Factor Analysis

The essential purpose of factor analysis is to describe, if possible, the covariance relationships among many variables in terms of a few underlying, but unobservable, random quantities called *factors*. Basically, the factor model is motivated by the following argument: Suppose variables can be grouped by their correlations. That is, suppose all variables within a particular group are highly correlated among themselves, but have relatively small correlations with variables in a different group. Then it is conceivable that each group of variables represents a single underlying construct, or factor, that is responsible for the observed correlations. For example, correlations from the group of test scores in classics, French, English, mathematics, and music collected by Spearman suggested an underlying "intelligence" factor. A second group of variables, representing physical-fitness scores, if available, might correspond to another factor. It is this type of structure that factor analysis seeks to confirm.

Factor analysis can be considered an extension of principal component analysis. Both can be viewed as attempts to approximate the covariance matrix $\mathbf{\Sigma}$. However, the approximation based on the factor analysis model is more elaborate. The primary question in factor analysis is whether the data are consistent with a prescribed structure.

- Measurement is necessary especially when it is unobservable
- Any behaviour pattern should be measured numerically
- Charles Spearman – Measuring of human intelligence using Factor analysis
- IQ level of Persons? - human intelligence is not visible- unobservable- how do you measure such unobservable?

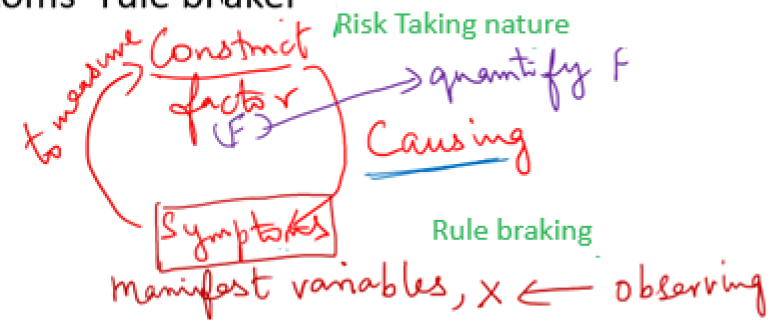
Few examples – cannot be measured easily

- A psychologist is interested to measure mental ability of a person
- A safety measure manager is interested to measure safety environment of his workplace
- A supply chain analyst is interested to measure supply chain coordination
- A marketing manager is interested to measure the purchase intension of customers
- A welfare officer is interested to measure the ability of worklife in factories

There are certain things which are hidden unobservables and are not directly measurable but those things are very important. Since they causes something to happen. Because of the presence of unobservable things, these manifested to different symptoms.

- Hence the hidden concepts/ constructs – not directly measured-
- concepts/ constructs - manifested into different symptoms
- Example : Construct / factor– Risk taking in nature causing

  Symptoms- rule braker



Observing numbers

$$X_{p\times 1} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ X_p \end{bmatrix},$$

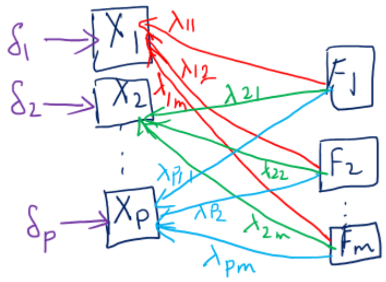
P manifest variables from population

mean vector of X, $\mu_{p\times 1} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$

$$Cov(X) = \Sigma = \begin{array}{c|ccc} & x_1 & x_2 & \cdots x_p \\ \hline x_1 & \sigma_{11} & \sigma_{12} & \cdots \sigma_{1p} \\ x_2 & \sigma_{12} & \sigma_{22} & \cdots \sigma_{2p} \\ x_3 & \vdots & & \\ x_p & \sigma_{1p} & \sigma_{2p} & \cdots \sigma_{pp} \end{array} \quad p\times p$$

Consider m factors

$$F_{m\times 1} = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix} \qquad m < p$$



m factors influencing

X, is measured through λ

whatever you observed in X, because of the causal factor F.

- The F_m factors collectively cannot explain everything about X, So add error term

No relationship, between Fs. They are independent Orthogonal

$$X_1 = \mu_1 + \lambda_{11} F_1 + \lambda_{12} F_2 + \dots \lambda_{1m} F_m + \delta_1$$

$$X_2 = \mu_2 + \lambda_{21} F_1 + \lambda_{22} F_2 + \dots \lambda_{2m} F_m + \delta_2$$

$$\vdots$$

$$X_j = \mu_j + \lambda_{j1} F_1 + \lambda_{j2} F_2 + \dots \lambda_{jm} F_m + \delta_j$$

$$\vdots$$

$$X_p = \mu_p + \lambda_{p1} F_1 + \lambda_{p2} F_2 + \dots \lambda_{pm} F_m + \delta_p$$

$$\underbrace{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_j \\ \vdots \\ X_p \end{bmatrix}_{p\times 1}}_{X} = \underbrace{\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_j \\ \vdots \\ \mu_p \end{bmatrix}_{p\times 1}}_{\mu} + \underbrace{\begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2m} \\ \vdots & & & \\ \lambda_{j1} & \lambda_{j2} & \cdots & \lambda_{jm} \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pm} \end{bmatrix}_{p\times m}}_{\lambda} \underbrace{\begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_k \\ \vdots \\ F_m \end{bmatrix}_{m\times 1}}_{F} + \underbrace{\begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_j \\ \vdots \\ \delta_p \end{bmatrix}_{p\times 1}}_{\delta}$$

$$\underset{(p\times 1)}{X} = \underset{(p\times 1)}{\mu} + \underset{(p\times m)}{\lambda} \underset{(m\times 1)}{F} + \delta_{(p\times 1)}$$

$$X - \mu = \lambda F + \delta \quad — (1) \text{ Factor model}$$

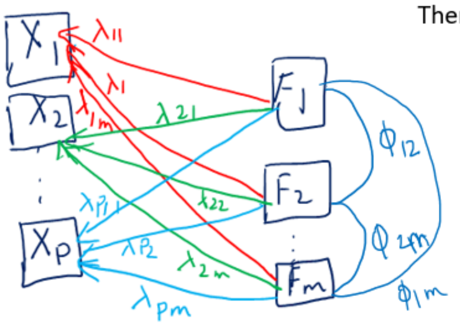
Assumptions:

$E(x) = \mu \qquad E(F) = 0 \qquad E(\delta) = 0$

$Cov(x) = \Sigma \qquad Cov(F) = E(FF^T) = I$

$Cov(F\delta) = 0 \qquad Cov(\delta) = \psi = \begin{bmatrix} \psi_{11} & 0 & \cdots & 0 \\ 0 & \psi_{22} & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \ddots & \psi_{pp} \end{bmatrix}$

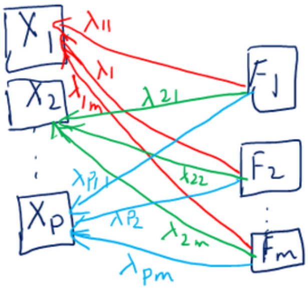
If the factor model hold all these assumptions, then it is called Exploratory orthogonal Factor model



Oblique factor model- When you allow the covariance between factor Then, it is called Oblique FM

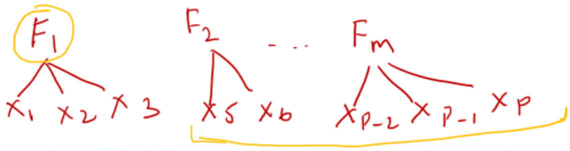
Factors are orthogonal, since they are independent

- Types of Factor Model: **1.Exploratory Factor Model** $X - \mu = \lambda F + \delta$



- **Each X is dependent on all F.**
- **Which are the X variables are basically coming out of F ?**
- **So allow all the full sets, find the factors that are linked to what type of variables**

2. **Confirmatory Factor Model**



Which of the manifest var representing which factor. X1,X2,X3 related F1 and so on -> need to confirm whether the connections are correct or wrong. : Hypothesis.
Through confirmatory FM we are going to prove hypothesis.

Key question 1

- What is common in the examples given?
  - The variables to be measured are unobservable or hidden or latent and known as constructs or factors : F
  - They all manifest some sympt oms which can be observed and measured and known as manifest variables : X

Factor analysis quantifies these constructs (factors with the help of the manifest variables)

We are doing this both in EFA, CFA

Key question 2

- The lessor the information contents, the better the design
  - When I have P number of variables, I have p dimensional information and I don’t want this much information. I want only 2 D only.
- FA also reduces dimensions

Purpose:

- The purpose of FA is to describe, if possible, the covariance relationships among many variables in terms of A FEW underlying, but unobservable, random quantities called FACTORS

Variability explanation:

- $Cov(X)=\Sigma$

$$X-\mu = \lambda F + \delta \quad \text{— (1) Factor model}$$

- Know: what is the variability you want to explain using FM?
- Whose variability? – X, what is variability? Cov(X)

$$
\begin{aligned}
Cov(X) &= E[(X-\mu)(X-\mu)^T] \\
&= E\left[(\lambda F + \delta)(\lambda F + \delta)^T\right] \\
&= E\left[\lambda F F^T \lambda^T + \lambda F \delta^T + \delta F^T \lambda^T + \delta \delta^T\right]
\end{aligned}
$$

λ – constant, $= \lambda E(FF^T)\lambda^T + \lambda E(F\delta^T) + E(\delta F^T)\lambda^T + E(\delta\delta^T)$

From assumptions, $= \lambda I \lambda^T + \lambda(0) + (0)\lambda^T + \psi$

$$\Sigma = \lambda\lambda^T + \psi \quad \text{— (2)}$$

if you have λ, you know ψ

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

λ is loading matrix

$$x_j = \lambda_{j1}F_1 + \lambda_{j2}F_2 + \cdots + \lambda_{jk}F_k + \cdots \lambda_{jm}F_m + \delta_j$$

$j = 1, 2 \ldots p$

$k = 1, 2 \ldots m$

general loading $= \lambda_{jk}F_k$

loading of k^{th} factor j^{th} x.

$$\lambda\lambda^T = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2m} \\ \vdots & & & \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pm} \end{bmatrix}_{p\times m} \begin{bmatrix} \lambda_{11} & \lambda_{21} & \cdots & \lambda_{p1} \\ \lambda_{21} & \lambda_{22} & & \lambda_{p2} \\ \vdots & \vdots & & \vdots \\ \cancel{\lambda_{p1}}\ \lambda_{1m} & \cancel{\lambda_{p2}}\ \lambda_{2m} & & \lambda_{pm} \end{bmatrix}_{m\times p}$$

$$= \begin{bmatrix} \sum_{k=1}^{m} \lambda_{1k}^2 & \sum_{k=1}^{m} \lambda_{1k}\lambda_{2k} & \cdots & \sum_{k=1}^{m} \lambda_{1k}\lambda_{pk} \\ \sum_{k=1}^{m} \lambda_{1k}\lambda_{2k} & \sum_{k=1}^{m} \lambda_{2k}^2 & \cdots & \sum_{k=1}^{m} \lambda_{2k}\lambda_{pk} \\ \vdots & & & \vdots \\ \sum_{k=1}^{m} \lambda_{1k}\lambda_{pk} & \sum_{k=1}^{m} \lambda_{2k}\lambda_{pk} & \cdots & \sum_{k=1}^{m} \lambda_{pk}^2 \end{bmatrix}$$

Variance?
Covariance?

$$\Sigma = \lambda\lambda^T + \psi = \begin{bmatrix} \sum_{k=1}^{m} \lambda_{1k}^2 & \sum_{k=1}^{m} \lambda_{1k}\lambda_{2k} & \cdots & \sum_{k=1}^{m} \lambda_{1k}\lambda_{pk} \\ \sum_{k=1}^{m} \lambda_{1k}\lambda_{2k} & \sum_{k=1}^{m} \lambda_{2k}^2 & \cdots & \sum_{k=1}^{m} \lambda_{2k}\lambda_{pk} \\ \vdots & & & \vdots \\ \sum_{k=1}^{m} \lambda_{1k}\lambda_{pk} & \sum_{k=1}^{m} \lambda_{2k}\lambda_{pk} & \cdots & \sum_{k=1}^{m} \lambda_{pk}^2 \end{bmatrix} + \begin{bmatrix} \psi_{11} & 0 & \cdots & 0 \\ 0 & \psi_{22} & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & \psi_{pp} \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}_{p\times p}$$

$$\sigma_{11} = \sum_{k=1}^{m} \lambda_{1k}^2 + \psi_{11}$$

$$\sigma_{22} = \sum_{k=1}^{m} \lambda_{2k}^2 + \psi_{22}$$

$$\sigma_{jj} = \sum_{k=1}^{m} \lambda_{jk}^2 + \psi_{jj}$$

$$\sigma_{jj} = \sum_{k=1}^{m} \lambda_{jk}^2 + \psi_{jj}$$

understand the variability part,

total variability

$$X_j : \boxed{\underbrace{\sum_{k=1}^{m} \lambda_{jk}^2}_{\text{Contributed by the } m \text{ common factors}} \ \Big|\ \underbrace{\psi_{jj}}_{\text{not contributed by the common factors}}}$$

The common factors are not able to explain this much of variability, this is unexplained

Contributed by the m Common factors

Commonality (h_j)

not contributed by the common factors

unique, specific variance of X_j

What is the contribution of first factor? λ_{11}^2

$$\mathrm{Cov}(XF) = E\left[(x-\mu)(F-E(F))^T\right]$$
$$= E\left[(x-\mu)F^T\right]$$

We know that
$x - \mu = \lambda F - \delta$

$$(x-\mu)F^T = (\lambda F - \delta)F^T$$
$$= \lambda F F^T - \delta F^T$$

$$\therefore E(x-\mu)F^T = E(\lambda F - \delta)F^T$$
$$= E(\lambda F F^T) - E(\delta F^T)$$
$$= \lambda E(F F^T) + 0$$
$$= \lambda$$

loading matrix.

An example

(Ref: Lawley and Maxwell, 1971; taken from Johnson and Wichern, Appl Mul Stat Ana, 2002)

- Lawley and Maxwell (1971) studied the general intelligence of 220 students

Student no.	Gaelic (X1)	English (X2)	History (X3)	Arithmatic (X4)	Algebra (X5)	Geometry (X6)
1						
2						
...						
220						

Factor ability criteria- can go for FA or not?

An example (contd.)

- The correlation matrix

	X1	X2	X3	X4	X5	X6
X1	1.00	0.44	0.41	0.29	0.33	0.25
X2		1.00	0.35	0.35	0.32	0.33
X3			1.00	0.16	0.19	0.18
X4				1.00	0.60	0.47
X5					1.00	0.46
X6						1.00

Exploratory Factor Analysis

- Model Estimation
- Model Adequacy Testing

Data : $X_{n \times p}$ on p variable(s)

$\Downarrow$

$S_{p \times p}$

$$X - \mu = \lambda F + \delta$$

$$\underset{(p\times p)}{\Sigma} = \underset{(p\times m)}{\lambda} \underset{(m\times p)}{\lambda^T} + \underset{(p\times p)}{\psi} \quad (\text{diagonal})$$

Estimate $\hat{\lambda}$, $\hat{\psi}$

What is known to us? We know $\hat{\Sigma} = S_{p \times p}$

If think that n is large sample space.
Appropriate Sampling strategy is applied

$$\hat{\Sigma} = \begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1p} \\ \hat{\sigma}_{12} & \hat{\sigma}_{22} & \cdots & \hat{\sigma}_{2p} \\ \vdots & & & \\ \hat{\sigma}_{1p} & \hat{\sigma}_{2p} & \cdots & \hat{\sigma}_{pp} \end{bmatrix} = S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{12} & s_{22} & \cdots & s_{2p} \\ \vdots & & & \\ s_{1p} & s_{2p} & \cdots & s_{pp} \end{bmatrix}$$

In general,

$$\hat{\sigma}_{jk} = s_{jk}, \quad j = 1, 2 \cdots p, \; k = 1, 2 \cdots p$$

- Now, we can assume that under root estimation that what will happen ultimately, if our sampling size is appropriate at recreating will have to capture properly, the behaviour the covalent structure.

$$\hat{\sigma}_{jk} = s_{jk}, \quad j = 1, 2 \cdots p, \; k = 1, 2 \cdots p$$

- each of the element in ∑-very close.

$$\hat{\Sigma} = \hat{\lambda}\hat{\lambda}^T + \hat{\psi}$$

We know S only. So Initialize $\hat{\lambda}, \hat{\psi}$ then find out $\hat{\Sigma}$. Then Compare this value with the individual elements of S. And we find out a situation where the element of these matrices (s,$\hat{\Sigma}$) will be as small as possible with certain threshold limit.

We want to estimate : $\hat{\Sigma} = \hat{\lambda}\hat{\lambda}^T + \hat{\psi}$

S ?   ?

- so we initialize values for this
- And then we find out this value, then we compare this value with s individual elements of the vector.
- And if we find out a situation where the difference between the element of these two matrices will be as small as possible within certain threshold limit.
- Trace (∑)= $\sum_{j=1}^{p} \sigma_{jj}$; Trace (S)= $\sum_{j=1}^{p} s_{jj}$
- Error – difference between S (sample cov) - ∑ (population cov)
- sum square error should be Minimized

$$\therefore \quad (S-\Sigma)^2$$

minimize $G = tr\left[(S-\Sigma)^T(S-\Sigma)\right]$

This leads to 2nd norm

$$= \|(S-\Sigma)^2\|$$

$$= \|(S-\lambda\lambda^T+\psi)\|$$

We want

① Initial value of ψ

② The number of factors — m value