Exploratory Factor Analysis

Applied Multivariate Statistics – Spring 2013



Latent-variable models

- Large number of observed (manifest) variables should be explained by a few un-observed (latent) underlying variables
- E.g.: Scores on several tests are influenced by "general academic ability"
- Assumes local independence: Manifest variables are independent given latent variables

	Latent variables			
Manifest Variables	Continuous	Categorical		
Continuous	Factor Analysis	Latent Profile Analysis		
Categorical	Item Response Theory	Latent Class Analysis		

Overview

- Introductory example
- The general factor model for x and Σ
- Estimation
- Scale and rotation invariance
- Factor rotation: Varimax
- Factor scores
- Comparing PCA and FA

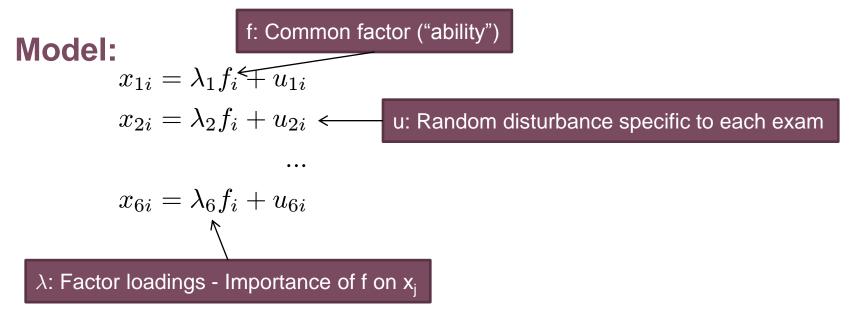
Introductory example: Intelligence tests

- Six intelligence tests (general, picture, blocks, maze, reading, vocab) on 112 persons
- Sample correlation matrix

generalpictureblocksmazereadingvocabgeneral1.0000000.46626490.55166320.34032500.57647990.5144058picture0.46626491.0000000.57243640.19309920.26292290.2392766blocks0.55166320.57243641.00000000.44509010.35402520.3564715maze0.34032500.19309920.44509011.00000000.18396450.2188370reading0.57647990.26292290.35402520.18396451.00000000.7913779vocab0.51440580.23927660.35647150.21883700.79137791.0000000

 Can performance in and correlation between the six tests be explained by one or two variables describing some general concept of intelligence?

Introductory example: Intelligence tests



Key assumption: u_1 , u_2 , u_3 are uncorrelated Thus x_1 , x_2 , x_3 are conditionally uncorrelated given f

R: Function "factanal" in package "stats"

General Factor Model

General model for one individual:

$$x_1 = \mu_1 + \lambda_{11}f_1 + \dots + \lambda_{1q}f_q + u_2$$

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

In matrix notation for one individual:

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

In matrix notation for n individuals:

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

- Assumptions:
 - $Cov(u_j, f_s) = 0$ for all j, s
 - E[u] = 0, $Cov(u) = \Psi$ is a diagonal matrix (diagonal elements = «uniquenesses»)
- Convention:

- E[f] = 0, Cov(f) = identity matrix (i.e. factors are scaled) Otherwise, Λ and μ are not well determined

To be determined from x:

- Number q of common factors
- Factor loadings Λ
- Specific variances Ψ
- Factor scores f

Representation in terms of covariance matrix

Using formulas and assumptions from previous slide:

 $x=\mu+\Lambda f+u \quad \Leftrightarrow \quad \Sigma=\Lambda\Lambda^T+\Psi$

- Factor model = particular structure imposed on covariance matrix "communality": variance
- Variances can be split up:

due to common factors

$$var(x_j) = \sigma_j^2 = \underbrace{\sum_{k=1}^q \lambda_{jk}^2}_{qk} + \underbrace{\psi_j}_{uniqueness}$$
 "specific variance"

- "Heywood case" (= kind of estimation error):
 - $\psi_j < 0$

Estimation: MLE

- Assume x_i follows multivariate normal distribution
- Choose Λ , Ψ to maximize the log-likelihood:

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

Iterative solution, difficult in practice (local maxima)

Number of factors

MLE approach for estimation provides test:

 $H_q: q - factor model holds$ vs $H_u: \Sigma$ is unconstrained

- Modelling strategy: Start with small value of q and increase successively until some H_q is not rejected.
- (Multiple testing problem: Significance levels are not correct)

Intelligence tests revisited: Number of factors

Part of output of R function "factanal":

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

Scale invariance of factor analysis

Suppose y_j = c_jx_j or in matrix notation y = Cx
 (C is a diagonal matrix); e.g. change of measurement units

$$Cov(y) = C\Sigma C^{T} =$$

$$= C(\Lambda\Lambda^{T} + \Psi)C^{T} =$$

$$= (C\Lambda)(C\Lambda)^{T} + C\Psi C^{T} =$$

$$= \hat{\Lambda}\hat{\Lambda}^{T} + \hat{\Psi}$$

I.e., loadings and uniquenesses are the same if expressed in new units

- Thus, using cov or cor gives basically the same result
- Common practice:
 - use correlation matrix or
 - scale input data
 - (This is done in "factanal")

Rotational invariance of factor analysis

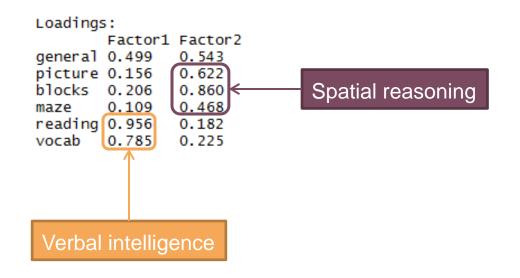
- Rotating the factors yields exactly the same model
- Assume MM^T and transform $f^* = M^T f$, $\Lambda^* = \Lambda M$
- This yields the same model:

 $x^* = \Lambda^* f^* + u = (\Lambda M)(M^T f) + u = \Lambda f + u = x$ $\Sigma^* = \Lambda^* \Lambda^{*T} + \Psi = (\Lambda M)(\Lambda M)^T + \Psi = \Lambda \Lambda^T + \Psi = \Sigma$

- Thus, the rotated model is equivalent for explaining the covariance matrix
- Consequence: Use rotation that makes interpretation of loadings easy
- Most popular rotation: Varimax rotation
 Each factor should have few large and many small loadings

Intelligence tests revisited: Interpreting factors

Part of output of R function "factanal":



Interpretation of factors is generally debatable

Estimating factor scores

- Scores are assumed to be random variables: Predict values for each person
- Two methods:
 - Bartlett (option "Bartlett" in R):
 - Treat f as fix (ML estimate)
 - Thompson (option "regression" in R): Treat f as random (Bayesian estimate)
- No big difference in practice

Case study: Drug use

Loadings:						
	Factor1	Factor2	Factor3	Factor4	Factor 5	Factor6
cigarettes	0.494				0.407	0.110
beer	0.776				0.112	
wine	0.786					
liquor	0.720	0.121	0.103	0.115	0.160	
cocaine		0.519		0.132		0.158
tranquillizers	0.130	0.564	0.321	0.105	0.143	
drug store medication		0.255				0.372
heroin		0.532	0.101			0.190
marijuana	0.429	0.158	0.152	0.259	0.609	0.110
hashish	0.244	0.276	0.186	0.881	0.194	0.100
inhalants	0.166	0.308	0.150		0.140	0.537
hallucinogenics		0.387	0.335	0.186		0.288
amphetamine	0.151	0.336	0.886	0.145	0.137	0.187

Social drugs AmphetamineSmoking

Hard drugs Hashish Inhalants ?

Test of the hypothesis that 6 factors are sufficient. The chi square statistic is 22.41 on 15 degrees of freedom. The p-value is 0.0975 Using different number of factors changes loadings completely

Loadings:

	Factor1	Factor2	Factor3
cigarettes	0.132	0.495	0.352
beer		0.778	0.150
wine		0.781	
liquor	0.193	0.717	0.192
cocaine	0.471		
tranquillizers	0.643	0.114	0.148
drug store medication	0.354		
heroin	0.502		
marijuana	0.237	0.394	0.806
hashish	0.474	0.261	0.395
inhalants	0.498	0.189	0.131
hallucinogenics	0.644		
amphetamine	0.705	0.155	0.208

Significance vs. Relevance:

Might keep less than six factors if fit of correlation matrix is good enough

Comparison: PC vs. FA

- PCA aims at explaining *variances*, FA aims at explaining *correlations*
- PCA is exploratory and without assumptions
 FA is based on statistical model with assumptions
- First few PCs will be same regardless of q First few factors of FA depend on q
- FA: Orthogonal rotation of factor loadings are equivalent This does not hold in PCA
- More mathematically: PCA: $x = \mu + \Gamma_1 z_1 + \Gamma_2 z_2 = \mu + \Gamma_1 z_1 + e$ FA: $x = \mu + \Lambda f + u$ Cov(u) is diagonal by assumption, Cov(e) is not
- Both PCA and FA only useful if input data is correlated !

Concepts to know

- Form of the general factor model
- Representation in terms of covariance matrix
- Scale and Rotation invariance, varimax
- Interpretation of loadings

R functions to know

Function "factanal"