Factor Analysis:

Objective: Describe covariance relationships among a large set of measured traits with a few linear combinations of underlying but unobservable traits.

$$\underset{\sim}{X}_i = \begin{bmatrix} X_{1i} \\ X_{2i} \\ \vdots \\ X_{pi} \end{bmatrix}$$

denotes the set of p measurements recorded for the i-th experimental unit

$$\underset{\sim}{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \quad \text{and} \quad Var(X) = \Sigma_{p \times p}$$

The big idea is to construct a linear model:

$$\underbrace{\begin{bmatrix} X_{1i} - \mu_1 \\ X_{2i} - \mu_2 \\ \vdots \\ X_{pi} - \mu_p \end{bmatrix}}_{p \text{ measurements}} = \underbrace{\begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1m} \\ l_{21} & l_{22} & \cdots & l_{2m} \\ \vdots & & & \vdots \\ l_{p1} & l_{p2} & \cdots & l_{pm} \end{bmatrix}}_{\text{matrix of factor loadings}} \begin{bmatrix} F_{1i} \\ F_{2i} \\ \vdots \\ F_{mi} \end{bmatrix} + \begin{bmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \vdots \\ \epsilon_{pi} \end{bmatrix}$$

unobserved values of factors which describe major features of members of the population

random "errors" corresponding to measurement error and variation not accounted for by the common factors (variation in specific factors).

Note that we want

$m =$ number of factors

to be much smaller than

$p =$ number of measured attributes.

Matrix of factor loadings

$$L = \begin{bmatrix} \ell_{11} & \ell_{12} & \cdots & \ell_{1j} & \cdots & \ell_{1m} \\ \ell_{21} & \ell_{22} & \cdots & \ell_{2j} & \cdots & \ell_{2m} \\ \vdots & & & & & \\ \ell_{p1} & \ell_{p2} & \cdots & \ell_{pj} & \cdots & \ell_{pm} \end{bmatrix}_{p \times m}$$

j-th column is the vector of loadings for the j-th factor

ℓ_{ij} is called the loading of the i-th variable on the j-th factor

$$\underset{\sim}{X}_i - \underset{\sim}{\mu} = L \underset{\sim}{F}_i + \underset{\sim}{\epsilon}_i$$

Restrictions:

$$E(\underset{\sim}{\epsilon}_i) = \underset{\sim}{0}$$

$$Var(\underset{\sim}{\epsilon}_i) = \Psi_{p \times p} = \begin{bmatrix} \psi_1 & & & \\ & \psi_2 & & \\ & & \ddots & \\ & & & \psi_p \end{bmatrix}$$

$\underset{\sim}{\epsilon}_i$'s and $\underset{\sim}{F}_i$'s are independent.

Orthogonal Factor model:

$$E(\underset{\sim}{F}_i) = \underset{\sim}{0}$$

$$Var(\underset{\sim}{F}_i) = I_{m \times m} = \begin{bmatrix} 1 & & \\ & 1 & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$$

The real restriction is that the factors are uncorrelated. Variances can be made equal to 1 by properly scaling the factor loadings

[an alternative is the Oblique factor model]

This model imposes a covariance structure on $\underset{\sim}{X}$

1) $\Sigma = Var(\underset{\sim}{X}) = Var(L\underset{\sim}{F} + \underset{\sim}{\epsilon})$

$= Var(L\underset{\sim}{F}) + Var(\underset{\sim}{\epsilon})$

$= L I L' + \Psi$

$= LL' + \Psi$

the off diagonal elements of LL' are σ_{ij}, the covariances in Σ.

$$Var(X_i) = \sum_{j=1}^{m} \ell_{ij}^2 + \Psi_i$$

$$Cov(X_i, X_j) = \sum_{k=1}^{m} \ell_{ik}\ell_{jk}$$

2) $Cov(\underset{\sim}{X}, \underset{\sim}{F}) = Cov(L\underset{\sim}{F}, \underset{\sim}{F}) = L$

so $Cov(X_i, F_j) = \ell_{ij}$

The proportion of the variance of the i-th measurement X_i contributed by the m factors $F_1, F_2, \ldots F_m$ is called the i-th communality.

The remaining proportion of the variance of the i-th measurement, associated with ϵ_i, is called the uniqueness or specific variance.

$$\underbrace{\sigma_{ii}}_{Var(X_i)} = \underbrace{\sum_{k=1}^{m} \ell_{ik}^2}_{\text{communality, call it } h_i^2} + \underbrace{\Psi_i}_{\text{specific variance}}$$

Then

$$\sigma_{ii} = \underline{h_i^2} + \underline{\Psi_i}$$

Limitations of the orthogonal factor model:

1. Linearity:

For a non-linear model, say

$$(X_i - \mu_i) = \left[\alpha_1 F_1 F_3 + \alpha_2 \ln(F_2)\right]^2 + \epsilon_i,$$

the covariance approximation $LL' + \Psi$ may be quite inadequate.

The amount of observed data is <u>not</u> adequate to check this assumption of linearity.

Linear combinations of the factors may provide a good approximation for non-linear relationships over a small range of factor values

2. The $\frac{p(p+1)}{2}$ elements of Σ are described with mp factor loadings in L and p specific variances $\{\psi_i\}$

The factor model is most useful when m is small, but in many cases $mp + p$ parameters are not adequate and Σ is not close to $LL' + \Psi$.

when $p = 12$, $\frac{p(p+1)}{2} = 78$

but for $m = 2$ factors,

$mp + p = 36$.

Look at example 9.2 in J+W for a 3x3 covariance matrix which cannot be exactly described by a one factor model.

3. The factor model is determined uniquely up to an orthogonal transformation of the factors.

Let $T_{m \times m}$ be an orthogonal matrix

Consider

$$\underset{\sim}{X} = L \underset{\sim}{F} + \underset{\sim}{\epsilon}$$

$$= \underbrace{L T} T' \underset{\sim}{F} + \underset{\sim}{\epsilon}$$

$$= L^* (T' \underset{\sim}{F}) + \underset{\sim}{\epsilon}$$

then

$$\Sigma = L L' + \Psi$$

$$= L T T' L' + \Psi = (L^*)(L^*)' + \Psi$$

where $L^* = LT$

Methods of Estimation:

L matrix of factor loadings (mp parameters)

$$\Psi = \begin{bmatrix} \psi_1 & & \\ & \ddots & \\ & & \psi_p \end{bmatrix}$$ specific variances (p parameters)

We will also estimate scores for the unobservable factors.

Obtain a random sample of n members of the population. Measure p attributes of each sampled member.

$$\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots, \underset{\sim}{X}_n \quad \text{where } \underset{\sim}{X}_j = \begin{bmatrix} X_{1j} \\ \vdots \\ X_{pj} \end{bmatrix}$$

$$S = \frac{1}{n-1} A = \frac{1}{n-1} \sum_{j=1}^{n} (\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})'$$

If standardized measurements

$$\underset{\sim}{z}_j = \begin{bmatrix} (x_{1j} - \bar{x}_1)/\sqrt{s_{11}} \\ \vdots \\ (x_{pj} - \bar{x}_p)/\sqrt{s_{pp}} \end{bmatrix}$$

are used, replace S with R, the matrix of sample correlations.

Methods of estimation:

1. Principal Component method
2. Principal Factor method
3. Maximum Likelihood method

Principal Component Method:

Spectral decomposition of $\Sigma_{p \times p}$

$$\Sigma = \lambda_1 \underset{\sim}{e}_1 \underset{\sim}{e}_1' + \lambda_2 \underset{\sim}{e}_2 \underset{\sim}{e}_2' + \cdots + \lambda_p \underset{\sim}{e}_p \underset{\sim}{e}_p'$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ are the eigenvalues for Σ.

Suppose $\lambda_1, \lambda_2, \ldots, \lambda_m$ are "large" and $\sum_{k=m+1}^{p} \lambda_k$ is "small." Then

$$\Sigma \doteq \sum_{k=1}^{m} \lambda_k \underset{\sim}{e}_k \underset{\sim}{e}_k'$$

$$= \left[\sqrt{\lambda_1}\, \underset{\sim}{e}_1 \quad \sqrt{\lambda_2}\, \underset{\sim}{e}_2 \quad \cdots \quad \sqrt{\lambda_m}\, \underset{\sim}{e}_m\right] \begin{bmatrix} \sqrt{\lambda_1}\, \underset{\sim}{e}_1' \\ \sqrt{\lambda_2}\, \underset{\sim}{e}_2' \\ \vdots \\ \sqrt{\lambda_m}\, \underset{\sim}{e}_m' \end{bmatrix}$$

$$= L L'$$

Then from the factor model

$$\Sigma = LL' + \Psi$$

we have

$$\begin{bmatrix} \psi_1 & & \\ & \ddots & \\ & & \psi_m \end{bmatrix} \doteq \Sigma - \left[\sqrt{\lambda_1}\,\underset{\sim}{e}_1 \cdots \sqrt{\lambda_m}\,\underset{\sim}{e}_m\right] \begin{bmatrix} \sqrt{\lambda_1}\,\underset{\sim}{e}_1' \\ \vdots \\ \sqrt{\lambda_m}\,\underset{\sim}{e}_m' \end{bmatrix}$$

so

$$\psi_i = \sigma_{ii} - \sum_{j=1}^{m} \ell_{ij}^2 = \sigma_{ii} - \sum_{j=1}^{m} \lambda_j e_{ij}^2$$

$$= \sigma_{ii} - h_i^2$$

Estimate L and Ψ by substituting estimated eigenvectors and eigenvalues from S or R.

$$\tilde{L} = \left[\sqrt{\hat{\lambda}_1}\,\hat{\underset{\sim}{e}}_1 \,\vdots\, \sqrt{\hat{\lambda}_2}\,\hat{\underset{\sim}{e}}_2 \,\vdots\, \cdots \,\vdots\, \sqrt{\hat{\lambda}_m}\,\hat{\underset{\sim}{e}}_m\right]$$

$$\tilde{\Psi} = \begin{bmatrix} \tilde{\psi}_1 & & & \\ & \tilde{\psi}_2 & & \\ & & \ddots & \\ & & & \tilde{\psi}_m \end{bmatrix} = \operatorname{diag}(S) - \operatorname{diag}(\tilde{L}\tilde{L}')$$

Estimated specific variance:

$$\tilde{\psi}_i = s_{ii} - \sum_{j=1}^{m} \hat{\lambda}_j \hat{e}_{ij}^2$$

Estimated communalities:

$$\tilde{h}_i^2 = \sum_{j=1}^{m} \tilde{\ell}_{ij}^2 = \sum_{j=1}^{m} \hat{\lambda}_j \hat{e}_{ij}^2$$

$$= s_{ii} - \tilde{\psi}_i$$

Example 9.4 (Page 489 in J+W)

Stock Price data consisting of $n = 100$ weekly rates of returns on $p = 5$ stocks.

The data were standardized and the factor analysis was performed on the sample correlation matrix R.

$$
\begin{array}{l}
\text{lied Chem} \\
\text{Du pont} \\
\text{ion Carb.} \\
\text{Exxon} \\
\text{Texaco}
\end{array}
\begin{bmatrix}
1 & .58 & .51 & .39 & .46 \\
 & 1 & .60 & .39 & .32 \\
 & & 1 & .44 & .42 \\
 & & & 1 & .52 \\
 & & & & 1
\end{bmatrix}
$$

Two factor ($m=2$) solution:

Loadings

Variable	Factor 1	Factor 2	Specific Variances $\tilde{\psi}_i = 1 - \tilde{h}_i^2$
Allied Chem	.783	−.217	.34
Dupont	.773	−.458	.19
Union Carbide	.794	−.234	.31
Exxon	.713	.472	.27
Texaco	.712	.524	.22
Cummulative proportion of total (standardized) sample variance	.571	.733	

Residual Matrix:

$$
R - (\tilde{L}\tilde{L}' + \tilde{\Psi}) = \begin{bmatrix}
0 & -.13 & -.16 & -.07 & .02 \\
 & 0 & -.12 & .06 & .01 \\
 & & 0 & -.02 & -.02 \\
 & & & 0 & -.23 \\
 & & & & 0
\end{bmatrix}
$$

Comments:

1. Estimated loadings on a factor do not change as the number of factors is increased.

2. Diagonal elements of S (or R) are exactly equal to the diagonal elements of $\tilde{L}\tilde{L}' + \tilde{\Psi}$, but sample covariances may not be exactly reproduced.

   Select the number of factors m to make off-diagonal elements <u>small</u> for the residual matrix

$$S - [\tilde{L}\tilde{L}' + \tilde{\Psi}]$$

Note that the sum of the squared elements of $S - (\tilde{L}\tilde{L}' + \tilde{\Psi})$ is $\leq \sum_{k=m+1}^{p} \hat{\lambda}_k^2$

choose m big enough to make this sum very small.

3. Contribution of the k-th factor to the total variance $tr(S) = \sum_{i=1}^{p} \hat{\lambda}_i$ is

$$\hat{L}_k' I \hat{L}_k = [\sqrt{\hat{\lambda}_k}\, \hat{e}_k]' [\sqrt{\hat{\lambda}_k}\, \hat{e}_k] = \hat{\lambda}_k$$

$$\begin{pmatrix} \text{Proportion of total} \\ \text{sample variance} \\ \text{due to } k\text{-th factor} \end{pmatrix} = \begin{cases} \dfrac{\hat{\lambda}_k}{tr(S)} & \text{using } S \\[2ex] \dfrac{\hat{\lambda}_k}{p} & \text{using } R \end{cases}$$

Principal Factor Method:

Consider the model for the correlation matrix

$$R = LL' + \Psi$$

Then

$$LL' = R - \Psi = \begin{bmatrix} h_1^2 & r_{12} & \cdots & r_{1p} \\ & h_2^2 & \cdots & r_{2p} \\ & & \ddots & \\ & & & h_p^2 \end{bmatrix}$$

where $h_i^2 = 1 - \Psi_i$

Suppose initial estimates are available for the communalities

$$(h_1^*)^2 \; (h_2^*)^2 \cdots\cdots (h_p^*)^2$$

Then

$$R_r = \begin{bmatrix} (h_1^*)^2 & r_{12} & \cdots & r_{1p} \\ & (h_2^*)^2 & & \vdots \\ & & \ddots & \\ & & & (h_p^*)^2 \end{bmatrix} \doteq L_r^* L_r^{*\prime}$$

where

$$L_r^* = \left[\sqrt{\hat{\lambda}_1^*}\, \hat{\underset{\sim}{e}}_1^* \;\vdots\; \cdots\cdots \;\vdots\; \sqrt{\hat{\lambda}_m^*}\, \hat{\underset{\sim}{e}}_m^* \right]$$

$$\Psi_i^* = 1 - \sum_{j=1}^{m} \hat{\lambda}_j^* \left[\hat{e}_{ij}^*\right]^2$$

and $\hat{\lambda}_1^* \geq \cdots \geq \hat{\lambda}_m^*$

are the largest eigenvalues for R_r

and $\hat{\underset{\sim}{e}}_1^*, \ldots, \hat{\underset{\sim}{e}}_m^*$ are the corresponding eigenvectors.

$$\left(\hat{h}_i^*\right)^2 = \sum_{j=1}^{m} \hat{\lambda}_j^* \left[\hat{e}_{ij}^*\right]^2 = 1 - \hat{\Psi}_i^*$$

Apply this procedure iteratively

1) Start with

$(h_i^*)^2 =$ R^2 value for the regression of X_i on the other variables

(2) Compute factor loadings from eigenvalues and eigenvectors of R_r

(3) Compute new $(h_i^*)^2$ values

[Repeat steps (2) and (3) until algorithm converges]

Problems:

- Some eigenvalues of R_r can be negative
- choice of m

If m is too large some communalities may become larger than one, and the iterations will terminate unless you use one of the following options:

HEYWOOD

Fix any communality that is larger than one equal to one and continue iterations with respect to the remaining variables

ULTRA HEYWOOD

Continue iterations with respect to all variables, regardless of the size of the communalities

	Factor 1	Factor 2	Specific Variance	h^2
Allied Chem.	.70	−.09	.50	.50
Dupont	.71	−.25	.44	.56
Union Carbide	.72	−.11	.47	.53
Exxon	.62	.23	.57	.43
Texaco	.62	.28	.54	.46
Variance explained	2.27	0.21		
	45%	4.2%		

$$R-\left(\hat{L}\hat{L}'+\hat{\Psi}\right) = \begin{bmatrix} 0 & .057 & -.005 & -.029 & .053 \\ & 0 & .060 & .007 & -.049 \\ & & 0 & .015 & .010 \\ & & & 0 & .073 \\ & & & & 0 \end{bmatrix}$$

Maximum Likelihood Method

A likelihood function is needed so additional assumptions are made

$$\underset{\sim}{X}_1, \underset{\sim}{X}_2, \ldots \underset{\sim}{X}_n \sim NID(\underset{\sim}{\mu}, \Sigma_{p\times p})$$

and

$$\underset{\sim}{X}_j = L\,\underset{\sim}{F}_j + \underset{\sim}{\epsilon}_j$$

where

$$\Sigma = LL' + \Psi$$

$$\underset{\sim}{F}_1, \underset{\sim}{F}_2, \ldots, \underset{\sim}{F}_n \sim NID(\underset{\sim}{0}, I_{m\times m})$$

$$\underset{\sim}{\epsilon}_1, \underset{\sim}{\epsilon}_2, \ldots, \underset{\sim}{\epsilon}_n \sim NID(\underset{\sim}{0}, \Psi_{p\times p})$$

the $\underset{\sim}{\epsilon}_i$'s and $\underset{\sim}{F}_i$'s are independent

A further restriction which specifies a unique solution

$$L' \Psi^{-1} L = \Delta$$

a diagonal matrix

[See supplement 9A for computational details of maximizing the likelihood function]

The m.l.e.'s are:

$$\hat{L}_{p \times m} = [\hat{\ell}_{ij}] \qquad \hat{\Psi} = \begin{bmatrix} \hat{\psi}_1 & & \\ & \ddots & \\ & & \hat{\psi}_p \end{bmatrix} \qquad \hat{\underset{\sim}{\mu}} = \bar{\underset{\sim}{X}}$$

For m factors:

estimated communalities

$$\hat{h}_i^2 = \sum_{k=1}^{m} \hat{\ell}_{ik}^2 \qquad \text{for } i = 1, 2, \ldots, p$$

$$\begin{bmatrix} \text{proportion of} \\ \text{total sample} \\ \text{variance due to} \\ \text{the k-th factor} \end{bmatrix} = \frac{\sum_{i=1}^{p} \hat{\ell}_{ik}^2}{\text{trace}(S_n)} \quad \text{using } S_n$$

$$= \frac{\sum_{i=1}^{p} \hat{\ell}_{ik}^2}{p} \quad \text{using } R$$

Maximum likelihood estimates for $m=2$ factor model for the example with 5 stocks.

Stock	Loadings for Factor 1	Loadings for Factor 2	Specific Variances
Allied Chem.	.684	.189	.50
Dupont	.694	.517	.25
Union Carbide	.681	.248	.47
Exxon	.621	−.073	.61
Texaco	.792	−.442	.18
Cummulative proportion of variance	.485	.598	

Residual Matrix

$$R - \hat{L}\hat{L}' - \hat{\Psi}$$

$$= \begin{bmatrix} 0 & .005 & -.004 & -.024 & -.004 \\ & 0 & -.003 & -.004 & .000 \\ & & 0 & .031 & -.004 \\ & & & 0 & -.000 \\ & & & & 0 \end{bmatrix}$$

Note:

The elements of the residual matrix are smaller for the m.l.e method.

A large sample chi-squared criterion for deciding if the number of factors m is sufficient.

$$H_0: \Sigma_{p\times p} = L_{p\times m} L'_{m\times p} + \Psi_{p\times p}$$

H_A: Σ is any positive definite covariance matrix

<u>Likelihood ratio test:</u>

Under H_0: $\hat{\Sigma} = \hat{L}\hat{L}' + \hat{\Psi}$ $\qquad \hat{\underset{\sim}{\mu}} = \bar{\underset{\sim}{X}}$

Under H_A: $\hat{\Sigma} = S_n = \frac{1}{n}A = \frac{n-1}{n}S$ $\qquad \hat{\underset{\sim}{\mu}} = \bar{\underset{\sim}{X}}$

$$-2\ln(\Lambda) = n\left[\ln|\hat{L}\hat{L}'+\hat{\Psi}| - \ln\left|\tfrac{n-1}{n}S\right|\right] + \underbrace{n\left[tr\left\{(\hat{L}\hat{L}'+\hat{\Psi})^{-1} S \tfrac{n-1}{n}\right\} - p\right]}_{\substack{\hat{S} \\ \text{this is zero}}}$$

$$= n\left[\ln|\hat{L}\hat{L}'+\hat{\Psi}| - \ln\left|\tfrac{n-1}{n}S\right|\right]$$

$$d.f. = \tfrac{1}{2}\left[(p-m)^2 - p - m\right]$$

<u>Bartlett correction:</u> Reject H_0 if

$$\left[n-1-\frac{2p+4m-5}{6}\right]\ln\left(\frac{|\hat{L}\hat{L}'+\hat{\Psi}|}{|\frac{n-1}{n}S|}\right) > \chi^2_{\frac{1}{2}[(p-m)^2-p-m],\alpha}$$

Must have
$m < \frac{1}{2}\left(2p+1-\sqrt{8p+1}\right)$ for positive d.f.

Determination of d.f.

Under $H_A \cup H_0$:

Estimate p means

$\frac{p(p+1)}{2}$ elements of Σ

Under H_0:

Estimate p means

$$\Sigma = LL' + \Psi$$

mp elements in L

p diagonal elements

these are subject to $\frac{m(m-1)}{2}$ constraints given by $L'\Psi L =$ diagonal matrix

$$df = \left[p + \frac{p(p+1)}{2}\right] - \left[p + mp + p - \frac{m(m-1)}{2}\right]$$

$$= \frac{1}{2}\left[(p-m)^2 - p - m\right]$$

Likelihood ratio test for

$$H_0: \Sigma_{p \times p} = L_{p \times m} L'_{p \times m} + \Psi_{p \times p}$$

diagonal matrix

tends to suggest too many factors (m values that are too large)

- non-normality
- outliers
- non-linearity
- correlated measurement errors

Consider AIC or SBC values

Kaiser Measure of Sampling Adequacy: (Psychometrika, 1970, 401-415)

(Kaiser + Rice, Ed. + Psych Meas. 1974, 34, 111-117)

$$MSA = 1 - \frac{\sum\sum_{j \neq k} q_{jk}^2}{\sum\sum_{j \neq k} r_{jk}^2}$$

r_{jk} = correlation between the j-th and k-th variables

q_{jk} = partial correlation between the j-th and k-th variables controlling for all other variables in the analysis

Guidelines:

$.9 < MSA$	excellent data
$.8 < MSA \leq .9$	very good
$.7 < MSA \leq .8$	good
$.6 < MSA \leq .7$	mediocre
$.5 < MSA \leq .6$	miserable
$MSA < .5$	unacceptable

Tucker and Lewis Reliability Coefficient

(Psychometrika, 1973, 38, 1-10)

From m.l.e. method, compute

$\hat{L}_m$ estimate of factor loadings

$$\hat{\Psi}_m = \text{diag}(R - \hat{L}_m \hat{L}_m')$$

$$G_m = \hat{\Psi}_m^{-1/2} (R - \hat{L}_m \hat{L}_m') \hat{\Psi}_m^{-1/2}$$

for the model with m factors

g_{mij} is called a "partial correlation" between variables X_i and X_j controlling for the m common factors

Sum of squared "partial correlations" between variables controlling for m common factors:

$$F_m = \sum_{i<j}\sum g_{mij}^2$$

Mean square:

$$M_m = F_m / df_m$$

where

$$df_m = \frac{1}{2}\left[(p-m)^2 - p - m\right]$$

are the d.f. for the log-likelihood ratio test of

H_0: m factors are sufficient

"Mean square" for model with zero common factors:

Then $G_0 = R$ and

$$M_0 = \left(\sum_{i<j}\sum r_{ij}^2\right) / df_0$$

where $df_0 = \frac{p(p-1)}{2}$

are d.f. for testing that all correlations are zero.

Reliability coefficient

$$\hat{\rho}_m = \frac{M_0 - M_m}{M_0 - \frac{1}{n_m}}$$

where $n_m = (n-1) - \frac{2p+5}{6} - \frac{2m}{3}$

is a Bartlett correction factor

Cornbach's Alpha

A set of observe items $\begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$ that "measure" the same latent trait should all have high positive correlations.

Define

$$\bar{r} = \frac{\frac{1}{p(p-1)/2} \sum\sum_{i<j} \widehat{Cov}(X_i, X_j)}{\frac{1}{p} \sum_{i=1}^{p} \widehat{Var}(X_i)}$$

$$= \frac{2}{p-1} \frac{\sum\sum_{i<j} s_{ij}}{\sum_i s_{ii}}$$

Cornbach's Alpha:

$$\alpha = \frac{p\,\bar{r}}{1 + (p-1)\,\bar{r}}$$

- $\alpha = 0$ if $\bar{r} = 0$
- $\alpha = 1$ if $\bar{r} = 1$
- when the analysis is done, be sure that all scales are orientated in the same direction, so all correlations are positive.

Rational:

Measure p items $\underset{\sim}{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix}$

Suppose each of the p items has a high positive correlation with a latent trait

Then each trait will have a high positive correlation with $\sum_{i=1}^{p} X_i$

$$Var\left(\sum_i X_i\right) = \sum_{i=1}^{p} Var(X_i) + 2 \sum_{i<j}\sum Cov(X_i, X_j)$$

$$= \sum_{i=1}^{p} Var(X_i) + 2 \sum_{i<j}\sum \rho_{ij} \sqrt{Var(X_i)\, Var(X_j)}$$

If $\rho_{ij} = 1$ for all $i \neq j$

then

$$Var(\Sigma X_i) = \sum_{i=1}^{p} Var(X_i) + 2 \sum_{i<j}\sum \sqrt{Var(X_i)\, Var(X_j)}$$

$$= \sum_{i}\sum_{j} \sqrt{Var(X_i)}\ \sqrt{Var(X_j)}$$

$$= \left[\sum_{i=1}^{p} \sqrt{Var(X_i)} \right]^2$$

$= p^2$ for standardized variables, $Var(X_i) = 1$ if $\rho_{ij} = 1$ for all $i \neq j$

This is the maximum value for $Var(\Sigma X_i)$

Also

$\sum_{i=1}^{p} Var(X_i) = p$ for standardized variables

In the extreme case where

$\rho_{ij} = 1$ for all $i \neq j$

$$\frac{\sum Var(X_i)}{Var(\sum X_i)} = \frac{p}{p^2} = \frac{1}{p}$$

and

$$\alpha = \frac{p}{p-1}\left[1 - \frac{\sum Var(x_i)}{Var(\sum_i x_i)}\right]$$

$$= \frac{p}{p-1}\left[1 - \frac{p}{p^2}\right] = 1$$

Note that:

$$\alpha = \frac{p}{p-1}\left[1 - \frac{\sum_i Var(x_i)}{Var(\sum_i X_i)}\right]$$

$$= \frac{p\bar{r}}{1+(p-1)\bar{r}} \quad \text{where } \bar{r} = \frac{\overline{Cov}}{\overline{Var}}$$

Factor Rotation:

Let $T_{m\times m}$ be an orthogonal matrix $(TT' = T'T = I_{m\times m})$.

Then, an orthogonal transformation of the factor loading matrix is

$$L_{p\times m} T_{m\times m} = \hat{L}^*_{p\times m}$$

and the decomposition of the covariance matrix is

$$\hat{L}\hat{L}' + \hat{\Psi} = \hat{L}TT'\hat{L}' + \hat{\Psi}$$

$$= \hat{L}^*(\hat{L}^*)' + \hat{\Psi}$$

Estimated specific variances and communalities are not altered by orthogonal transformations of $\hat{L}$.

The varimax criterion:

Define $\tilde{\ell}^*_{ij} = \hat{\ell}^*_{ij} / \hat{h}_i$

to be the "scaled" loading of the i-th variable on the j-th factor after rotation.

The varimax procedure selects the orthogonal transformation that maximizes

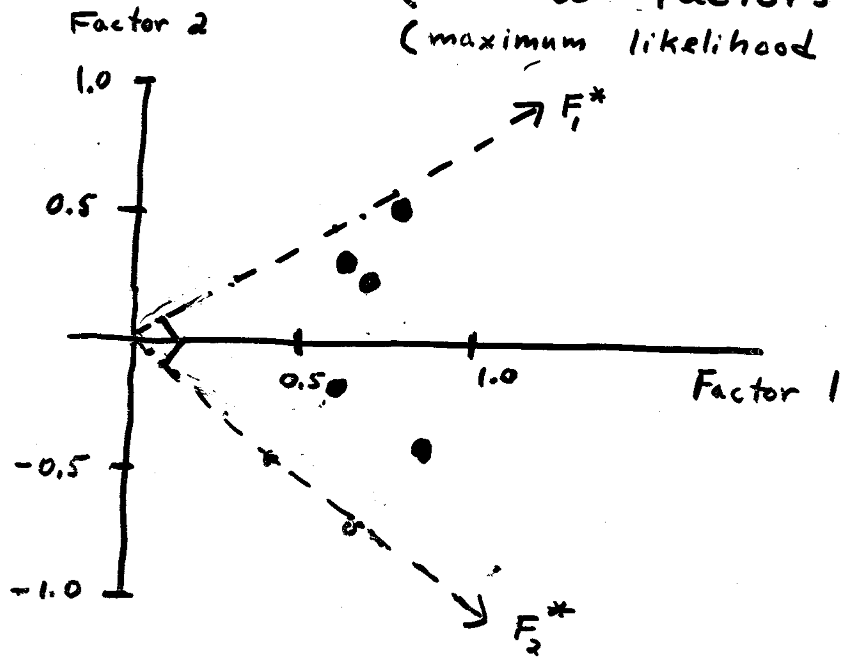
$$V = \frac{1}{p} \sum_{j=1}^{m} \left[\sum_{i=1}^{p} \left(\tilde{\ell}^*_{ij}\right)^4 - \frac{1}{p}\left[\sum_{i=1}^{p} \tilde{\ell}^{*2}_{ij}\right]^2 \right]$$

$$\propto \sum_{j=1}^{m} \left(\begin{array}{l} \text{variance of squares of} \\ \text{scaled loadings for j-th factor} \end{array} \right)$$

Scaling gives variables with smaller communalities more influence.

The stock market example

($m = 2$ factors)
(maximum likelihood)



After rotation each of the p variables (or measured traits) should have a high loading on only one factor. (This is not always possible).

Varimax rotation of $m=2$ factors for the stock market example: (maximum likelihood method)

Variable	m.l.e.'s F_1	F_2	Varimax rotation F_1^*	F_2^*
Allied Chem.	.68	.19	.60	.38
Du pont	.69	.52	.85	.16
Union Carbide	.68	.25	.64	.34
Exxon	.62	-.07	.36	.51
Texaco	.79	-.44	.21	.88

- the general market factor was destroyed by the rotation (it is possible to keep some factors fixed while rotating)
- relationship to pattern of correlations

Quartimax criterion:

Scaled loadings:

$$\tilde{\ell}_{ij}^{*} = \hat{\ell}_{ij}^{*} / \hat{h}_i$$

where $\hat{h}_i^2 = \sum_{j=1}^{m} \hat{\ell}_{ij}^2$

Select an **orthogonal** transformation (rotation) of the loadings to maximize

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

Objectives:

(1) Each of the p variables (or traits) should have a fairly high loading on the same factor.

(2) Each variable (or trait) should have a high loading on at most one other factor and near zero loadings on remaining factors.

Quartimax rotation of m=2 factors for the stock returns data (maximum likelihood method)

	m.l.e.'s		Quartimax rotation	
Variable	F_1	F_2	F_1^*	F_2^*
Allied Chem.	.68	.19	.71	.05
Dupont	.69	.52	.83	−.26
Union Carbide	.68	.25	.72	−.01
Exxon	.62	−.07	.56	.27
Texaco	.79	−.44	.60	.68

PROMAX METHOD:

An "oblique" or non-orthogonal transformation

(step 1) First do the varimax rotation to obtain the loadings $L^*_{p\times m}$

(step 2) Construct another $p\times m$ matrix Q where

$$q_{ij} = |\ell^*_{ij}|^{k-1}\,\ell^*_{ij}\,, \quad \ell^*_{ij}\neq 0$$
$$= 0 \qquad\qquad , \quad \ell^*_{ij} = 0$$

($k>1$ is chosen by trial and error, usually $k<4$)

(step 3) Find a matrix U such that each column of L^*U is close to the corresponding column of Q.

Choose the j-th column of U to minimize

$$(\underset{\sim}{q}_j - L^*\underset{\sim}{u}_j)'(\underset{\sim}{q}_j - L^*\underset{\sim}{u}_j)$$

This yields

$$U = (L^{*\prime}L^*)^{-1}L^{*\prime}Q$$

(step 4) Rescale U so transformed factors have unit variance

$$D^2 = \text{diag}\left[(U'U)^{-1}\right]$$

$$M = UD$$

PROMAX transformation yields factors with loadings

$$L^{P} = L^{*} M$$

Also,

$$\Phi = (M'M)^{-1}$$

is the correlation matrix for the new factors

$$L^{*}\underset{\sim}{F} = L^{*}M\underbrace{M^{-1}\underset{\sim}{F}} = L^{P}\underset{\sim}{F}^{P}$$

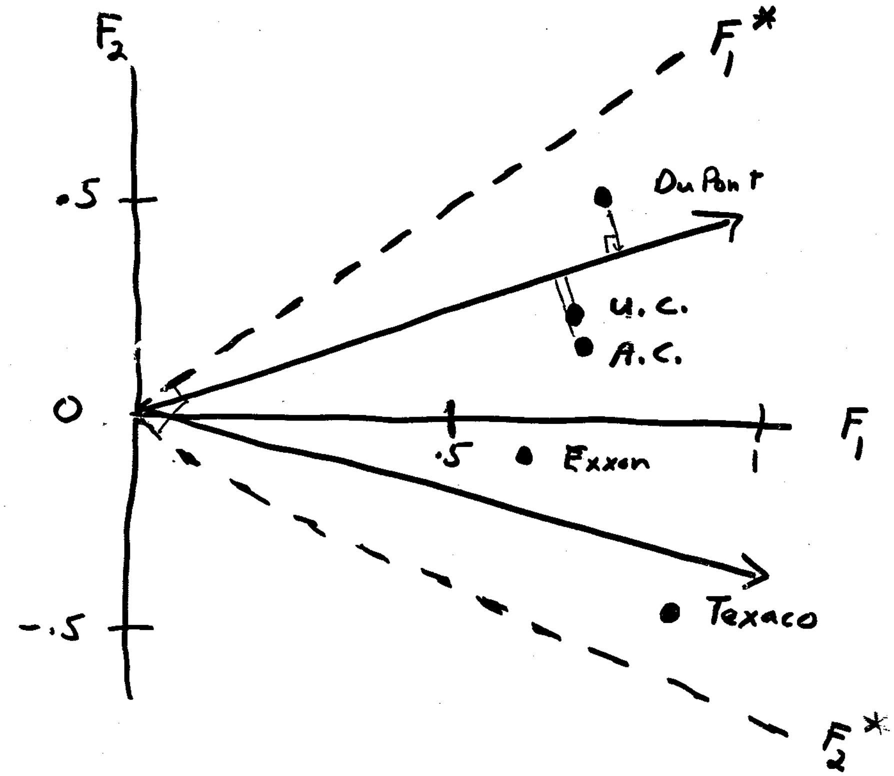
$\Rightarrow$

$$\begin{aligned} Var(\underset{\sim}{F}^{P}) &= Var(M^{-1}\underset{\sim}{F}) \\ &= M^{-1}\, Var(\underset{\sim}{F})\, (M^{-1})' \\ &= M^{-1}\, I\, (M^{-1})' \\ &= (M'M)^{-1} \end{aligned}$$

	Factor 1	Factor 2
Allied Chem	.56	.24
Dupont	.90	-.08
Union Carbide	.62	.18
Exxon	.26	.45
Texaco	-.02	.92

Correlation between factors is

0.49



The green axes correspond to an oblique rotation.

misuse of the word

Estimation of Factor Scores:

$$(\underset{\sim}{X}_j - \underset{\sim}{\mu})_{p\times 1} = L_{p\times m} \underset{\sim}{F}_j + \underset{\sim}{\epsilon}_j$$

If this model is correct

$$Var(\underset{\sim}{\epsilon}_j) = \Psi = \begin{bmatrix} \psi_1 & & \\ & \ddots & \\ & & \psi_p \end{bmatrix}$$

Weighted least squares estimation of $\underset{\sim}{F}_j$

$$\hat{\underset{\sim}{F}}_j = (L'\Psi^{-1}L)^{-1} L'\Psi^{-1}(\underset{\sim}{X}_j - \underset{\sim}{\mu})$$

$$\doteq (\hat{L}'\hat{\Psi}\hat{L})^{-1} \hat{L}'\hat{\Psi}^{-1}(\underset{\sim}{X}_j - \bar{\underset{\sim}{X}})$$

Ordinary (unweighted) least squares estimation is sometimes used when factor loadings are obtained from the principal component method (since specific variances tend to be more nearly equal, i.e. $\hat{\psi}_1 \doteq \cdots \doteq \hat{\psi}_p$)

$$\hat{\underset{\sim}{F}}_j = (\tilde{L}'\tilde{L})^{-1}\tilde{L}'(\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})$$

$$= \begin{bmatrix} \frac{1}{\sqrt{\hat{\lambda}_1}} \hat{\underset{\sim}{e}}_1'(\underset{\sim}{x}_j - \bar{\underset{\sim}{x}}) \\ \vdots \\ \frac{1}{\sqrt{\hat{\lambda}_m}} \hat{\underset{\sim}{e}}_m'(\underset{\sim}{x}_j - \bar{\underset{\sim}{x}}) \end{bmatrix}$$

$$\tilde{L} = \left[\sqrt{\hat{\lambda}_1}\, \hat{\underset{\sim}{e}}_1 \,\vdots\, \cdots \,\vdots\, \sqrt{\hat{\lambda}_m}\, \hat{\underset{\sim}{e}}_m \right]$$

Regression Method:

Consider the joint distribution of $(\underset{\sim}{X}_j - \underset{\sim}{\mu})$ and $\underset{\sim}{F}_j$

Assume multivariate normality as in the maximum likelihood approach to factor analysis.

$$\begin{bmatrix} \underset{\sim}{X}_j - \mu \\ \underset{\sim}{F}_j \end{bmatrix} \sim N_{p+m}\left(\underset{\sim}{0}, V\right)$$

where

$$V = \left[\begin{array}{c|c} LL' + \Psi & L \\ \hline L' & I_{m \times m} \end{array}\right]$$

if the m factor model is correct.

Obtain the conditional mean of $\underset{\sim}{f}_j$ given $\underset{\sim}{X}_j - \underset{\sim}{\mu} = (\underset{\sim}{x}_j - \underset{\sim}{\mu})$

$$E(\underset{\sim}{F}_j \mid \underset{\sim}{X}_j - \underset{\sim}{\mu})$$

$$= L'(LL' + \Psi)^{-1}(\underset{\sim}{x}_j - \underset{\sim}{\mu})$$

Use the estimated conditional mean vector as the estimate of the factor scores

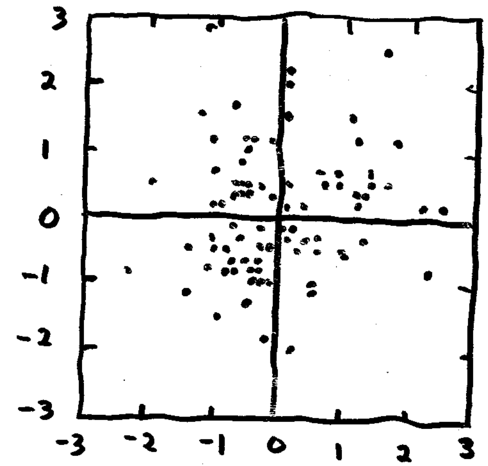
$$\hat{\underset{\sim}{F}}_j = \hat{L}'(\hat{L}\hat{L}' + \hat{\Psi})^{-1}(\underset{\sim}{x}_j - \bar{\underset{\sim}{X}})$$

To reduce bad effects of an incorrect choice of m replace $\hat{L}\hat{L}' + \hat{\Psi}$ with $S = \frac{1}{n-1}\sum_{j=1}^{n}(\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})'$ to obtain

$$\hat{\hat{\underset{\sim}{F}}}_j = \hat{L}' S^{-1}(\underset{\sim}{x}_j - \bar{\underset{\sim}{x}})$$

Estimated factor Scores for stock market example (maximum likelihood solution, $m = 2$, varimax rotation)

(Regression estimates)



Recall $\underset{\sim}{F}_j \sim NID_m(\underset{\sim}{0}, I_{m \times m})$

- check for outliers
- multivariate normality (circular contours)
- univariate tests of normality for the factor scores