Exploratory Factor Analysis I

Mengqian Lu

FACTOR ANALYSIS

- A class of procedures to identify underlying dimensions explaining the correlation structure of variables
- Why FA? – Data Reduction and Data Summarization
- What special about FA? – understand the patterns of relationships among DVs, simultaneously discover how IDVs affect them
- Good for problems with too many correlated variables that need to be reduced to a manageable level
  - E.g. Psychologists use FA to understand the profile of a person by analyzing his/her lifestyle statements.
  - E.g. In market research, FA helps to identify customers' groups.

Exploratory Factor Analysis

- Exploratory Factor Analysis is used to determine the number of latent variables that are needed to explain the correlations among a set of observed variables.
- EFA is to discover the factor structure of a measure and to examine its internal reliability.
- Recommended when researchers have no hypotheses about the nature of the underlying factor structure of their measure.
- Three main decision points of EFA: (1) number of factors; (2) extraction method; (3) rotation method.

Some terms to clarify first

- A latent variable: not directly observable, but affect the response variable (manifest variable)
- Latent variable model: (1) represent the effect of unobservable covariates/factors; (2) account for the unobserved heterogeneity between subjects; (3) account for measurement errors – the latent variables represent the "true" outcomes and the manifest variables represent their "disturbed" versions; (4) summarize different measurements of the same unobservable characteristics (e.g. happiness index for qualify-of-life?)
- Factor analysis models (EFA, CFA) are latent variable models

Latent variable models

- Assume that large number of observed random variables $X_1, \ldots, X_p$ can be explained by a smaller set of unobservable (latent) underlying variables; e.g. PCA, EFA, Canonical Correlation Analysis (CCA), Structural Equation Modeling (ESEM; *Asparouhov & Muthén, 2009a*) and Independent Component Analysis (ICA)

In Exploratory Factor Analysis:

- Factors: (continuous) latent variables
- Factor indicators: observed variables (continuous, censored, binary, ordered/ordinal categorical, counts or mixture of these)

Example: Ability and Intelligence Tests

- Six intelligence tests (general, picture, blocks, maze, reading, vocabulary) were given to 112 individuals.
- The covariance matrix is given:

```
> ability.cov
$cov
        general picture  blocks   maze reading   vocab
general  24.641   5.991  33.520  6.023  20.755  29.701
picture   5.991   6.700  18.137  1.782   4.936   7.204
blocks   33.520  18.137 149.831 19.424  31.430  50.753
maze      6.023   1.782  19.424 12.711   4.757   9.075
reading  20.755   4.936  31.430  4.757  52.604  66.762
vocab    29.701   7.204  50.753  9.075  66.762 135.292

$center
[1] 0 0 0 0 0 0

$n.obs
[1] 112
```

Question: Can you find one or two "summary" variable(s)/factor(s) describing some general concept of intelligence out of this correlation?

Example: Ability and Intelligence Tests

- Model:

$$x_{1i} = \lambda_1 f_i + u_{1i}$$
$$x_{2i} = \lambda_2 f_i + u_{2i}$$
$$\ldots$$
$$x_{6i} = \lambda_6 f_i + u_{6i}$$

  - f: common factor ('ability')
  - u: random disturbance s.t. each test
  - λ: factor loadings, give weights/importance of f on x_j

- Underlying assumption: $u_1, u_2, \ldots, u_6$ are uncorrelated → x_1, x_2, x_3 are conditionally uncorrelated given f

$$x_{1i} - \lambda_1 f_i = u_{1i}; x_{2i} - \lambda_2 f_i = u_{2i}, \ldots; x_{6i} - \lambda_6 f_i = u_{6i}$$
$$\mathrm{corr}(u_{1i}, u_{2i}) = 0 = \mathrm{corr}(x_{1i} - \lambda_1 f, x_{2i} - \lambda_2 f)$$

Factor Models

- One-factor model: $\boldsymbol{X}_i = \lambda F_i + \boldsymbol{U}_i, \quad 1 \le i \le n$

  $$X_{ji} = \lambda_j F_i + U_{ji}, \quad 1 \le j \le p, 1 \le i \le n.$$

  where X_{ji} is the j^{th} random variable in a p-dimensional random vector $\boldsymbol{X}_i$, the i^{th} subject.

- F_i is the univariate factor variable defined for each subject i
- $\lambda = (\lambda_1, ..., \lambda_p)^T$ is the factor loadings explain the relation between the factor F and the observed $\boldsymbol{X}$.
- U_{ij} is the error term.

Matrix form: $\mathbf{X}_{pxn} = \boldsymbol{\lambda}_{px1} \mathbf{F}_{1xn} + \mathbf{U}_{pxn}$

Example: Ability and Intelligence Tests

- General model for one individual:

$$x_1 = \mu_1 + \lambda_{11} f_1 + \ldots + \lambda_{1q} f_q + u_1$$

$$\ldots$$

$$x_p = \mu_p + \lambda_{p1} f_p + \ldots + \lambda_{pq} f_q + u_p$$

$$\mathrm{cov}(x_i) = \Sigma = \Lambda\Lambda^T + \Psi$$

To be determined from x:

1. q : no. of common factors
2. Factor loadings Λ
3. Ψ, the cov($\mathbf{u}$)
4. Factor scores f

- Matrix notation for one individual:

$$x = \mu + \Lambda f + u$$

- Matrix notation for *n* individuals:

$$x_i = \mu + \Lambda f_i + u_i \quad (i = 1, \ldots, n)$$

Factor Models (Cont'd)

- Generalized k-factor model:

$$X_{ji} = \mu_j + \sum_{\ell=1}^{k} \lambda_{j\ell} F_{\ell i} + U_{ji}, \quad 1 \le j \le p, 1 \le i \le n$$

Or

$$\boldsymbol{X}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}_{(p\times k)} \boldsymbol{F}_{i(k\times 1)} + \boldsymbol{U}_i, \quad 1 \le i \le n$$

where X_{ji} is the j^{th} random variable in a p-dimensional random vector $\boldsymbol{X}_i$, the i^{th} subject.

- $\boldsymbol{\mu} \in \mathbb{R}^p$ constant
- $\boldsymbol{\Lambda}_{(p\times k)}$ is factor loading matrix, explaining the relation between the factors in $\boldsymbol{F}_i$ and the observed $\boldsymbol{X}_i$
- Errors $\boldsymbol{U}_i$ with $E(\boldsymbol{U}_i) = 0$; $Cov(\boldsymbol{U}_i) = diag(\psi_1, \dots, \psi_p) = \Psi$
- $Cov(\boldsymbol{F}_i, \boldsymbol{U}_i) = \mathbf{0}$, **or** $Cov(u_j, f_s) = 0$ for all j,s

Assumptions

Convention: factors are scaled, otherwise Λ and μ are not well determined.

- $\boldsymbol{F}_{i(k\times 1)}$ is a k-variate random vector with $E(\boldsymbol{F}_i) = 0$; $Cov(\boldsymbol{F}_i) = I_k$

FA: Covariance matrix

$$x = \mu + \Lambda f + u \Leftrightarrow \mathrm{cov}(x) = \Sigma = \Lambda\Lambda^T + \Psi$$

- Factor model is essentially a particular structure imposed on covariance matrix

Still remember the decomposition of covariance matrix in PCA?

$$\boxed{\begin{aligned} Cov(x) &= \frac{1}{n}\sum_i x_i x_i^T = C_x \\ C_x &= U\Lambda U^T \end{aligned}}$$

- Rotations: PCs
- Scores: describe relation between individuals and PCs

FA:

$$\boxed{\mathrm{cov}(x) = \Sigma = \Lambda\Lambda^T + \Psi}$$

FA: Covariance matrix

$$x = \mu + \Lambda f + u \Leftrightarrow \operatorname{cov}(x) = \Sigma = \Lambda\Lambda^T + \Psi$$

- Factor model is essentially a particular structure imposed on covariance matrix
- Instead of spectral decomposition in PCA, we look for a way to split up the variances:

$$\operatorname{var}(x_j) = \sigma_j^2 = \operatorname{var}\left(\sum_{k=1}^{q} \lambda_{jk}^2 F_k\right) + \operatorname{var}(u_j) = \sum_{k=1}^{q} \lambda_{jk}^2 + \psi_j$$

$\sum_{k=1}^{q} \lambda_{jk}^2$	Describe the "Communality", variance due to common factors
ψ_j	Describe the "uniqueness", variance specific to X_j

- "Heywood case" in FA, when $\psi_j \le 0$ (Fabrigar et al.)

Factor Analysis in R: `factanal()`

data "ability.cov" in R

general	a non-verbal measure of general intelligence using Cattell's culture-fair test
picture	a picture-completion test
blocks	block design
maze	mazes
reading	reading comprehension
vocab	vocabulary

Ref: Bartholomew, D. J. (1987) and Bartholomew, D. J. and Knott, M. (1990)

In R: `factanal()` perform maximum-likelihood factor analysis on a covariance matrix or data matrix

Example: Ability and Intelligence Tests

In R: `factanal()` perform maximum-likelihood factor analysis on a covariance matrix or data matrix

Maximum Likelihood Estimation (MLE):

Assume X_i follows Multivariate Normal Distribution (MVN)

The log-likelihood with parameters μ, Λ, Ψ:

$$ll = \log(L) = -\frac{n}{2}\log(|\Sigma|) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^T \Sigma^{-1}(x_i - \mu)$$

$$= -\frac{n}{2}\log(|\Lambda\Lambda^T + \Psi|) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^T \Sigma^{-1}(x_i - \mu)$$

μ is estimated with sample mean, thus choose Λ,Ψ to maximize the log-likelihood

Some notes

- Likelihood function: a product of likelihood functions across all observations

$$L(\mathbf{x}|\mu) = f(X_1|\mu) \times f(X_2|\mu) \times \ldots \times f(X_N|\mu)$$

$$L(\mathbf{x}|\mu) = \prod_{i=1}^{N} f(X_i|\mu) = \left(\frac{1}{2\pi(1)}\right)^{N/2} \exp\left(-\frac{\sum_{i=1}^{N}(X_i-\mu)^2}{2(1)^2}\right)$$

- The value of μ that maximizes f(**X**|μ) is the MLE of population mean, e.g. here μ is the sample mean
- For unknown mean **μ** and variance **Σ**, the likelihood function is:

$$L(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \prod_{i=1}^{N} \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\left(\mathbf{x}_i-\boldsymbol{\mu}\right)'\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}_i-\boldsymbol{\mu}\right)/2\right)$$

- More commonly, it is written as

$$L(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2}}\frac{1}{|\Sigma|^{n/2}} \exp\left(-\sum_{i=1}^{N}\left(\mathbf{x}_i-\boldsymbol{\mu}\right)'\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}_i-\boldsymbol{\mu}\right)/2\right)$$

Number of factors

- MLE approach for estimation provides test:

H_q: q-factor model is sufficient; vs. *H_u: Σ is unconstrained*

- Practical strategy:
  - Start with a small value of q (e.g. q=1) and check hypothesis test; increase q by 1 at a time until some H_q is not rejected

R codes for Intelligence Tests Example

```
ability.FA = factanal(factors = 1, covmat=ability.cov)
update(ability.FA, factors=2)
update(ability.FA, factors=2, rotation="promax")
```

```
> ability.FA
Call:
factanal(factors = 1, covmat = ability.cov)

Uniquenesses:
general picture blocks maze reading vocab
  0.535   0.853  0.748  0.910 0.232  0.280

Loadings:
        Factor1
general 0.682
picture 0.384
blocks  0.502
maze    0.300
reading 0.877
vocab   0.849

               Factor1
SS loadings      2.443
Proportion Var 0.407

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 75.18 on 9 degrees of freedom.
The p-value is 1.46e-12
```

```
> update(ability.FA, factors=2)

Call:
  factanal(factors = 2, covmat = ability.cov)

Uniquenesses:
  general picture  blocks    maze reading   vocab
0.455   0.589   0.218   0.769   0.052   0.334

Loadings:
        Factor1 Factor2
general 0.499   0.543
picture 0.156   0.622
blocks  0.206   0.860
maze    0.109   0.468
reading 0.956   0.182
vocab   0.785   0.225

               Factor1 Factor2
SS loadings      1.858   1.724
Proportion Var   0.310   0.287
Cumulative Var   0.310   0.597

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 6.11 on 4 degrees of freedom.
The p-value is 0.191
```

Hypothesis can not be rejected; for simplicity, we thus use two factors