19 February 2018 23:34

Factor Analysis

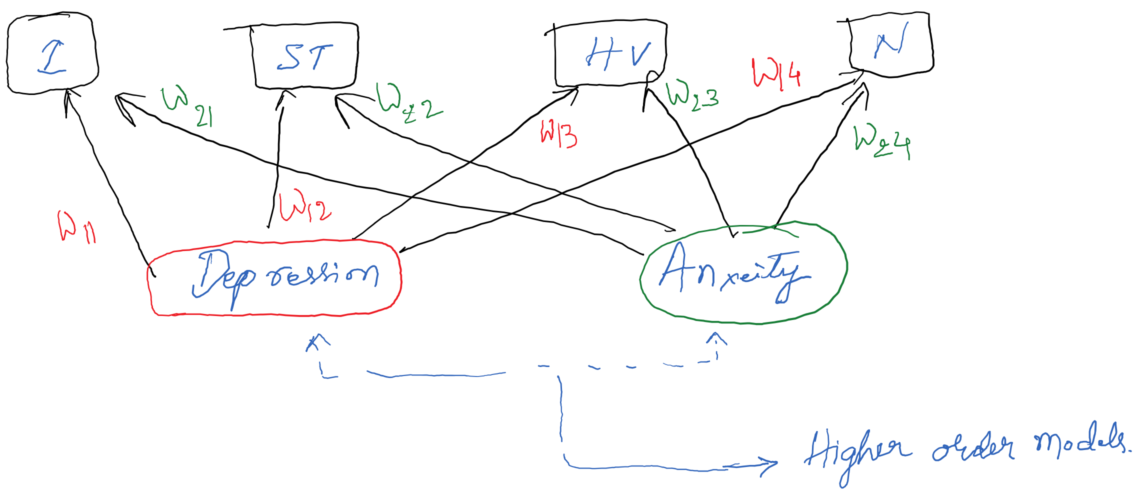
The purpose is to estimate a model which explains Variance / Co-var between a set of observed variables (in a population) by a set of (fewer) unobserved factors + weights

Data set of observed "psychiatric" characters of individuals

Insomnia   Suicidal Thoughts   Hyperventilation   Nausea

$Cov(I, S) = 0.3$, $cov(S, N) = 0.7$ etc we are trying to come up with a model that explains the co-var in that population

If our observed variance & co-variance b/n variables could be because of some unobserved factors.



The factors D & A may be responsible for variance & co-variance b/n the above variables

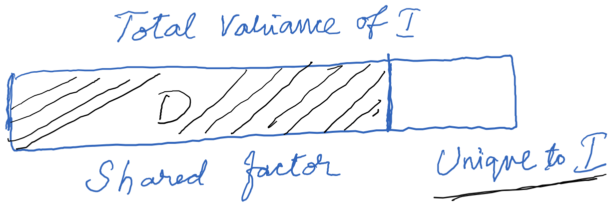
There is a weightage of D on I, ST, HV, N &

———— " ———— A on I, ST, HV, N

The weightage these unobserved characters have on these observed characters are different ($w_{11}, w_{12}, w_{13}, w_{14}, w_{21}, w_{22}, w_{23}, w_{24}$)

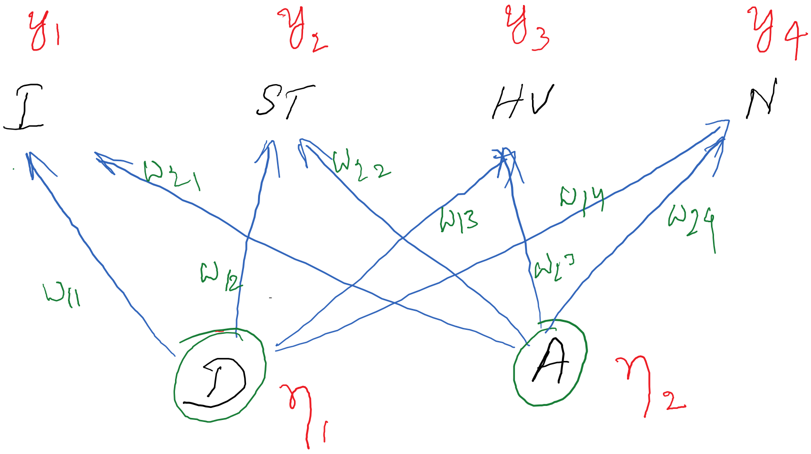
We are trying to estimate these weights & unobserved factors

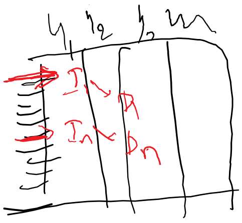
We are trying to estimate these weights & Unobserved factors



Major use of Factor Analysis is → Dimensionality reduction

Model representation in FA





$$y_1 = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \end{bmatrix} \cdots \quad y_4 = \begin{bmatrix} y_{41} \\ y_{42} \\ \vdots \\ y_{4n} \end{bmatrix}$$

2 ways of model selection

1 → Vector of Regression eqns

2 → Matrix Representation

1 → Vector of equations → Not to Consider individual values, we Consider Vectors alone

$$\Rightarrow y_1 = \lambda_{11} \cdot \eta_1 + \lambda_{21} \cdot \eta_2 + \varepsilon_1$$

Weights Weight Error term (Unique Variation of y_1)

$$y_4 = \underbrace{\lambda_{41} \cdot \eta_1 + \lambda_{42} \cdot \eta_2}_{} + \underbrace{\varepsilon_4}_{}$$

Unique Variance

$$y_4 = \underbrace{\lambda_{41} \cdot \eta_1 + \lambda_{42} \cdot \eta_2}_{\text{Commonality variance}} + \underbrace{\varepsilon_4}_{\text{Unique Variance}}$$

η are different

$$\Rightarrow \eta_1 = \begin{pmatrix} \eta_{11} \\ \eta_{12} \\ \vdots \\ \eta_{1n} \end{pmatrix} \quad \& \quad \eta_2 = \begin{pmatrix} \eta_{21} \\ \eta_{22} \\ \vdots \\ \eta_{2n} \end{pmatrix}$$

Combine

y_1, y_2, y_3, y_4

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{12} \\ \vdots & \ddots & \\ \lambda_{41} & & \lambda_{42} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

$\to \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$

$y = \Lambda \eta + \varepsilon$ -

$y_1 \Rightarrow \lambda_{11} \cdot \eta_1 + \lambda_{12} \cdot \eta_2 \ldots + \varepsilon_1$

2 -

3 -

4 -

$y_1 = y_{11} = \lambda_{11}\eta_1 + \varepsilon_1$

$y_{12} = \lambda_{12}\eta_2 + \varepsilon_2$

$y_{13} =$

$\vdots$

$y_{1n} =$

$$\boxed{y = \Lambda \eta + \varepsilon}$$

$\varepsilon \to$ Vector of Unique variance.

$\Lambda \to$ Weights

$\eta \to$ Variance of Unobserved factor

The above method hides information $\to$ Vector inside a vector.

---

(2) Matrix representation $\to$ Allows us to keep more information

Adding index $\to$ i

$$\begin{bmatrix} y_{i1} = \lambda_{11}\eta_{i1} + \lambda_{12}\eta_{i2} + \varepsilon_{i1} \\ \vdots \\ y_{i4} = \lambda_{41}\eta_{i1} + \lambda_{42}\eta_{i2} + \varepsilon_{i4} \end{bmatrix} \quad i = 1 \ldots n$$

$i \to 1 - n$
$j \to 1 - 4$

$$\Rightarrow \quad y_{ij} = \lambda_{j1} \cdot \eta_{i1} + \lambda_{j2}\eta_{i2} + \varepsilon_{ij}$$

$$\begin{bmatrix} y_{11} & \cdots & y_{14} \\ \vdots & \ddots & \vdots \\ y_{41} & \cdots & y_{44} \end{bmatrix} = \begin{bmatrix} \eta_{11} & \eta_{12} \\ \vdots & \vdots \\ \eta_{N1} & \eta_{N2} \end{bmatrix} \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{41} \\ \lambda_{12} & \cdots & \lambda_{42} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} & \cdots & \varepsilon_{14} \\ \vdots & \ddots & \vdots \\ \varepsilon_{N1} & \cdots & \varepsilon_{N4} \end{bmatrix}$$

$$\Rightarrow \boxed{Y_{NV} = F_{Nf} \, P'_{fV} + \varepsilon_{NV}} \rightarrow \text{Unique var or error Term}$$

Rows, Variance, factors, Transpose of weights.

$N = \#$ of individuals, $f = \#$ of factors, $V = \#$ of observed characters.

Procedure
→ Load data
→ Correlation Matrix
→ Find eigen Values.
→ Uniqueness & Commonality
$U + C = 1$

If all factors jointly explain large variance in a given Variable they are highly Common & less Unique

→ Visualise the data

(*) → Rotation of factors → we use Varimax Rotation
→ Orthogonally rotates the factor axis with the goal of Maximizing the Variance of Squared loadings of a factor on the Variable in the factor matrix
Weights.

the Variance of squared loadings of a factor matrix

weights.

→ Create Composite Variables → Spurious factors are introduced or Try to Combine Synonyms in to composites.