Displays for Statistics 5401/8401

Lecture 29

November 14, 2005

Christopher Bingham, Instructor

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

Class Web Page

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

There are two important and different types of factor analysis - **exploratory factor analysis** and **confirmatory factor analysis**.

Exploratory factor analysis estimates the factor analysis model with some purely mathematical restriction (e.g. $\mathbf{L}'\mathbf{L}$ diagonal) on the loading matrix $\mathbf{L}$ to assure uniqueness. Then $\hat{\mathbf{L}}$ is "rotated" to have an "interpretable" structure.

For example a rotated $\hat{\mathbf{L}}$ might be

$$\hat{\mathbf{L}} = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array} \begin{array}{c} \begin{array}{cc} f_1 & f_2 \end{array} \\ \begin{bmatrix} .99 & -.08 \\ -.10 & -.81 \\ .09 & .58 \end{bmatrix} \end{array} \approx \begin{bmatrix} 1 & 0 \\ 0 & -.8 \\ 0 & .6 \end{bmatrix}$$

Interpretation:

- x_1 is affected only by factor 1
- x_2 and x_3 are affected only by factor 2.

You use **confirmatory factor analysis** with a pre-determined form for $\mathbf{L}$ like

$$\mathbf{L} = \begin{bmatrix} * & 0 \\ 0 & * \\ 0 & * \end{bmatrix} \quad (* \text{ means non-zero})$$

The pattern might come from an earlier exploratory factor analysis or from subject-matter knowledge.

You estimate $\mathbf{L}$ by finding the best fitting loadings *of this form*, that is with $\ell_{12} = \ell_{21} = \ell_{31} = 0$.

A goodness-of-fit test of $\hat{\mathbf{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}$ to $\mathbf{S}$ or $\hat{\boldsymbol{\rho}}$ to $\mathbf{R}$ tests the null hypothesis

$$H_0: \ell_{12} = \ell_{21} = \ell_{31} = 0$$

vs

$$H_1: \text{no special pattern of 0's.}$$

When you cannot reject H_0, then the proposed pattern is *confirmed*.

Confirmatory factor analysis is a special case of **structural equation modeling**.

My focus is on exploratory factor analysis.

Recapitulate

The factor analytic model for $\mathbf{x}$ is

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{L}\mathbf{f} + \boldsymbol{\varepsilon}.$$

$\mathbf{\Sigma} = V[\mathbf{X}]$ has the factor analytic form

$$\mathbf{\Sigma} = \underset{\text{rank } m}{\mathbf{V}} + \underset{\text{diagonal}}{\mathbf{\Psi}} = \mathbf{L}\mathbf{\Gamma}\mathbf{L}' + \mathbf{\Psi}, \ \mathbf{\Gamma} = V[\mathbf{f}]$$

where

- $\mathbf{\Psi} = \text{Var}[\boldsymbol{\varepsilon}] = \text{diag}[\psi_1, \psi_2, \ldots, \psi_p]$ is p by p diagonal with $\psi_i \geq 0$
- $\mathbf{L}$ is p by m
- $\mathbf{\Gamma} = V[\mathbf{f}]$ is m by m
- $\mathbf{V} = \mathbf{L}\mathbf{\Gamma}\mathbf{L}'$ has rank m < p

Only $\mathbf{V}$ and $\mathbf{\Psi}$ are uniquely defined. Given $\mathbf{V}$ and $\mathbf{\Psi}$, there are infinitely many p by m $\mathbf{L}$ and m by m $\mathbf{\Gamma}$ with $\mathbf{V} = \mathbf{L}\mathbf{\Gamma}\mathbf{L}'$.

$\mathbf{\Gamma} = \mathbf{I}_m$ characterizes the orthogonal factor model.

Exploratory factor analysis usually consists of two phases

1 *Factor extraction* - compute estimates $\hat{\Psi}$ and $\hat{V} = \hat{L}\hat{L}'$, where $\hat{L}$ satisfies a mathematical restriction to achieve uniqueness and may not be interpretable.

This is followed by

2 *Factor rotation* - modifies $\hat{L}$ without changing $\hat{V} = \hat{L}\hat{L}'$ so as to try to achieve an interpretable form for $\hat{L}$.

With oblique factor analysis, $\hat{\Gamma}$ and $\hat{L}$ are found so that $\hat{V} = \hat{L}\hat{\Gamma}\hat{L}'$ and $\hat{L}$ has interpretable structure and $\hat{\Gamma}$ provides information about how the factors are related. You can often get a "simpler" $\hat{L}$ than with $\Gamma = I_m$.

You start with the sample variance matrix S or correlation matrix R.

Usually

$$S = (n-1)^{-1}\sum_{1\le i\le n}(x_i - \bar{x})(x_i - \bar{x})'$$

comes from a random sample $x_1, \ldots, x_n$.

S can also be a "pooled" matrix

$$S = f_e^{-1}E$$

where E is the error matrix from a MANOVA or MANACOVA.

The estimated correlation matrix is

$$R = DSD$$

with

$$D = \mathrm{diag}[1/\sqrt{s_{11}}, 1/\sqrt{s_{22}}, \ldots, 1/\sqrt{s_{pp}}].$$

For two methods of estimation, GLS and MLE (but not ULS), you get essentially the same results whether you start with S or R. This contrasts strongly with Principal Components Analysis.

In **covariance-based** *factor extraction* you try to approximate S by finding

$$p \text{ by } m\ \hat{L} \text{ and diagonal } p \text{ by } p\ \hat{\Psi}$$

such that

$$[\hat{\sigma}_{ij}] = \hat{\Sigma} \equiv \hat{L}\hat{L}' + \hat{\Psi} = \hat{V} + \hat{\Psi} \cong S$$

When the factor analytic model is correct, $\hat{\Sigma}$ will be a "better" estimate of Σ than S, because $\hat{\Sigma}$ has factor analytic structure while S does not.

Correlation-based factor analysis tries to approximate

$$R = \begin{bmatrix} 1 & r_{12} & r_{13} & \ldots & r_{1p} \\ r_{21} & 1 & r_{23} & \ldots & r_{2p} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ r_{p1} & r_{p2} & \ldots & r_{p,p-1} & 1 \end{bmatrix}, \quad r_{ij} = r_{ji}$$

by a matrix of the form $\hat{\rho} \equiv \hat{L}\hat{L}' + \hat{\Psi}$ which should be a better estimator of ρ than R.

Usually factor analysis starts with R.

Because $\hat{V} = \hat{L}\hat{L}'$ appears in the approximations for Σ and ρ, the factor extraction phase is based on an orthogonal factor model, but this is only for mathematical convience.

Finding $\hat{L}$ in this phase is just the way you find the rank m piece $\hat{V} = \hat{L}\hat{L}'$ of the factor analytic estimate $\hat{\Sigma} = \hat{V} + \hat{\Psi}$ or $\hat{\rho} = \hat{V} + \hat{\Psi}$.

I will focus on methods starting with S. They can be applied to R by substituting ρ and R for Σ and S in what follows.

The most usual mathematical restrictions on $\hat{L}$ are

- $\hat{L}'\hat{L}$ is diagonal (columns of $\hat{L}$ are orthogonal), assumed for **ULS** (unweighted least squares) estimation
- $\hat{L}'\hat{\Psi}^{-1}\hat{L}$ is diagonal (columns of $\hat{\Psi}^{-1/2}\hat{L}$ are orthogonal), assumed for **GLS** (generalized least squares) and **MLE** (maximum likelihood) estimation.

Neither restriction on $\hat{L}$ appears to have an interpretable meaning.

Both are particular cases (when $Q = I_p$ and $Q = \hat{\Psi}$) of

- $L'Q^{-1}L$ is diagonal
- Q positive definite, possibly depending on $\hat{\Psi}$.

Fact:
Suppose you have found $\hat{V}$ by some means or other. Then you can find $\hat{L}_Q$ such that $\hat{V} = \hat{L}_Q\hat{L}_Q'$ and $\hat{L}_Q$ satisfies

$$\hat{L}_Q'Q^{-1}\hat{L}_Q \text{ is diagonal}$$

where Q is a specific positive definite symmetric matrix, say $\hat{\Psi}$.

$$\hat{L}_Q = [\hat{\ell}_{1Q}, \ldots, \hat{\ell}_{mQ}],\ \hat{\ell}_{iQ} = \sqrt{\delta_i}Qe_i$$

where

- $e_1, \ldots, e_m$ are the first m eigenvectors of $\hat{V}$ *relative* to Q
- $\delta_1, \ldots, \delta_m$ the first m relative eigenvalues

With this choice $\hat{V} = \hat{L}_Q\hat{L}_Q'$ and

$$\hat{L}_Q'Q^{-1}\hat{L}_Q = \text{diag}[\delta_1, \delta_2, \ldots, \delta_m]$$

is diagonal.

1. Diagonal $\hat{L}'\hat{L}$ corresponds to $Q = I_p$ when relative eigenvalues and vectors are ordinary eigenvalues $\hat{\lambda}_i$ and eigenvectors $\hat{v}_i$ of $\hat{V}$, so the columns of $\hat{L}$ are $\hat{\ell}_j = \sqrt{\hat{\lambda}_i}\hat{v}_i$.
2. Diagonal $\hat{L}'\hat{\Psi}^{-1}\hat{L}$ corresponds to $Q = \hat{\Psi}$. In this case $\hat{\ell}_j = \sqrt{\gamma_i}u_i$, where γ_i and u_i are eigenvalue and eigenvector of $\hat{V}$ *relative* to $\hat{\Psi}$.

We will look at several ways to estimate V, including the ULS, GLS and MLE (or ML) methods

ULS assumes $L'L$ is diagonal ($Q = I_p$).

GLS and ML assumes $L'\Psi^{-1}L$ is diagonal ($Q = \Psi$).

To summarize exploratory factor analysis:

Phase 1: ***Extraction***:
$(x_1,\ldots,x_n) \rightarrow (S \text{ or } R) \rightarrow (\hat{L},\hat{\Psi}) \rightarrow (\hat{V}, \hat{\Psi})$,
$\hat{V} = \hat{L}\hat{L}'$
where $\hat{L}$ satisfies a mathematical restriction that makes it unique.

Phase 2: ***Rotation***:
Find H (and possibly $\hat{\Gamma}$) so that
$\hat{L}^* = \hat{L}H$ is interpretable
with $\hat{V} = \hat{L}^*\hat{L}^{*\prime}$, (or $\hat{V} = \hat{L}^*\hat{\Gamma}\hat{L}^{*\prime}$)

The choice of $\hat{L}^*$ (or H) must be based on substantive scientific reasons, not mathematical convenience.

All methods of rotation share the idea of finding $\mathbf{H}$ so that $\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{H}$ has *simple structure*.

The left matrix is "not simple" (* means non-zero); the right matrix would generally be considered simpler

$$\hat{\mathbf{L}} = \begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \end{bmatrix}, \quad \hat{\mathbf{L}}^* = \begin{bmatrix} * & 0 & 0 \\ * & 0 & 0 \\ * & * & 0 \\ 0 & * & 0 \\ 0 & 0 & * \end{bmatrix}$$

- In $\hat{\mathbf{L}}$ *all* p = 5 variables load on *all* m = 3 factors. This is <u>not simple</u>.
- In $\hat{\mathbf{L}}^*$, variables x_1 and x_2 load *only* on f_1, x_3 loads on f_1 *and* f_2, x_4 loads *only* on f_2, and x_5 loads *only* on f_3. And factor f_1 affects only x_1, x_2 and x_3, f_2 affects only x_3 and x_4, and f_3 affects only x_5.

The simpler pattern *might* be more interpretable *by a subject matter expert*, especially if you can identify a feature that is shared by x_1 and x_2.

For example, if both x_1 and x_2 were concerned with a subject's <u>logical thinking</u> skills, then it *might* be possible to identify f_1 as a "clear thinking" factor. You would need subject matter knowledge to do this with confidence.

Most computational rotation methods such as *varimax* and *quartimax* find $\mathbf{H}$

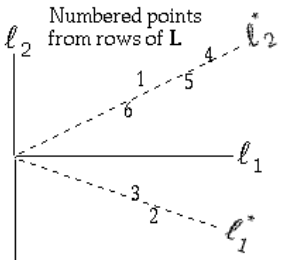
- to *maximize* $I_S(\hat{\mathbf{L}}\mathbf{H})$ where $I_S(\mathbf{L})$ is a "simplicity" index
- or to *minimize* $I_C(\hat{\mathbf{L}}\mathbf{H})$, where $I_C(\mathbf{L})$ is a "complexity" index.

For varimax and quartimax the criterion function is a degree 4 polynomial in ℓ_{ij}.

For oblique (non-orthogonal) factor rotation, you "rotate" with a matrix $\mathbf{H}$ that is not a rotation matrix, that is $\mathbf{H}'\mathbf{H} \neq \mathbf{I}_m$ and $\mathbf{\Gamma} = V[\mathbf{f}] \neq \mathbf{I}_m$.

Here is an m = 2 example when this might be appropriate. No *orthogonal* rotation will produce simple form.

Using coordinates based on the non perpendicular dashed lines, variables 3 and 2 will load primarily on factor 1 and variables 1, 4, 5, and 6 on factor 2, a simple structure.



That is, $\mathbf{L}^*$ has the form

$$\mathbf{L}^* = \begin{bmatrix} 0 & * \\ * & 0 \\ * & 0 \\ 0 & * \\ 0 & * \\ 0 & * \end{bmatrix}$$

The rotated factors f_j^* will *not* be uncorrelated ($\mathbf{\Gamma}$ not diagonal).

We will look at four or five methods of factor extraction:

- PC = Principal Components
- IPF = Iterated principal factor
- ULS = Unweighted Least Squares
- GLS = Generalized Least Squares
- ML = Maximum likelihood assuming normality.

All except PC require <u>iteration</u>.

IPF is sometimes an option on computer programs. It is a simple iterative method that should lead to the same solution as the ULS method. IPF is generally not recommended because it may need a huge number of steps. There are similar inferior iterative methods for GLS and MLE.

ULS, GLS and MLE estimates each attempt to <u>minimize a specific criterion</u> which measures the discrepancy between $\mathbf{S}$ (or $\mathbf{R}$) and the estimate $\hat{\mathbf{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}$.

Principal Component Factor Estimation

The PC method approximates $\mathbf{x}$ in terms of the first few sample principal components $\hat{z}_j = \hat{\mathbf{v}}_j'(\mathbf{x} - \overline{\mathbf{x}})$

Summary of method

- Find the eigenvectors $\hat{\mathbf{v}}_j$ and eigenvalues $\hat{\lambda}_j$, j = 1,...,p of the sample covariance matrix $\mathbf{S}$ or (more usually) the sample correlation matrix $\mathbf{R}$.
- For a suitable choice of m (usually the number of eigenvalues greater than a threshold such as 1), the *unrotated* estimated loading matrix is

$$\hat{\mathbf{L}} = [\sqrt{\hat{\lambda}_1}\hat{\mathbf{v}}_1,\ldots,\sqrt{\hat{\lambda}_m}\hat{\mathbf{v}}_m] = [\hat{\boldsymbol{\ell}}_1, \ldots, \hat{\boldsymbol{\ell}}_m],\ \hat{\ell}_{ij} = \sqrt{\hat{\lambda}_j}\times\hat{v}_{ij}$$

  Because eigenvectors are orthonormal, $\hat{\mathbf{L}}'\hat{\mathbf{L}} = \mathrm{diag}[\hat{\lambda}_1,\ldots,\hat{\lambda}_m]$ is diagonal.
- The estimated factors are $\hat{f}_j = \hat{z}_j/\sqrt{\hat{\lambda}_j}$

- You estimate $\mathbf{V}$ by $\hat{\mathbf{V}} = \hat{\mathbf{L}}\hat{\mathbf{L}}'$. $\hat{\mathbf{V}}$ is the best (least squares) rank m approximation to $\mathbf{S}$ (or $\mathbf{R}$).
- You find $\hat{\boldsymbol{\Psi}}$ so that the diagonal of $\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}$ *exactly matches* the diagonal of $\mathbf{S}$. That is

  $\hat{\boldsymbol{\Psi}} = \mathrm{diag}[\mathbf{S} - \hat{\mathbf{L}}\hat{\mathbf{L}}'],\ = \mathrm{diag}[\mathbf{S} - \hat{\mathbf{V}}]$

  or explicitly

$$\hat{\psi}_i = s_{ii} - \sum_{1\le j\le m}\hat{\ell}_{ij}^2,\ i = 1,\ldots,p.$$

The estimated communalities are

$$\hat{h}_i^2 = s_{ii} - \hat{\psi}_i = \sum_{1\le j\le m}\hat{\ell}_{ij}^2.$$

When $\mathbf{S} = \mathbf{R}$, $\hat{h}_i^2 = 1 - \hat{\psi}_i$

Notation note:
When $\mathbf{A} = [a_{ij}]$, the notation diag[$\mathbf{A}$] means the diagonal matrix $\mathrm{diag}[a_{11},\ldots,a_{pp}]$.

Example: Artifical data with m = 2

```
Cmd> s <- tabs(y,covar:T); s # sample variance matrix
(1,1)     3.1903      0.6777     -2.2794     -0.62891     0.71555
(2,1)     0.6777      1.3874     -1.2398     -0.3807      0.29635
(3,1)    -2.2794     -1.2398      7.7395      1.7365     -1.8761
(4,1)   -0.62891     -0.3807      1.7365     0.73894    -0.56029
(5,1)    0.71555     0.29635     -1.8761    -0.56029     0.87565

Cmd> eigs <- eigen(s)

Cmd> eigs$values # eigenvalues of S
(1)      9.8495      2.2704      1.1149     0.46384     0.23323

Cmd> m <- 2

Cmd> Lhat <- sqrt(eigs$values[run(m)]') * eigs$vectors[,run(m)]

Cmd> Lhat # unrotated estimated loadings
(1,1)       1.1223       1.3747
(2,1)      0.54182      0.19703
(3,1)      -2.7107      0.57841
(4,1)     -0.66079     0.062869
(5,1)      0.71535    -0.056138

Cmd> psihat <- diag(s) - diag(Lhat %*% L'); psihat
(1)    0.040853       1.055    0.056965     0.29834     0.36078
```

Columns are $\sqrt{\hat{\lambda}_j}\hat{\mathbf{v}}_j$

These are the estimated uniquenesses $\hat{\psi}_j$.

```
Cmd> Vhat <- Lhat %*% Lhat'; Vhat # rank 2 piece
(1,1)     3.1495     0.87894      -2.247    -0.65517     0.72565
(2,1)    0.87894     0.33239     -1.3547    -0.34564     0.37653
(3,1)     -2.247     -1.3547      7.6826      1.8276     -1.9716
(4,1)   -0.65517    -0.34564      1.8276     0.44059    -0.47622
(5,1)    0.72565     0.37653     -1.9716    -0.47622     0.51488

Cmd> h <- diag(vhat); h #c Estimated omunalities
(1)      3.1495     0.33239      7.6826     0.44059     0.51488

Cmd> sigmahat <- Vhat + dmat(psihat)

Cmd> sigmahat # Estimated variance matrix;diags exactly match S
(1,1)     3.1903     0.87894      -2.247    -0.65517     0.72565
(2,1)    0.87894      1.3874     -1.3547    -0.34564     0.37653
(3,1)     -2.247     -1.3547      7.7395      1.8276     -1.9716
(4,1)   -0.65517    -0.34564      1.8276     0.73894    -0.47622
(5,1)    0.72565     0.37653     -1.9716    -0.47622     0.87565
```


Advantages of PC factor estimation

- Estimated uniqueness $\hat{\psi}_i$ and unrotated loadings $\hat{\mathbf{L}}$ are explicit functions of eigenvalues and eigenvectors of $\mathbf{S}$ or $\mathbf{R}$. *No iteration is required.*
- You can obtain loadings for any $m \le p$
- $\hat{\boldsymbol{\ell}}_j$ doesn't change when m changes.

Disadvantage:

- The PC method does not actually estimate either $\boldsymbol{\Psi}$ or $\mathbf{V} = \mathbf{L}\mathbf{L}'$.

  Even when $\boldsymbol{\Sigma} = \mathbf{V} + \boldsymbol{\Psi}$ and $\mathbf{S} = \boldsymbol{\Sigma}$, the principal component method does not reproduce $\mathbf{V}$ and $\boldsymbol{\Psi}$. This is because in the Principal Component "model"

  $\mathbf{x} = \boldsymbol{\mu} + \sum_{1\le j\le m} f_j\boldsymbol{\ell}_j + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} = \sum_{m+1\le j\le p} z_j\mathbf{v}_j$

  $f_j = z_j/\sqrt{\lambda_j},\ \boldsymbol{\ell}_j = \sqrt{\lambda_j}\mathbf{v}_j.$

But $V[\boldsymbol{\varepsilon}] = \sum_{m+1\le j\le p}\lambda_j\mathbf{v}_j\mathbf{v}_j'$ is *not* diagonal.

Iterated Principal Factor Estimation

The *iterated principal factor* (IPF) method of factor extraction is iterative. Each step or stage is quite similar to the PC method.

At the i^{th} stage you have a trial value $\hat{\Psi}_{(i)}$. If $\hat{\Psi}_{(i)}$ is close to the true Ψ then $V^* = S - \hat{\Psi}_{(i)}$ should be close to the true rank m matrix V.

Then you find a p by m $\hat{L}_{(i+1)}$ such that $\hat{V}_{(i+1)} = \hat{L}_{(i+1)}\hat{L}_{(i+1)}'$ is the best rank m approximation to $V_{(i)}^* \equiv S - \hat{\Psi}_{(i)}$.

But this is

$$\hat{L}_{(i+1)} = [\sqrt{\delta_1}\mathbf{e}_1, \ldots, \sqrt{\delta_m}\mathbf{e}_m].$$

where δ_j and $\mathbf{e}_j$ are the eigenvalues and eigenvectors of $V_{(i)}^*$. This is like the PC method but with $V_{(i)}^*$ replacing S.

Because the eigenvectors $\mathbf{e}_j$ are orthonormal, $\hat{L}_{(i+1)}'\hat{L}_{(i+1)} = \text{diag}[\delta_1, \ldots, \delta_m]$ and thus is diagonal.

You then get a new $\hat{\Psi}$ as

$$\hat{\Psi}_{(i+1)} = \text{diag}[S - \hat{L}_{(i+1)}\hat{L}_{(i+1)}']$$

just as in the PC method.

To summarize: At the i^{th} iteration:

- $\hat{\Psi}_{(i)}$ and $\hat{L}_{(i)}$ are current trial values.
- $\hat{\Sigma}_{(i)} = \hat{L}_{(i)}\hat{L}_{(i)}' + \hat{\Psi}_{(i)} = \hat{V}_{(i)} + \hat{\Psi}_{(i)}$ is the current approximation to S based on $\hat{\Psi}_{(i)}$ and $\hat{L}_{(i)}$.

After iteration i+1

- $\hat{L}_{(i+1)} = [\sqrt{\delta_1}\mathbf{e}_1, \ldots, \sqrt{\delta_m}\mathbf{e}_m]$, δ_j and $\mathbf{e}_j$ eigenvalues and vectors of $V_{(i)}^* = S - \hat{\Psi}_{(i)}$
- $\hat{\Psi}_{(i+1)} = \text{diag}[S - \hat{L}_{(i+1)}\hat{L}_{(i+1)}']$.

Substitute R for S when working with correlations.

You continue the iteration until it converges (if it does), that is,

$$\hat{\Psi}_{(i+1)} - \hat{\Psi}_{(i)} \approx 0 \text{ or } \hat{\Sigma}_{(i+1)} - \hat{\Sigma}_{(i)} \approx 0.$$

In practice, it may converge very slowly. Or it can abort if at some point $\delta_m < 0$ so that $\sqrt{\delta_m}$ can't be calculated.

Note:

- At each stage, $\hat{\Psi}_{(i)}$ is *all* you need to go to the next stage, since the next stage is computed from the eigenvalues and vectors of $V^* = S - \hat{\Psi}_{(i)}$.
- To start the iteration, you must provide an *initial value* $\hat{\Psi}_{(0)}$ for Ψ.

The most usual is

$$\hat{\Psi}_{(0)} = \text{diag}[1/s^{11}, 1/s^{22}, \ldots, 1/s^{pp}]$$

where $S^{-1} = [s^{k\ell}]$.

$1/s^{kk}$ is essentially $1 - R^2$ in a regression of x_k on the other x's.