbf{f} + \varepsilon = \mu + \mathbf{L}^*\mathbf{f}^* + \varepsilon$
  $\Sigma = \mathbf{V} + \Psi = \mathbf{V}^* + \Psi$

We have now

- A different factorization $\mathbf{L}^*\mathbf{L}^{*\prime}$ of
  $$\mathbf{V} = \Sigma - \Psi = \mathbf{L}\mathbf{L}' = \mathbf{L}^*\mathbf{L}^{*\prime}$$
  with $\mathbf{L}^* = \mathbf{L}\mathbf{H} \neq \mathbf{L}$
- A new representation of **x** in terms of factors f_k^* with loading matrix $\mathbf{L}^*$:
  $$\mathbf{x} = \mu + \mathbf{L}^*\mathbf{f}^* + \varepsilon,\ \mathbf{L}^* \neq \mathbf{L},\ \mathbf{f}^* \neq \mathbf{f}$$
- The f_j^*'s are orthonormal factors that are linear combinations of f_j's with coefficients taken from the *columns* of **H**, that is $\mathbf{f}^* = \mathbf{H}'\mathbf{f}$.
- Conversely, the f_j's are linear combinations of the f_j^*'s with coefficients taken from the *rows* of **H**: $\mathbf{f} = \mathbf{H}\mathbf{f}^*$.

Fact:
det(**H**) = ±1 for any orthogonal **H**.

When det(**H**) = +1, **H** and **H**′ are *rotation matrices* that correspond to rigid rotations of m-dimensional space

Suppose $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_N$ are N vectors of m factor scores. Then you can view them as points in m-dimensional space. If **H** is a rotation matrix then the transformation

$$\mathbf{f}_j \to \mathbf{f}_j^* = \mathbf{H}'\mathbf{f}_j$$

amounts to rotating the m-dimensional space of factor scores about the origin **0** = [0, ..., 0]′ in the process of which point $\mathbf{f}_j$ is moved to a new point $\mathbf{f}_j^*$.

There another set of entities that are rotated. These are the points $\mathbf{l}_i$ whose m coordinates come from the rows of

$$\mathbf{L} = \begin{bmatrix} \mathbf{l}_1' \\ \mathbf{l}_2' \\ \cdots \\ \mathbf{l}_p' \end{bmatrix}, \; \mathbf{l}_k' = [\ell_{k1}, \ell_{k2}, \ldots, \ell_{km}]$$

Then the change $\mathbf{L} \rightarrow \mathbf{L}^* = \mathbf{LH}$ rotates $\mathbf{l}_k$ into $\mathbf{l}_k^* = \mathbf{H}'\mathbf{l}_k$.

If you view $\mathbf{l}_1, \ldots, \mathbf{l}_p$ as defining p points in m-dimensional space, then $\mathbf{l}_1^*, \ldots, \mathbf{l}_m^*$ are the same points after the space of loadings is rotated by $\mathbf{H}$.

When m = 2, for every rotation matrix $\mathbf{H}$ ($\det(\mathbf{H}) = +1$) there is an angle θ, $-\pi < \theta \leq \pi$ such that

$$\mathbf{H} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

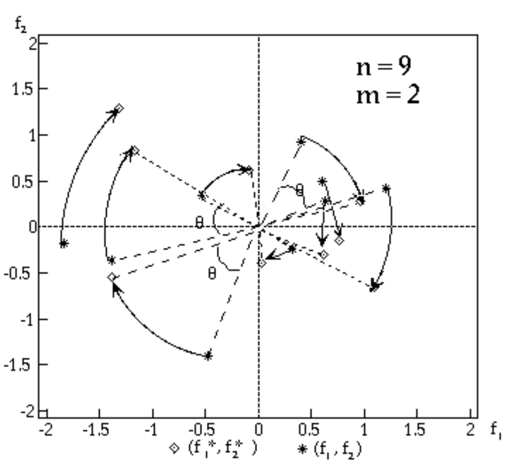
This corresponds to a rotation by angle θ.

When you combine a rotation with a change of sign of one coordinate, you get

$$\tilde{\mathbf{H}} = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}, \; -\pi < \theta < \pi$$

$\tilde{\mathbf{H}}$ is orthogonal, but is not a rotation matrix since $\det(\tilde{\mathbf{H}}) = -1$. It carries out a rotation followed by a "reflection" in one of the coordinate axes.

Here is a plot of factor scores (f_1,f_2) (∗) and rotated factors scores (f_1^*,f_2^*) (⋄) for n = 9 cases, with m = 2.



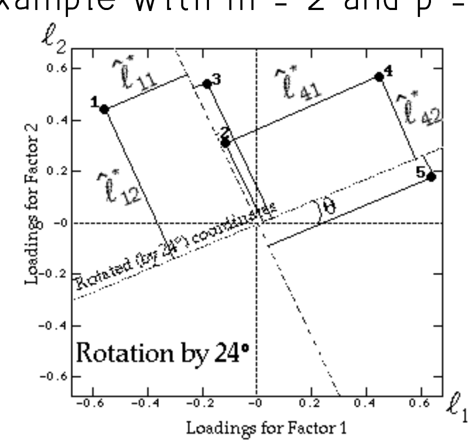
Lines and curves connect corresponding $\mathbf{f}$ and $\mathbf{f}^*$ points.

All the rotation angles are θ.

Rotation from $\mathbf{L}$ to $\mathbf{L}^* \equiv \mathbf{LH}$ rotates the loadings. There are p points in m-dimensional space defined by the rows $\mathbf{l}_k' = [\ell_{k1}, \ell_{k2},\ldots,\ell_{km}]'$ of $\mathbf{L}$, <u>one for each variable</u>.

When $\mathbf{H}$ is a rotation matrix, the change $L \rightarrow \mathbf{L}^* = LH$ describes a rigid rotation of points in that space with each $\mathbf{l}_k \rightarrow \mathbf{l}_k^* = \mathbf{H}'\mathbf{l}_k$, k = 1,..., p.

Example with m = 2 and p = 5.



Thus the factor analytic decomposition of $\mathbf{\Sigma}$ in terms of $\mathbf{\Psi}$ and $\mathbf{L}$ (or of $\mathbf{x}$ in terms of $\mathbf{L}$, $\mathbf{f}$, and $\boldsymbol{\varepsilon}$) *is not unique*.

Question

If $\mathbf{L}$ and $\mathbf{f}$ are not unique, what, if anything, *is* unique?

Answer

The decomposition $\mathbf{\Sigma} = \mathbf{V} + \mathbf{\Psi}$, rank m $\mathbf{V}$ and diagonal $\mathbf{\Psi}$

You can estimate $\mathbf{V}$ and $\mathbf{\Psi}$ from data in an *unambiguous* manner.

You can estimate $\mathbf{L}$ unambiguously *only* when you introduce some further principles or restrictions to eliminate the non-uniqueness.

Thus, the factor extraction stage is the process of estimating $\mathbf{V}$ and $\mathbf{\Psi}$. Usually $\mathbf{V}$ is estimated by finding an $\hat{\mathbf{L}}$ and setting $\hat{\mathbf{V}} = \hat{\mathbf{L}}\hat{\mathbf{L}}'$, but $\hat{\mathbf{L}}$ cannot be interpreted.