Factor Analysis

An alternative technique for studying correlation and covariance structure

Let $\boldsymbol{X}$ be observable random vector which has a p-variate Normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$

The *Factor Analysis* Model:

Let $F_1, F_2, \ldots, F_m$ be some unobservable random variables called *the common factors*

Let $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$ be random variables called *errors* or specific factors.

Suppose that there exist constants λ_{ij} (*the loadings*) such that:

$$x_1 = \mu_1 + \lambda_{11}F_1 + \lambda_{12}F_2 + \ldots + \lambda_{1m} F_m + \varepsilon_1$$

$$x_2 = \mu_2 + \lambda_{21}F_1 + \lambda_{22}F_2 + \ldots + \lambda_{2m} F_m + \varepsilon_2$$

$$\ldots$$

$$x_p = \mu_p + \lambda_{p1}F_1 + \lambda_{p2}F_2 + \ldots + \lambda_{pm} F_m + \varepsilon_p$$

Factor Analysis Model in Matrix Notation

$$\boldsymbol{X} - \boldsymbol{\mu} = \boldsymbol{L}\boldsymbol{F} + \boldsymbol{\varepsilon}$$

where

$\boldsymbol{X}$ is $p \times 1$, $\boldsymbol{L}$ is $p \times m$, $\boldsymbol{F}$ is $m \times 1$, and $\boldsymbol{\varepsilon}$ is $p \times 1$

Assume: $\text{cov}(\boldsymbol{F}) = \boldsymbol{I}_{m \times m}$, and $\text{cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}$,
where

$$\Psi = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix}$$

Note:

$$\boldsymbol{\Sigma} = \text{cov}(\boldsymbol{X}) = \boldsymbol{LL}' + \boldsymbol{\Psi}$$

Hence

$$\sigma_{ii} = \text{Var}(X_i) = \sum_{j=1}^{m} \lambda_{ij}^2 + \psi_i$$

and

$$\sigma_{ik} = \text{cov}(X_i, X_k) = \sum_{j=1}^{m} \lambda_{ij}\lambda_{kj}$$

$h_i^2 = \sum_{j=1}^{m} \lambda_{ij}^2$ is called the i^{th} *communality*

i.e. the component of variance of x_i that is due to the common factors $F_1, F_2, \ldots, F_m$

ψ_i is called the *specific* variance

i.e. the component of variance of x_i that is **specific** only to that variable

$F_1, F_2, \ldots, F_m$ are called the **common factors**

$\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$ are called the **specific factors**

$$\lambda_{ij} = \mathrm{cov}\left(x_i, F_j\right)$$

$$= \text{the correlation between } x_i \text{ and } F_j.$$

Extracting the Factors

Several methods of estimation – we consider two:

1. Principal Component Method
2. Maximum Likelihood Method

Principle Component Method

Recall

$$\Sigma = \left[\vec{a}_1, \cdots, \vec{a}_p\right] \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_p \end{bmatrix} \begin{bmatrix} \vec{a}_1' \\ \vdots \\ \vec{a}_p' \end{bmatrix} = PDP'$$

where $\vec{a}_1, \cdots, \vec{a}_p$ are eigenvectors of $\boldsymbol{\Sigma}$ of length 1 and

$$\lambda_i \geq \ldots \geq \lambda_p \geq 0$$

are eigenvalues of $\boldsymbol{\Sigma}$.

Hence

$$\Sigma = \left[\sqrt{\lambda_1}\vec{a}_1, \cdots, \sqrt{\lambda_p}\vec{a}_p\right] \begin{bmatrix} \sqrt{\lambda_1}\vec{a}_1' \\ \vdots \\ \sqrt{\lambda_p}\vec{a}_p' \end{bmatrix} = LL' + \underset{p\times p}{0}$$

Thus

$$L = \left[\sqrt{\lambda_1}\vec{a}_1, \cdots, \sqrt{\lambda_p}\vec{a}_p\right] \text{ and } \Psi = \underset{p\times p}{0}$$

This is the ***Principal Component Solution*** with p factors

Note: The specific variances, ψ_i, are all zero.

The objective in Factor Analysis is to explain the correlation structure in the data vector with as few factors as necessary

It may happen that the latter eigenvalues of $\boldsymbol{\Sigma}$ are small.

$$\lambda_i \geq \ldots \geq \lambda_p \geq 0$$

$$\Sigma = \left[\vec{a}_1, \cdots, \vec{a}_p\right] \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_p \end{bmatrix} \begin{bmatrix} \vec{a}_1' \\ \vdots \\ \vec{a}_p' \end{bmatrix}$$

$$= \lambda_1 \vec{a}_1 \vec{a}_1' + \cdots + \lambda_p \vec{a}_p \vec{a}_p'$$

$$\approx \lambda_1 \boldsymbol{a}_1 \boldsymbol{a}_1' + \cdots + \lambda_m \boldsymbol{a}_m \boldsymbol{a}_m' = \boldsymbol{L}_m \boldsymbol{L}_m'$$

where $\boldsymbol{L}_m = [\sqrt{\lambda_1}\boldsymbol{a}_1, \ldots, \sqrt{\lambda_m}\boldsymbol{a}_m]$

In addition let

$$\psi_i = \sigma_{ii} - h_i^2 = i^{th} \text{ diagonal element of } \boldsymbol{\Sigma} - \boldsymbol{L}_m \boldsymbol{L}_m'$$

$$= \sigma_{ii} - \sum_{j=1}^{m} \lambda_{ij}^2$$

In this case

$$\boldsymbol{\Sigma} \approx \boldsymbol{L}_m \boldsymbol{L}_m{}' + \boldsymbol{\Psi}$$

The equality will be exact along the diagonal

where

$$\Psi = \begin{bmatrix} \psi_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \psi_p \end{bmatrix}$$

Maximum Likelihood Estimation

Let $\vec{x}_1, \cdots, \vec{x}_n$ denote a sample from $N_p\left(\vec{\mu}, \Sigma\right)$

where $\underset{p\times p}{\Sigma} = \underset{p\times k}{L}\,\underset{k\times p}{L'} + \underset{p\times p}{\Psi}$

The joint density of $\vec{x}_1, \cdots, \vec{x}_n$ is

$$L\left(\vec{\mu}, \Sigma\right) = L\left(\vec{\mu}, L, \Psi\right)$$

$$= \frac{1}{\left(2\pi\right)^{np/2}\left|\Sigma\right|^{n/2}} \exp\left\{-\tfrac{1}{2}\left[tr\left(\Sigma^{-1}A + n\Sigma^{-1}\left(\bar{\vec{x}} - \vec{\mu}\right)\left(\bar{\vec{x}} - \vec{\mu}\right)'\right)\right]\right\}$$

where $A = \left(n-1\right)S = \sum_{i=1}^{n}\left(\vec{x}_i - \bar{\vec{x}}\right)\left(\vec{x}_i - \bar{\vec{x}}\right)'$

The Likelihood function is

$$L(\vec{\mu},\Sigma) = L(\vec{\mu},L,\Psi)$$

$$= \frac{1}{(2\pi)^{(n-1)p/2}\left|\Sigma\right|^{(n-1)/2}}\exp\left\{-\tfrac{n-1}{2}\left[tr\left(\Sigma^{-1}S\right)\right]\right\}$$

$$\times\frac{1}{(2\pi)^{p/2}\left|\Sigma\right|^{1/2}}\exp\left\{-\tfrac{n}{2}\left[\left(\bar{\vec{x}}-\vec{\mu}\right)\Sigma^{-1}\left(\bar{\vec{x}}-\vec{\mu}\right)'\right]\right\}$$

with $\underset{p\times p}{\Sigma} = \underset{p\times k}{L}\ \underset{k\times p}{L'} + \underset{p\times p}{\Psi}$

The maximum likelihood estimates $\hat{\vec{\mu}},\hat{L}$ and $\hat{\Psi}$

Are obtained by numerical maximization of $L(\vec{\mu},L,\Psi)$

Example 9.6: *Olympic decathlon Scores*

Data was collected for n = 280 starts from 1960 to 2004 for the ten decathlon events (*100-m run, Long Jump, Shot Put, High Jump, 400-m run, 110-m hurdles, Discus, Pole Vault, Javelin, 1500-m run*). The sample correlation matrix is given on the next slide

Correlation Matrix

	X100.m	LongJump	ShotPut	HighJump	X400.m	X110.m.hurdles	Discus	PoleVault	Javelin	X1500.m
[1,]	1.0000	0.6386	0.4752	0.3227	0.5520	0.3262	0.3509	0.4008	0.1821	-0.0352
[2,]	0.6386	1.0000	0.4953	0.5668	0.4706	0.3520	0.3998	0.5167	0.3102	0.1012
[3,]	0.4752	0.4953	1.0000	0.4357	0.2539	0.2812	0.7926	0.4728	0.4682	-0.0120
[4,]	0.3227	0.5668	0.4357	1.0000	0.3449	0.3503	0.3657	0.6040	0.2344	0.2380
[5,]	0.5520	0.4706	0.2539	0.3449	1.0000	0.1546	0.2100	0.4213	0.2116	0.4125
[6,]	0.3262	0.3520	0.2812	0.3503	0.1546	1.0000	0.2553	0.4163	0.1712	0.0002
[7,]	0.3509	0.3998	0.7926	0.3657	0.2100	0.2553	1.0000	0.4036	0.4179	0.0109
[8,]	0.4008	0.5167	0.4728	0.6040	0.4213	0.4163	0.4036	1.0000	0.3151	0.2395
[9,]	0.1821	0.3102	0.4682	0.2344	0.2116	0.1712	0.4179	0.3151	1.0000	0.0983
[10,]	-0.0352	0.1012	-0.0120	0.2380	0.4125	0.0002	0.0109	0.2395	0.0983	1.0000

	PC1	PC2	PC3	PC4	h2	u2	com
X100.m	0.70	0.02	**-0.47**	-0.42	0.88	0.12	2.5
LongJump	0.79	0.08	-0.25	-0.11	0.71	0.29	1.3
ShotPut	0.77	**-0.43**	0.20	-0.11	0.83	0.17	1.8
HighJump	0.71	0.18	0.00	0.37	0.67	0.33	1.6
X400.m	0.60	**0.55**	-0.05	-0.40	0.83	0.17	2.7
X110.m.hurdles	0.51	-0.08	**-0.37**	0.56	0.72	0.28	2.8
Discus	0.69	**-0.46**	**0.29**	-0.08	0.77	0.23	2.2
PoleVault	0.76	0.16	0.02	0.30	0.70	0.30	1.4
Javelin	0.52	-0.25	**0.52**	-0.07	0.61	0.39	2.5
X1500.m	0.22	**0.75**	**0.49**	0.09	0.85	0.15	2.0

In this example, $p = 10$, $m = 4$

The columns PC1 to PC4 are the loadings λ_{ij}

h2 are the communalities

u2 are the psi's

Identification of the factors

Principal components

Factor	*Description*
1	General athletic ability
2	Contrast of running ability with throwing ability
3	Contrast of endurance with speed
4	Mystery

Maximum Likelihood

Factor	*Description*
1	Running Endurance (1500m)
2	Strength
3	Running endurance (400m & 1500m)
4	Leg strength (jumping)