

FACTOR ANALYSIS

STATISTICS 407, ISU

1

EXPLANATION

In many practical situations it is not possible to observe the concepts or processes that are major interest. An infamous example is the measurement of intelligence, but more practical examples are in environmental problems such as measuring forest health, or in sociology measuring social factors. In these cases the researcher will collect information on variables likely to be indicators of the concepts in question and try to find the relationship between these observed variables to determine if they are consistent with them being measures of a small number of unobservable _____ variables. These unobservables are called _____.

2

THE ORIGINS

Spearman(1904) published a paper which is believed to be the FA original. He was working with exam scores (Classics, French, English, Math, Discourse, Music) of children and noticed a systematic effect in the _____, of the scores children made on different tests.

$$R = \begin{bmatrix} 1 & 0.83 & 0.78 & 0.70 & 0.66 & 0.63 \\ 0.83 & 1 & 0.67 & 0.67 & 0.65 & 0.57 \\ 0.78 & 0.67 & 1 & 0.64 & 0.54 & 0.51 \\ 0.70 & 0.67 & 0.64 & 1 & 0.45 & 0.51 \\ 0.66 & 0.65 & 0.54 & 0.45 & 1 & 0.40 \\ 0.63 & 0.57 & 0.51 & 0.51 & 0.40 & 1 \end{bmatrix}$$

3

ORIGINS

He proposed that a randomly selected child's score could be modeled by:

$$\text{Classics} = \lambda_1 f + \varepsilon_1$$

$$\text{French} = \lambda_2 f + \varepsilon_2$$

$$\text{English} = \lambda_3 f + \varepsilon_3$$

$$\text{Math} = \lambda_4 f + \varepsilon_4$$

$$\text{Discs} = \lambda_5 f + \varepsilon_5$$

$$\text{Music} = \lambda_6 f + \varepsilon_6$$



4

THE COMMON FACTOR MODEL

A set of observed variables, X_1, \dots, X_p , is explained by a smaller number of unobservable, latent variables, f_1, f_2, \dots, f_k . The model is as follows:

$$\begin{aligned} X_{i1} &= \mu_1 + \lambda_{11}f_{i1} + \dots + \lambda_{1k}f_{ik} + \varepsilon_{i1} \\ X_{i2} &= \mu_2 + \lambda_{21}f_{i1} + \dots + \lambda_{2k}f_{ik} + \varepsilon_{i2} \\ &\vdots \\ X_{ip} &= \mu_p + \lambda_{p1}f_{i1} + \dots + \lambda_{pk}f_{ik} + \varepsilon_{ip} \end{aligned}$$

$$i=1, \dots, n; k < p$$

5

VARIANCES

The variance for each variable can be written as:

$$\text{Var}(\mathbf{X}_j - \boldsymbol{\mu}_j) = \sigma_j^2 = \sum_{i=1}^k \lambda_{ji}^2 + \psi_j$$

The covariance between variables is:

$$\text{Cov}(\mathbf{X}_i, \mathbf{X}_j) = \sum_{h=1}^k \lambda_{ih} \lambda_{jh}$$

6

ESTIMATION

The loadings and specific variances can be estimated either by

- _____ or
- _____.

7

ESTIMATION BY PRINCIPAL COMPONENTS

- _____:
 $\hat{\lambda}_j = \text{eigenvalue}_j \times \text{eigenvector}_j, j = 1, \dots, k.$

- _____:
$$\psi_j = s_{jj} - \sum_{i=1}^k \hat{\lambda}_{ji}^2, j = 1, \dots, p$$

8

ESTIMATION BY MAXIMUM LIKELIHOOD

Find the estimates for

$$\mu_j, \lambda_j, \psi_j, j = 1, \dots, p$$

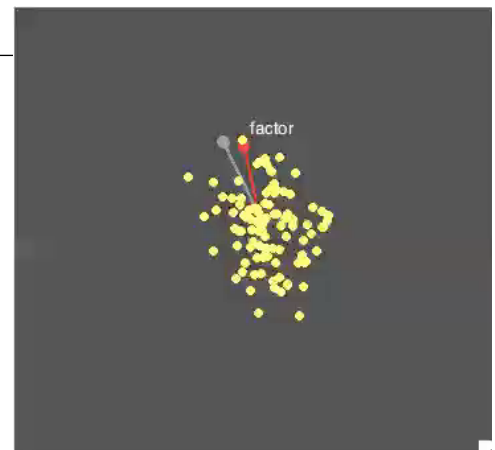
which maximize the likelihood equation.

Requires _____ assumptions in order to write out the likelihood equation.

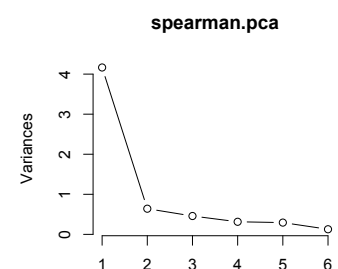
9

SPEARMAN'S EXAMPLE

- “True” factor (_____)
- Estimated factor by PCA (_____) and by MLE (_____).



Variable	PCA		MLE	
	Loadings	Specific	Loadings	Specific
Classics	0.88	0.23	0.85	0.27
French	0.76	0.42	0.69	0.52
English	0.88	0.22	0.85	0.27
Math	0.94	0.12	0.96	0.07
Discourse	0.82	0.33	0.77	0.40
Music	0.70	0.52	0.62	0.61
Variance	4.17	1.83	3.85	2.15
% Cumulative	70	100	64	100



10

FACTOR ROTATION ($K > 1$)

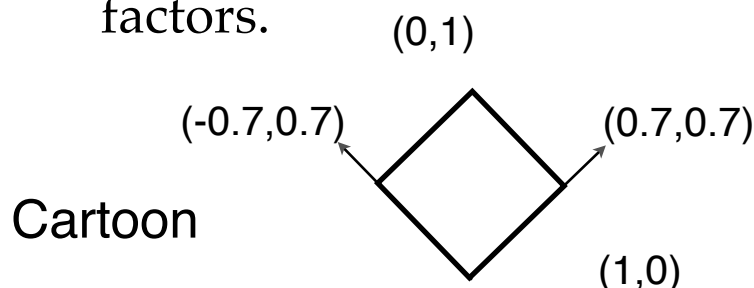
The loadings matrix is an orthonormal basis defining the factor plane. The choice of this basis is not unique, and often the interpretation of the coefficients for the basis is not as intuitive or immediate as desired. So an important part of factor analysis is to rotate the loadings matrix to get more _____ coefficients.

$$\hat{\Lambda}^* = \hat{\Lambda} \mathbf{T}, \quad \text{where } \mathbf{T}'\mathbf{T} = \mathbf{T}\mathbf{T}' = I_k$$

11

ROTATION GOALS

- Try to make as many of the loadings near _____ and make the others as large as possible.
- Each variable has a high loading on one factor, and low or zero loadings on other factors.



12

ROTATION METHODS

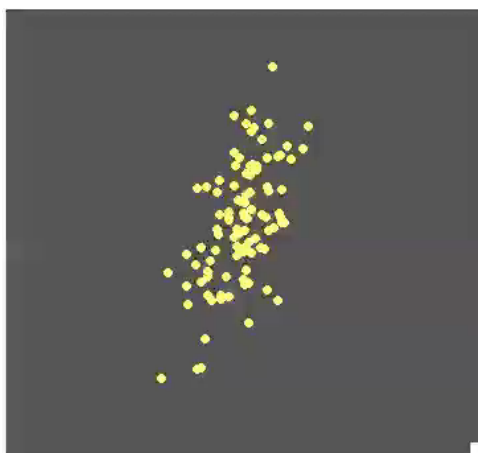
Orthogonal

- _____: Maximize the sum of the variances of the squared loadings within a factor. Has the effect of spreading out the loadings, forcing some towards 0 and others towards 1.
- *Quartimax*: Maximizes the variance of the squared loadings within the variables.

Oblique: Loadings lose orthonormality, eg *Oblimin*, *Promax*, *Oblimax*

13

SPEARMAN'S EXAMPLE (EXTENDED)



$$\text{Classics} = \lambda_{11}f_1 + \lambda_{12}f_2 + \varepsilon_1$$

$$\text{French} = \lambda_{21}f_1 + \lambda_{22}f_2 + \varepsilon_2$$

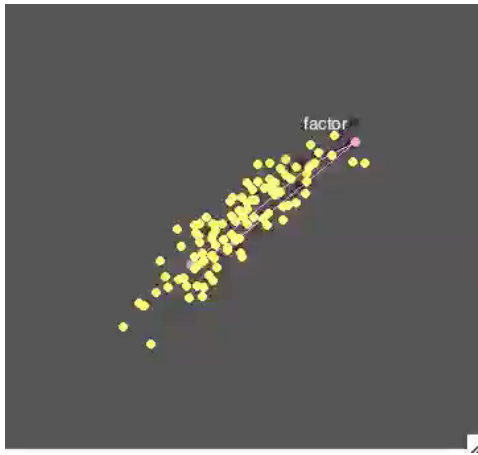
$$\text{English} = \lambda_{31}f_1 + \lambda_{32}f_2 + \varepsilon_3$$

$$\text{Math} = \lambda_{41}f_1 + \lambda_{42}f_2 + \varepsilon_4$$

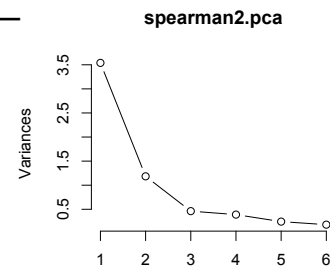
$$\text{Discs} = \lambda_{51}f_1 + \lambda_{52}f_2 + \varepsilon_5$$

$$\text{Music} = \lambda_{61}f_1 + \lambda_{62}f_2 + \varepsilon_6$$

14



True = _____
 Estimated = _____
 Rotated = _____



Variable	MLE Loadings		Varimax Loadings		Specific
Classics	0.77	0.34	0.59	0.61	0.28
French	0.35	0.65	0.73	0.10	0.46
English	0.51	0.73	0.86	0.23	0.20
Math	0.99	-0.07	0.27	0.95	0.02
Discourse	0.51	0.64	0.78	0.26	0.32
Music	0.64	-0.12	0.11	0.64	0.58
Variance	2.63	1.51	2.31	1.83	1.9
% Cumulative	44	69	38	69	100

15

SUMMARY

- These methods are called exploratory factor analysis.
- Confirmatory factor analysis fits a more rigid model.
- Maximum likelihood is an iterative fit, and may not converge, to give estimates.
- Maximum likelihood allows for formal testing for the number of factors needed. Compare this with the scree plot from PCA analysis.

16

This work is licensed under the Creative Commons Attribution-Noncommercial 3.0 United States License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/3.0/us/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.