Latent Variable Models: Main features and applications in Social Sciences Part A

Irini Moustaki
Athens University of Economics and Business

Outline

- Some historical background
- Motivation through examples
- Objectives
- Theoretical Framework
- Sufficiency principle
- Models for continuous responses
- Applications

History of Test Scoring

- Harold Gulliksen, 1950 — Theory of Mental Tests
- Luis Guttman, 1950 — Psychometrika, Review of Gulliksen's book
- Fred Lord, 1952, Psychometrika monograph — First to develop a statistical framework for test scoring.
- Classical Test Theory: linear models, ANOVA + Regression

  OBSERVED VARIANCE = TRUE VARIANCE + ERROR

  OBSERVED SCORE = TRUE SCORE + ERROR
- Item Response Theory: Scaling Tradition

Example 1: The Law School Admission Test, Section VI.

The test consisted of 5 items taken by 1000 individuals.

```
MOST FREQUENT RESPONSE PATTERNS
 11 00011          80 10111
 16 01011          16 11000
 15 01111          56 11001
 10 10000          21 11010
 29 10001         173 11011
 14 10010          11 11100
 81 10011          61 11101
 28 10101          28 11110
 15 10110         298 11111
```

```
MARGINS
ITEM     ONES    ZEROS    MISSINGS

    1    0.924    0.076     0.000
    2    0.709    0.291     0.000
    3    0.553    0.447     0.000
    4    0.763    0.237     0.000
    5    0.870    0.130     0.000
```

Aim of the analysis:

- Check whether the five items form a scale.
- Score the individuals based on their responses.

Example 2: Social life feelings study, Schuessler (1982)

Scale used: Economic self-determination, Sample size: 1490 Germans

Yes or no responses were obtained to the following five questions:

1. Anyone can raise his standard of living if he is willing to work at it.
2. Our country has too many poor people who can do little to raise their standard of living.
3. Individuals are poor because of the lack of effort on their part.
4. Poor people could improve their lot if they tried.
5. Most people have a good deal of freedom in deciding how to live.

Example 3

Workplace Industrial Relation Survey dealing with management /worker consultation in firms.

Construct: High commitment management.

Sample size: 1005 firms, concerns non-manual workers.

Please consider the most recent change involving the introduction of new plant, machinery and equipment. Were discussions or consultations of any type on this card held either about the introduction of the change or about the way it was to be implemented?

1. Informal discussions with individual workers.
2. Meetings with group of workers.
3. Discussions in established joint consultative committee.
4. Discussions in specially constituted committee to consider the change.
5. Discussions with unions representatives at the establishment.
6. Discussions with paid union officials from outside.

Example 4: Subject marks, n=220 boys

Table 1: Pairwise correlation coefficients between subject marks

	Gaelic	English	History	Arithmetic	Algebra	Geometry
Gaelic	1.00					
English	0.44	1.00				
History	0.41	0.35	1.00			
Arithmetic	0.29	0.35	0.16	1.00		
Algebra	0.33	0.32	0.19	0.59	1.00	
Geometry	0.25	0.33	0.18	0.47	0.46	1.00

- There is a general tendency for those who do well in one subject to do well in others.

- What is hidden under those correlations?

Example 5: The Height test

1. I bump my head quite often
2. The seat of my bicycle is quite low
3. In bed I often suffer from cold feet
4. When the school picture was taken I was always asked to stand at the back
5. As a police office I would not make much of an impression
6. In airplanes, I usually sit comfortably
7. In libraries, I often has to use a ladder to reach books

Measurement Models

1. Many theories in behavioral and social sciences are formulated in terms of theoretical constructs that are not directly observed or measured.

   Prejudice, ability, radicalism, motivation, wealth.

2. The measurement of a construct is achieved through one or more observable indicators (questionnaire items).
3. The purpose of a measurement model is to describe how well the observed indicators serve as a measurement instrument for the constructs also known as latent variables.
4. Measurement models often suggest ways in which the observed measurements can be improved.
5. In some cases, a concept may be represented by a single latent variable, but often they are multidimensional in nature and so involve more than one latent variable.

Psychology Intelligence Verbal ability Visual Perception	spelling writing word fluency reading punctuation
Education Academic performance	children books library visits TV watching
Sociology Socio-economic status Attitudes towards sex-roles	shovelling snow cleaning the house washing the car making beds

Summarize the objectives

1. Scale construction (Item Response Theory)

2. Study the relationships among a set of observed indicators. Identify the underlying factors that explain the relationships among the observed items.

3. Reduction of dimensionality

   - Fit a latent variable model with one or more factors
   - Fit a latent class model with two or more latent classes

4. Scale individuals on the identified latent dimensions

Latent Variable Models

		Manifest variables		
		Metrical	Categorical	Mixed
Latent variables	Metrical	factor analysis	latent trait analysis	latent trait analysis
	Categorical	latent profile analysis	latent class analysis	latent class analysis

- Categorical: binary, nominal, ordinal
- Factor analysis: classical normal linear factor model
- Latent Trait analysis (Item Response Theory): took its name from Psychometrics
- Latent Class analysis: model based clustering

Types of analysis

- Exploratory Latent Variable Analysis (No theory is known in advance about the data)
- Confirmatory Latent Variable Analysis (validate a theory)

General ideas

• LVM are closely related to the standard regression model. The regression relationship is between a manifest variable and the latent variables.

• Distributional assumptions are made about the residual or error terms which enable us to make inferences.

• The issue is to invert the regression relationships to learn about the latent variables when the manifest variables are given. Since we can never observe the latent variables, we can only ever learn about this relationship indirectly.

• Several manifest variables will usually depend on the same latent variable, and this dependence will induce a correlation between them. The existence of a correlation between two indicators may be taken as evidence of a common source of influence. As long as any correlation remains, we may therefore suspect the existence of a further common source of influence.

References

- Bartholomew, D.J., Steele, F., Moustaki, I. and Galbraith, J. (2002) The Analysis and Interpretation of Multivariate Data in Social Sciences. Chapman and Hall/CRC.
- Bartholomew, D.J. and Knott, M. (1999) Latent Variable Models and Factor Analysis, Kendall Library of Statistics, Arnold.
- Skrondal, A. and Rabe-Hesketh, S. (2005) Generalized Latent Variable Models. Chapman and Hall/CRC.

Item Response Theory (IRT)

IRT consists of a family of models that are useful in the design, construction and evaluation of educational and psychological tests.

1. What IRT models are available? (Rasch Model, Guttman model, Two-parameter model, Partial Credit model, Three-parameter model, Grade of membership model, latent class models etc.)

2. How do we estimate model parameters? (E-M algorithm, Newthon-Raphson, MCMC).

3. How do we assess the fit of the models fitted? (Pearson, Log-likelihood ratio, Residual analysis, Model selection criteria).

4. What software can we use? (BILOG, MULTILOG, PARSCAL, GLAMM, GENLAT)

5. How can we use IRT to construct tests/ select items?

6. How can we use IRT to evaluate the consequences of introducing new items in a test?

7. How can we compare abilities?

   - Use different tests
   - Tests might have different difficulty levels (cannot compare across groups)
   - Tests might have been calibrated using a group of examinees which is different for other groups of examinees (group-dependent)

Characteristics of IRT models

- Measurement Invariance: The instrument used is invariant across groups. (Ability estimates obtained from different sets of items will be the same, also parameter estimates for items obtained in different groups of examinees will be the same).

- Unidimensionality

- Local Independence

Theoretical Framework

Manifest or observed variables or items are denoted by: $x_1, x_2, \ldots, x_p$.

Latent variables / factors / unobserved constructs are denoted by: $y_1, y_2, \ldots, y_q$.

As only $\mathbf{x}$ can be observed any inference must be based on the joint distribution of $\mathbf{x}$:

$$f(\mathbf{x}) = \int_{R_\mathbf{y}} g(\mathbf{x} \mid \mathbf{y}) \phi(\mathbf{y}) d\mathbf{y}$$

$\phi(\mathbf{y})$: prior distribution of $\phi(\mathbf{y})$

$g(\mathbf{x} \mid \mathbf{y})$: conditional distribution of $\mathbf{x}$ given $\mathbf{y}$.

What we want to know: $\phi(\mathbf{y} \mid \mathbf{x}) = \phi(\mathbf{y}) g(\mathbf{x} \mid \mathbf{y}) / f(\mathbf{x})$

$\phi(\mathbf{y})$ and $g(\mathbf{x} \mid \mathbf{y})$ are not uniquely determined.

Conditional Independence

If correlations among the x's can be explained by a set of latent variables then when all y's are accounted for the x's will be independent.

q must be chosen so that:

$$g(\mathbf{x} \mid \mathbf{y}) = \prod_{i=1}^{p} g(x_i \mid \mathbf{y})$$

$\mathbf{y}$ is sufficient to explain the dependencies among the $\mathbf{x}$ variables.

Does $f(\mathbf{x})$ admit the presentation for some small value of q:

$$f(\mathbf{x}) = \int_{R_{\mathbf{y}}} \prod_{i=1}^{p} g(x_i \mid \mathbf{y}) \phi(\mathbf{y}) d\mathbf{y}$$

Conditional Independence: an example

Conditional on the latent variables the responses to items are independent.

	Items		Same ability
	1	2	
1			θ
2			θ
.			θ
.			θ
.			θ
N			θ
	$b_1 < b_2$		(difficulty)

Since all individuals have the same ability response to item 1 does not give any information regarding the response to item 2 in other words there is no systematic reason why their responses differ.

Consider the example of children's writing:

if x_1 is foot size, x_2 is writing ability and y is the single variable, age, then x_1 and x_2 are positively correlated, but $conditional$ on y they are uncorrelated:

$$\begin{aligned} \mathsf{Corr}(x_1, x_2) &> 0, \\ \mathsf{Corr}(x_1, x_2 \,|\, y) &= 0. \end{aligned}$$

Differences in age fully account for the apparent correlation between foot size and writing ability.

Normal Linear Factor Model for Continuous Responses

$\mathbf{x}$ and $\mathbf{y}$ continuous.

$$\mathbf{x} = \boldsymbol{\mu} + \Lambda\mathbf{y} + \mathbf{e}$$

The prior distribution for the latent variables:

$$\mathbf{y} \sim N_q(\mathbf{0}, I)$$

Assumptions:

$$\mathbf{e} \sim N_p(\mathbf{0}, \Psi)$$

$$\mathrm{Cov}(\mathbf{y}, \mathbf{e}) = 0$$

$$E(x_i x_j \mid \mathbf{y}) = 0$$

$$\mathbf{x} \mid \mathbf{y} \sim N_p(\boldsymbol{\mu} + \Lambda \mathbf{y}, \Psi) \tag{1}$$

$$\mathbf{x} \sim N_p(\boldsymbol{\mu}, \Lambda\Lambda' + \Psi)$$

- The $\Lambda\Lambda'$ is the part of the variance of the observed items explained by the factors (communality).
- The Ψ part is the residual or specific variance.
- The covariances between the xs depend only on the factor loadings.

$$\mathbf{y} \mid \mathbf{x} \sim N_q(\Lambda'(\Lambda\Lambda' + \Psi)^{-1}(\mathbf{x} - \boldsymbol{\mu}), (\Lambda'\Psi^{-1}\Lambda + I)^{-1})$$

Arbitrariness of the model

- Note that

$$f(\mathbf{x}) = \int_{R_{\mathbf{y}}} g(\mathbf{x} \mid \mathbf{y})\phi(\mathbf{y})d\mathbf{y}$$

is not unique. A one-to-one transformation of the factor space from $\mathbf{y}$ to $\mathbf{z}$ will leave the $f(\mathbf{x})$ unchanged but will change both the g and ϕ functions.

- However, some transformations will be more interpretable than others.

- The indeterminacy of ϕ leave us free to adopt a metric for $\mathbf{y}$.

- If $\mathbf{x}$ is normal there is an important transformation which leaves the form of ϕ unchanged and which leaves a degree of arbitrariness for g.

Suppose $q \geq 2$, then the orthogonal transformation $\mathbf{z} = \mathbf{My}$, $(\mathbf{M}'\mathbf{M} = \mathbf{I})$ gives

$$\mathbf{z} \sim N_q(\mathbf{0}, \mathbf{I})$$

which has the same distribution as $\mathbf{y}$.

The conditional distribution is now:

$$\mathbf{x} \mid \mathbf{z} \sim N_p(\boldsymbol{\mu} + \Lambda M'\mathbf{z}, \Psi)$$

That model cannot be distinguished from the one with weights $\Lambda M'$. The joint distribution of $\mathbf{x}$ is unaffected. In both cases the covariance matrix is $\Lambda\Lambda' + \Psi$.

The advantage is that it allows the researcher to choose among different solutions the most interpretable one.

Interpretation: naming latent variables

• Look at the magnitude of the factor loadings α_{ij}, where $(+)$ large positive loadings and (.) small loadings

$$A = \begin{pmatrix} + & . & . \\ + & . & . \\ + & . & . \\ . & + & . \\ . & + & . \\ . & + & . \\ . & . & + \\ . & . & + \\ . & . & + \end{pmatrix}$$

• Components: $X_j = \sum_{i=1}^{p} \alpha_{ij} u_i(x_i)$

Sufficiency Principle

Aim: Reduce the dimensionality of $\mathbf{x}$ from p to q where q is much less than p.

Find q functions of $\mathbf{x}$, $X_1, X_2, \ldots, X_q$ so that the conditional distribution given X does not depend on $\mathbf{y}$.

Barankin and Maitra (1963):

A necessary and sufficient condition subject to weak regularity conditions is that at least $p - q$ of the g_i shall be of the exponential family.

The X_j, $j = 1, \ldots, q$ are called components.

Sufficiency Principle, continued

All the information about the latent variable in $\mathbf{x}$ can be found by the posterior distribution:

$$\phi(\mathbf{y} \mid \mathbf{x}) = \phi(\mathbf{y}) \prod_{i=1}^{p} g_i(x_i \mid \mathbf{y}) / f(\mathbf{x})$$

Substitute the exponential family density for $g_i(x_i \mid \mathbf{y})$:

$$\phi(\mathbf{y} \mid \mathbf{x}) \propto \phi(\mathbf{y}) \left[\prod_{i=1}^{p} G_i(\theta_i) \right] \exp \sum_{j=1}^{q} y_j X_j$$

where $X_j = \sum_{i=1}^{p} \alpha_{ij} u_i(x_i)$, $j = 1, \ldots, q$.

• The posterior distribution of $\mathbf{y}$ depends on $\mathbf{x}$ through the q-dimensional vector $\mathbf{X}' = (X_1, \ldots, X_q)$, $\mathbf{X}$ is a minimal sufficient statistic for $\mathbf{y}$.

- The reduction does not depend on $\phi(\mathbf{y})$.

Example

x_i: Bernoulli random variable, $x_i = 0$ or 1.

$g(x_i \mid y) = \pi_i^{x_i}(1-\pi_i)^{1-x_i} = (1-\pi_i)\exp(x_i \text{logit}\pi_i)$.

$g(x_i \mid \theta_i) = F_i(x_i)G_i(\theta_i)\exp(\theta_i u_i(x_i))$

$\theta_i = \text{logit}\pi_i = \log_e \frac{\pi_i}{1-\pi_i} = \alpha_{i0} + \alpha_{i1}y_1 + \cdots + \alpha_{iq}y_q$

$G_i(\theta_i) = 1 - \pi_i$

$u_i(x_i) = x_i$

The GLLVM: $\text{logit}\pi_i = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij}y_j$

Components: $X_j = \sum_{i=1}^{p} \alpha_{ij}x_i$

π_i: probability of a positive response given y.

Model Interpretation

- Covariance between $\mathbf{x}$ and $\mathbf{y}$

$$E(\mathbf{x}-\boldsymbol{\mu})\mathbf{y}' = E[E(\mathbf{x}-\boldsymbol{\mu})\mathbf{y}' \mid \mathbf{y}] = E[E\{(\mathbf{x}-\boldsymbol{\mu}) \mid \mathbf{y}\}\mathbf{y}'] = E(\Lambda\mathbf{y}\mathbf{y}') = \Lambda I = \Lambda$$

Factor loadings are covariances between individual manifest variables and factors.

The correlations are given by: $(diag\Sigma)^{-1/2}\Lambda$

Model estimation

The estimation procedure is based on the maximization of the marginal likelihood of the manifest variables given by:

$$\log f(\mathbf{x}_h) = \log \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\mathbf{x}_h \mid \mathbf{y}) \phi(\mathbf{y}) d\mathbf{y}$$

where $\mathbf{x}_h$ represents a vector with all the responses to the p manifest variables of the h^{th} individual.

Here $\mathbf{\Lambda}$ is $p \times q$ matrix of factor loadings and $\mathbf{\Psi}$ is a $p \times p$ diagonal matrix of specific variances for the continuous items.

The matrix $\mathbf{\Lambda}_{p \times q}$ contains the covariances between elements of $\mathbf{y}$ and $\mathbf{x}$. The parameters λ_{ij} can be standardized in order to express the correlation between the observed variable i and the latent variable j.

E-M algorithm

Since the latent variables are unobserved we use the E-M algorithm.

For a random sample of size n, the complete likelihood is:

$$l = \prod_{h=1}^{n} f(\mathbf{x}_h, \mathbf{y}_h)$$

The log-likelihood is:

$$\log l = L = \sum_{h=1}^{n} [\log g(\mathbf{x}_h \mid \mathbf{y}_h) + \log \phi(\mathbf{y}_h)]$$

Using the assumption of conditional independence:

$$g(\mathbf{x} \mid \mathbf{y}) = \prod_{i=1}^{p} g(x_i \mid \mathbf{y})$$

The E-M algorithm requires to compute the expected value of the score function. The expectation is taken with respect to the posterior distribution of the latent variables based on the observed variables. $(\phi(\mathbf{y} \mid \mathbf{x}))$.

The score function of the ML are the first derivatives:

$$E\left(\frac{\partial L}{\partial \mu_i}\right) = \int \cdots \int \frac{\partial L}{\partial \mu_i} \phi(\mathbf{y} \mid \mathbf{x}) d\mathbf{y} \tag{2}$$

$$E\left(\frac{\partial L}{\partial \lambda_{ij}}\right) = \int \cdots \int \frac{\partial L}{\partial \lambda_{ij}} \phi(\mathbf{y} \mid \mathbf{x}) d\mathbf{y} \tag{3}$$

$$E\left(\frac{\partial L}{\partial \Psi_{i}}\right) = \int \cdots \int \frac{\partial L}{\partial \Psi_{i}} \phi(\mathbf{y} \mid \mathbf{x}) d\mathbf{y} \tag{4}$$

The integrals can be approximated to any practical degree of accuracy by Gauss-Hermite quadrature, Laplace approximation, Monte Carlo, Adaptive quadrature.

E-M steps

1. Give initial values to the model parameters.

2. Compute the expected score functions.

3. Obtain new estimates of the parameters from the MLE using the values of the expectation step. (M-step)

4. Check convergence.

Initial values of the parameters are chosen ad hoc. Different initial values are used to check the convergence of the EM algorithm to a global maximum.

Adequacy of the model and choice of the number of factors

1. *Percentage of variance explained by the factors*

   The communalities $(\hat{\Lambda}'\hat{\Lambda})$ are used to check that the individual observable variables are adequately explained by the factors.

2. *Reproduced correlation matrix*

   Compare the fitted (reproduced) correlation matrix of the xs with the correlation matrix computed from the sample data.

3. *Goodness-of-fit test:* If q is specified a priori

$$Ho : \Sigma = \Lambda\Lambda' + \Psi$$

$$H_1 : \Sigma \text{ is unconstrained } (\hat{\Sigma} = S)$$

Using a likelihood ratio statistic

$$W = -2\{L(Ho) - L(H_1)\} = n\{\log|\hat{\Sigma}| + trace\hat{\Sigma}^{-1}S - \log|S| - p\}$$

where $\hat{\Sigma} = \hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}$

If $\Psi > 0$ then $-2\{L(Ho) - L(H_1)\} \sim \chi^2$ with

$$d.f. = \frac{1}{2}p(p+1) - \{pq + p - \frac{1}{2}q(q-1)\} = \frac{1}{2}\{(p-q)^2 - (p+q)\}$$

Failure to reject this null hypothesis would imply a good fit.

4. The number of factors, q, must be small enough for the degrees of freedom $[(p-q)^2 - (p+q)/2]$ to be greater than or equal to zero. So when $p = 3$ or $p = 4$, q cannot be greater than one, but when $p = 20$, q could be as large as 14.

Factor scores

- Posterior mean for each response pattern:

$$E(y_j \mid \mathbf{x}), j = 1, \cdots, q$$

- Component scores:

$$X_j = \sum_{i=1}^{p} \frac{\lambda_{ij}}{\sqrt{\Psi_{ii}}} x_i$$

- Regression scores.

Example: Subject marks, n=220 boys

Table 2: Pairwise correlation coefficients between subject marks

	Gaelic	English	History	Arithmetic	Algebra	Geometry
Gaelic	1.00					
English	0.44	1.00				
History	0.41	0.35	1.00			
Arithmetic	0.29	0.35	0.16	1.00		
Algebra	0.33	0.32	0.19	0.59	1.00	
Geometry	0.25	0.33	0.18	0.47	0.46	1.00

• There is a general tendency for those who do well in one subject to do well in others.

Factor loadings, subject marks

Subject	$\hat{\lambda}_{i1}$	$\hat{\lambda}_{i2}$
Gaelic	0.56	0.43
English	0.57	0.29
History	0.39	0.45
Arithmetic	0.74	−0.28
Algebra	0.72	−0.21
Geometry	0.60	−0.13

- The first factor measures overall ability in the six subjects.
- The second factor contrasts humanities and mathematics subjects.

Communalities, subject marks

Communality of a standardized observable variable is the proportion of the variance that is explained by the common factors.

	Communalities
Gaelic	0.49
English	0.41
History	0.36
Arithmetic	0.62
Algebra	0.56
Geometry	0.37

- 49% of the variance in Gaelic scores is explained by the two common factors ($0.56^2 + 0.43^2 = 0.49$).
- The larger the communality, the better does the variable serve as an indicator of the associated factors.

• The sum of the communalities is the variance explained by the factor model. For the example is 2.81 or 47% of 6 which is the total variance for the subject marks data.

Table 3: Reproduced correlations and communalities (top section) for a linear two-factor model fitted to the subject marks data, and discrepancies between observed and reproduced correlations (bottom section), subject marks data

Correlation	Gaelic	English	History	Arithmetic	Algebra	Geometry
Gaelic	0.49	0.44	0.41	0.29	0.31	0.28
English	0.44	0.41	0.35	0.34	0.35	0.30
History	0.41	0.35	0.36	0.16	0.19	0.17
Arithmetic	0.29	0.34	0.16	0.62	0.59	0.48
Algebra	0.31	0.35	0.19	0.59	0.56	0.46
Geometry	0.28	0.30	0.17	0.48	0.46	0.37
Discrepancy						
Gaelic		0.00	0.00	0.00	0.02	−0.03
English	0.00		0.00	0.01	−0.03	0.03
History	0.00	0.00		0.00	0.00	0.00
Arithmetic	0.00	0.01	0.00		0.00	0.00
Algebra	0.00	−0.03	0.00	0.00		0.00
Geometry	−0.03	0.03	0.00	0.00	0.00	

Rotation in two-factor model

1. Rotation does not change the fit of the model.
2. Rotation does not change the reproduced correlation matrix or the goodness-of-fit test statistic.
3. The communalities remain unchanged.
4. This is because rotation has not changed the relative positions of the loadings.
5. Since rotation alters the loadings, the interpretation of the new factors will be different. Also, although the overall percentage of variance explained by the common factors remains the same after rotation, the percentage of variance explained by each factor will change. Rotation redistributes the explained variance across the factors.

Ways of doing it

- Orthogonal rotation

Some procedures have been developed to search automatically for a suitable rotation. For example, the VARIMAX procedure attempts to find an orthogonal rotation that is close to simple structure by finding factors with few large loadings and as many near-zero loadings as possible.

- Non-orthogonal (oblique) rotation.

This type of rotation requires us to relax the original assumption of the linear factor model that the latent variables be uncorrelated. An oblique rotation leads to correlated factors.

The correlation between these transformed factors is 0.515.

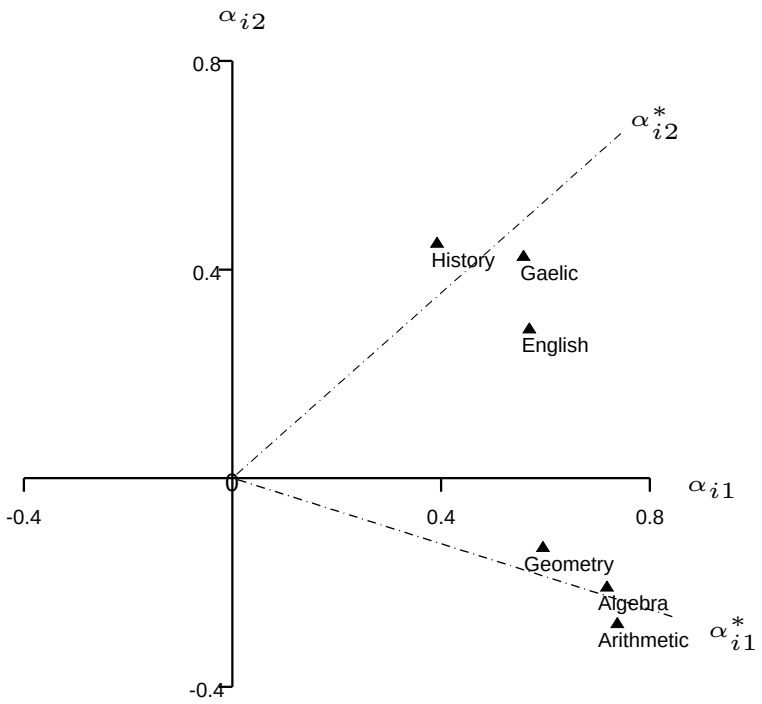

Figure 1: Plot of unrotated and rotated factor loadings for the subject marks data

Latent Variable Models - Part B

- Factor models for categorical responses
- Applications
- Latent class models
- Applications
- Software and references

Models for Binary Data - Notation

Suppose there are p items or questions to which the respondent is required to give a binary response: right/wrong, agree/disagree, yes/no.

With p variables, each having two outcomes, there are 2^p different response patterns which are possible.

The binary observed variables are denoted with $(x_1, \cdots, x_p)$.

x_i: independent Bernoulli variables taking values 0 and 1.

The latent variables are denoted with $y_1, y_2, \ldots y_q$ where q is much less than p.

The individuals in the sample are denoted with h where $h = 1, \ldots, n$.

Factor analysis principles

For a given set of response variables $x_1, \ldots, x_p$ one wants to find a set of latent factors $y_1, \ldots, y_q$, fewer in number than the observed variables, that contain essentially the same information.

If both the response variables and the latent factors are normally distributed with zero means and unit variances, this leads to the model

$$E(x_i \mid y_1, y_2, \ldots, y_q) = \lambda_{i1} y_1 + \lambda_{i2} y_2 + \cdots + \lambda_{iq} y_q \ ,$$

If the response variables are binary we specify instead the probability of each response pattern as a function of $y_1, y_2, \ldots, y_q$:

$$Pr(x_1 = a_1, x_2 = a_2, \ldots, x_p = a_p \mid y_1, y_2, \ldots, y_q) = f(y_1, y_2, \ldots, y_q)$$

Literature Approaches

Approach A

Item Response Theory Approach: response function that gives the probability of a positive response for an individual with latent position $\mathbf{y}$.

$$P(x_i = 1 \mid \mathbf{y})$$

Approach B:

Underlying variable approach, supposes that the binary x's have been produced by dichotomizing underlying continuous variables.

The connection between the binary variable x_i and the underlying variable x_i^* is

$$x_i = 0 \iff \infty < x_i^* \leq \tau_{(i)}$$

$$x_i = 1 \iff \tau_{(i)} < x_i^* \leq +\infty$$

The τ are called threshold values.

The connection between the ordinal variable x_i and the underlying variable x_i^* is

$$x_i = a \quad \Longleftrightarrow \quad \tau_{a-1}^{(i)} < x_i^* \leq \tau_a^{(i)},\ a = 1, 2, \ldots, m_i\ ,$$

where

$$\tau_0^{(i)} = -\infty\ ,\ \tau_1^{(i)} < \tau_2^{(i)} < \ldots < \tau_{m_i-1}^{(i)}\ ,\ \tau_{m_i}^{(i)} = +\infty\ ,$$

For variable x_i with m_i categories, there are $m_i - 1$ threshold parameters.

Since only ordinal information is available about x_i^*, the mean and variance of x_i^* are not identified and are therefore set to zero and one, respectively

Note: Model parameters equivalence exist between the two approaches.

Underlying Variable Approach - Structural Equation Modelling

All the variables are treated as metric through assumed underlying and normal variables and by using ML, GLS or WLS as the estimation method.

Contributors

- Muthén and Muthén (M-Plus)
- Jöreskog and Sörbom (LISREL)
- Bentler (EQS)

Their work covers a wide range of models that allows relationships among the latent variables, inclusion of exogenous (explanatory) variables, multilevel analysis, analysis of panel data.

Item Response Theory Approach

The response function is modelled through a logistic model:

$$\text{logit}\pi_i(\mathbf{y}) = \alpha_{i0} + \alpha_{i1}y_1 + \alpha_{i2}y_2 + \cdots + \alpha_{iq}y_q$$

where

$$\pi_i(\mathbf{y}) = P(x_i = 1 \mid \mathbf{y})$$

is the response function and $y_1, y_2, \ldots, y_q$ are independently and normally distributed variables with mean 0 and variance 1.

Note: When $\mathbf{y}$ is unidimensional, $\pi_i(y)$ is referred to as item characteristic curve or item response function and the model is known as the two-parameter logistic model (2PL).

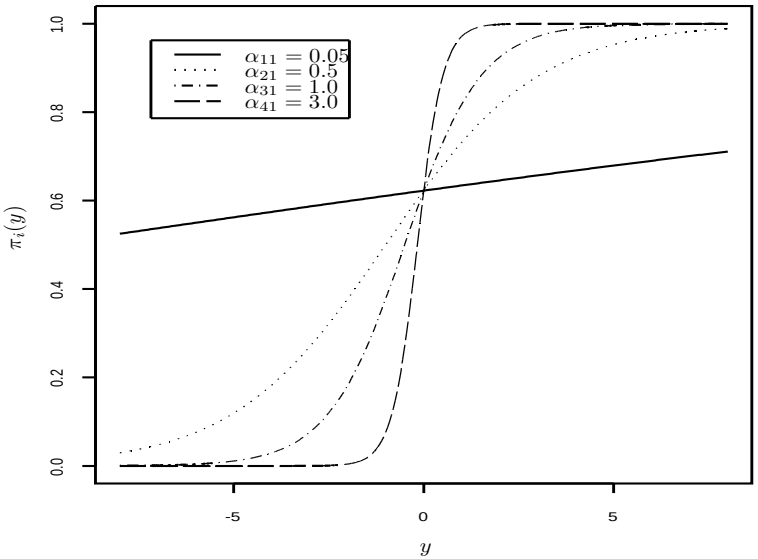

Figure 2: Item characteristic curves for different values of the discrimination coefficient α_{i1} and $\alpha_{i0} = 0.5$

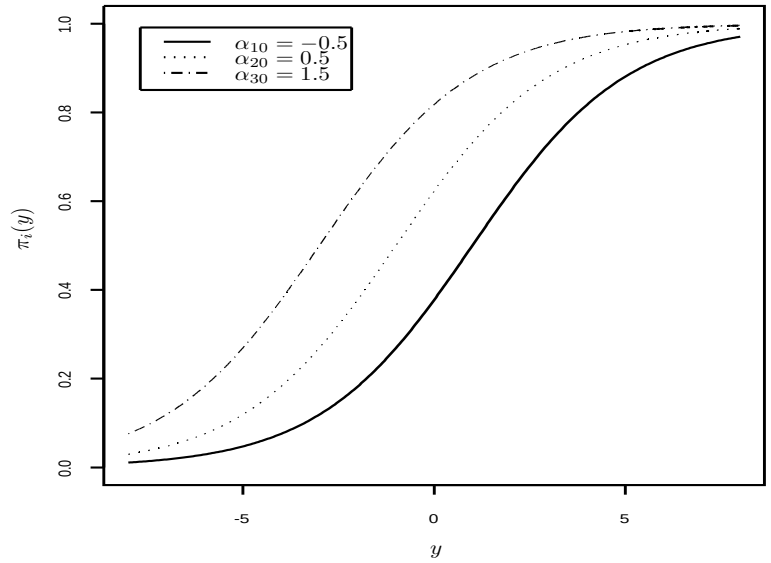

Figure 3: Item characteristic curves for different values of the "difficulty" parameter α_{i0} and $\alpha_{i1} = 0.5$

When this model is used for the form of the response function we can show that the $\phi(\mathbf{y} \mid \mathbf{x})$ depends on $\mathbf{x}$ through the component:

$$X_j = \sum_{i=1}^{p} \alpha_{ij} x_i, \qquad j = 1, \cdots, q$$

For the latent variables we choose the standard normal distribution because the factor axes can be rotated without affecting the model. In other words an orthogonal transformation of the factor loadings (factor coefficients) will leave the value of the likelihood unchanged.

The probit response function: Bock and Aitkin (1981):

$$x^*_{ih} = \alpha_{i1}y_1 + \alpha_{i2}y_2 + \cdots + \alpha_{iq}y_q + \epsilon_{ih}$$

where $i = 1, \ldots, p$ and $h = 1, \ldots, n$.

The model describes not an observed variable, but an unobservable 'response process'. The process generates a positive response for item i from an individual h when the x^*_{ih} equals or exceeds a threshold τ_i and gives a negative response otherwise.

On the assumption that $\epsilon_{ih} \sim N(0, \sigma^2_i)$:

$$
\begin{aligned}
P(x_i = 1 \mid \mathbf{y}) &= \frac{1}{(2\pi)^{1/2}\sigma} \int_{\tau_i}^{\infty} \exp\left\{ -\frac{1}{2} \left(\frac{x_i^* - \sum_{j=1}^{q} \alpha_{ij} y_j}{\sigma_i} \right)^2 \right\} dx_i^* \\
&= \Phi\left(-\frac{\tau_i - \sum_{j=1}^{q} \alpha_{ij} y_j}{\sigma_i} \right) \\
&= \Phi(\mathbf{y}).
\end{aligned}
$$

This is the Normal ogive model. In practice the difference is small:

$$
\text{logit}(u) = \frac{\pi}{\sqrt{3}} \Phi^{-1}(u)
$$

Interpretation of Parameters

The coefficient α_{i0} is the value of logit$\pi_i(\mathbf{y})$ at $\mathbf{y} = \mathbf{0}$. The probability of a positive response from the median individual.

In educational testing, α_{i0} is called **'difficulty'** parameter.

$$\pi_i(\mathbf{0}) = P(x_i = 1 \mid \mathbf{0}) = \frac{\exp(\alpha_{i0})}{1 + \exp(\alpha_{i0})}$$

The coefficients $\alpha_{ij}, \quad j = 1, \ldots, q$ are called **'discrimination'** coefficients.

α_{ij} are the weights used in the component function to weight the individual's responses to the p observed items.

They also measure the extent to which the latent variable y_j discriminates between individuals.

Standardized α_{ij} 's

$$\alpha_{ij}^{*} = \frac{\alpha_{ij}}{\sqrt{\sum_{j=1}^{q} \alpha_{ij}^{2} + 1}}$$

This standardization brings the interpretation close to factor analysis (factor loadings express correlation between the observed items and the latent variables).

Summary of the model

- Assumption of Conditional Independence: Responses to the p observed items are independent given the vector of the latent variables.

$$g(\mathbf{x} \mid \mathbf{y}) = \prod_{i=1}^{p} g(x_i \mid \mathbf{y})$$

- Independent latent variables with standard normal distributions:

$$\phi(\mathbf{y}) = \phi(y_1)\phi(y_2)\cdots\phi(y_q)$$

- Bernoulli distribution for $x_i \mid \mathbf{y}$

$$g(x_i \mid \mathbf{y}) = \pi_i(\mathbf{y})^{x_i}(1 - \pi_i(\mathbf{y}))^{1-x_i}$$

where $\pi_i(\mathbf{y}) = P(x_i = 1 \mid \mathbf{y})$

- link function: logit or probit
  - logit$\pi_i(\mathbf{y}) = \alpha_{i0} + \alpha_{i1} z_1 + \cdots + \alpha_{iq} y_q$
  - $\Phi^{-1}(\pi_i(\mathbf{y})) = \alpha_{i0} + \alpha_{i1} y_1 + \cdots + \alpha_{iq} y_q$
- Component Scores: $\sum_{i=1}^{p} \alpha_{ij} x_i$, $j = 1, \ldots, q$.

Model estimation

The estimation procedure is based on the maximization of the marginal likelihood of the manifest variables given by:

$$\log f(\mathbf{x}_h) = \log \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\mathbf{x}_h \mid \mathbf{y}) \phi(\mathbf{y}) d\mathbf{y}$$

where $\mathbf{x}_h$ represents a vector with all the responses to the p manifest variables of the h^{th} individual and $\phi(\mathbf{y})$ is the prior distribution of the latent variables, assumed to have independent standard normal distributions.

$$g(x_i \mid \mathbf{y}) = [\pi_i(\mathbf{y})]^{x_i} [1 - \pi_i(\mathbf{y})]^{1-x_i} \qquad i = 1, \cdots, p$$

where $\pi_i(\mathbf{y}) = Pr(x_i = 1 \mid \mathbf{y})$ is the response function for binary item i.

For a random sample of n individuals the loglikelihood is written as:

$$\log L = \sum_{h=1}^{n} \log f(\mathbf{x}_h)$$

The estimation is done with the E-M algorithm.

Goodness-of-Fit

Compare the observed (O) and expected (E) frequencies of the 2^p response patterns by means of a X^2 Pearson Goodness-of-fit or a likelihood ratio test G^2.

$$X^2 = \sum_{i=1}^{2^p} \frac{(O_i - E_i)^2}{E_i}$$

$$G^2 = 2 \sum_{i=1}^{2^p} O_i \log \frac{O_i}{E_i}$$

When n is large and p small the above statistics follow a chi-square distribution with degrees of freedom equal to: $2^p - p(q+1) - 1$.

As the number of items increases the chi-square approximation to the distribution of either goodness-of-fit statistic ceases to be valid. Parameter estimates are still valid but it is difficult to assess the model.

Example:
$p = 10$ $2^p = 1024$ $n = 1000$. With this data we expect that there will be many response patterns with $E_i \leq 1.0$.

Solutions

1. Group the response patterns with expected frequencies less than 5.0. There is a danger of being left out with no degrees of freedom.

2. Compute a measure of the total amount of association explained by the model.

$$\frac{G^2(H_o) - G^2(H_1)}{G^2(H_o)} \times 100\%$$

$G^2(H_o)$ is the likelihood ratio statistic under the assumption that the responses are mutually independent and

$G^2(H_1)$ is the likelihood ratio statistics under the fitted latent variable model.

3. Examination of residuals.

   Compare the observed and expected frequencies for pair and triplets of responses. If these differences are small it means that the associations between all pairs of responses are well predicted by the model. Check whether pairs or triples of responses occur more or less, often than the model predicts. The above given discrepancy measures can be used to measure discrepancies in the margins. The residuals are not independent and so not a formal test can be applied. However, if we consider the distribution of each residual as a chi-square with 1 degree-of-freedom then a residual with a X^2 or G^2 value greater than 4 will indicate a poor fit.

Diagnostics procedures based on residuals:

- Give reasons for poor fit.
- Suggest ways in which the scales may be improved.

Example 1

The Law School Admission Test, Section VI.

Sample size =1000

This is a classical example in educational testing. The test consisted of 5 items taken by 1000 individuals. The main interest is whether the 5 items form a scale or in other words whether their interrelationships can be explained by a single factor named ability.

Example 1: Analysis

item	$\hat{\alpha}_{i0}$	s.e.	$\hat{\alpha}_{i1}$	s.e.	st$\hat{\alpha}_{i1}$	$\hat{\pi}_i(0)$
1	2.77	(0.20)	0.83	(0.25)	0.64	0.94
2	0.99	(0.09)	0.72	(0.19)	0.59	0.73
3	0.25	(0.08)	0.89	(0.23)	0.67	0.56
4	1.28	(0.10)	0.69	(0.19)	0.57	0.78
5	2.05	(0.13)	0.66	(0.20)	0.55	0.89

- All items have similar factor loadings (discrimination power)

- The easiest item is the first one.

Table 4: Factor scores in increasing order, LSAT data

Observed frequency	Expected frequency	$\hat{E}(y \mid \mathbf{x})$	$\hat{\sigma}(y \mid \mathbf{x})$	Component score (X_1)	Total score	Response pattern
3	2.3	−1.90	0.80	0.00	0	00000
6	5.9	−1.47	0.80	0.66	1	00001
2	2.6	−1.45	0.80	0.69	1	00010
1	1.8	−1.43	0.80	0.72	1	01000
10	9.5	−1.37	0.80	0.83	1	10000
1	0.7	−1.32	0.80	0.89	1	00100
11	8.9	−1.03	0.81	1.35	2	00011
8	6.4	−1.01	0.81	1.38	2	01001
29	34.6	−0.94	0.81	1.48	2	10001
16	13.6	−0.55	0.82	2.07	3	01011
81	76.6	−0.48	0.82	2.17	3	10011
56	56.1	−0.46	0.82	2.21	3	11001
21	25.7	−0.44	0.82	2.24	3	11010
28	25.0	−0.35	0.82	2.37	3	10101
15	11.5	−0.33	0.82	2.40	3	10110
11	8.4	−0.30	0.82	2.44	3	11100
173	173.3	0.01	0.83	2.89	4	11011
15	13.9	0.05	0.84	2.96	4	01111
80	83.5	0.13	0.84	3.06	4	10111
61	62.5	0.15	0.84	3.10	4	11101
28	29.1	0.17	0.84	3.13	4	11110
298	296.7	0.65	0.86	3.78	5	11111

FIRST AND SECOND ORDER OBSERVED AND EXPECTED MARGINS

RESPONSE (1,1) to ITEMS (I,J)

I	J	OBSER	EXPECT	OBS-EXP	((O-E)**2)/E
1	1	924	924.0009	-0.0009	0.0000
2	1	664	663.1141	0.8859	0.0012
2	2	709	708.9945	0.0055	0.0000
3	1	524	521.4167	2.5833	0.0128
..	..	...	.....	.....	.....
5	1	806	808.3290	-2.3290	0.0067
5	2	630	626.9241	3.0759	0.0151
5	3	490	494.5676	-4.5676	0.0422
5	4	678	672.4921	5.5079	0.0451
5	5	870	869.9991	0.0009	0.0000

THIRD ORDER OBSERVED AND EXPECTED MARGINS
RESPONSE (1,1,1) to ITEMS (I,J,J1)

I	J	J1	OBSER	EXPECT	OBS-EXP	((O-E)**2)/E
1	2	3	398	396.7774	1.2226	0.0038
1	2	4	520	524.7850	-4.7850	0.0436
1	2	5	588	588.6264	-0.6264	0.0007
1	3	4	421	420.8115	0.1885	0.0001
1	3	5	467	467.7153	-0.7153	0.0011
1	4	5	632	630.0993	1.9007	0.0057
2	3	4	343	341.7304	1.2696	0.0047
2	3	5	377	377.4815	-0.4815	0.0006
2	4	5	502	497.4927	4.5073	0.0408
3	4	5	397	400.0829	-3.0829	0.0238

Example 2: Women's mobility in Bangladesh

The particular dimension that we shall focus on here is women's mobility or social freedom. Women were asked whether they could engage in the following activities alone (1=yes, 0=no).

1. Go to any part of the village/town/city.
2. Go outside the village/town/city.
3. Talk to a man you do not know.
4. Go to a cinema/cultural show.
5. Go shopping.
6. Go to a cooperative/mothers' club/other club.
7. Attend a political meeting.
8. Go to a health centre/hospital.

Example 2: Goodness-of-fit measures

- The one-factor model gives a G^2 equal to 364.5 on 39 degrees of freedom indicating a bad fit.

- The two-factor model is still rejected based on a G^2 equal to 263.41 on 33 degrees of freedom.

- The percentage of G^2 explained increases only slightly from 94.98% to 96.92%.

Table 5: Chi-squared residuals greater than 3 for the second and the (1,1,1) third order margins for the one-factor model, women's mobility data

Response	Items	O	E	$O-E$	$(O-E)^2/E$
(0,1)	3, 2	187	229.19	−42.19	7.76
	7, 6	532	596.04	−64.04	6.88
	8, 5	194	245.15	−51.15	10.67
(1,0)	2, 1	52	117.29	-65.29	36.35
	5, 1	13	3.02	9.99	32.92
	6, 2	274	196.34	77.66	30.71
	7, 1	6	1.13	4.87	20.97
	7, 2	62	36.82	25.18	17.21
	7, 6	41	93.69	−52.69	29.63
	8, 1	28	7.15	20.85	60.83
	8, 3	38	22.74	15.26	10.24
(1,1)	6, 2	665	756.15	−91.15	10.99
	7, 6	407	356.45	50.55	7.17
(1,1,1)	1, 2, 3	2433	2338.67	94.33	3.80
	1, 2, 6	659	751.02	−92.02	11.27
	2, 4, 6	637	704.12	−67.12	6.40
	6, 7, 8	318	267.09	50.91	9.70

Table 6: Chi-squared residuals greater than 3 for the second and the (1,1,1) third order margins for the two-factor model, women's mobility data

Response	Items	O	E	$O-E$	$(O-E)^2/E$
(0,1)	8, 5	194	239.58	−45.58	8.67
	8, 7	108	137.09	−29.09	6.17
(1,0)	4, 3	226	253.70	−27.70	3.02
	5, 1	13	7.12	5.88	4.86
	5, 4	19	33.25	−14.25	6.10
	6, 1	15	30.37	−15.37	7.78
	7, 2	62	78.03	−16.03	3.29
	7, 6	41	67.28	−26.28	10.26
	8, 1	28	14.42	13.58	12.78
	8, 5	340	388.56	−48.56	6.07
(1,1)	8, 5	392	355.73	36.27	3.70
(1,1,1)	1, 5, 8	392	353.37	38.63	4.22
	2, 5, 8	351	316.27	34.73	3.81
	3, 5, 8	389	348.32	40.68	4.75
	4, 5, 8	386	347.28	38.72	4.32
	5, 7, 8	276	245.75	30.25	3.72
	6, 7, 8	318	287.55	30.45	3.23

Table 7: Estimated difficulty and discrimination parameters with standard errors in brackets and standardized factor loadings for the two-factor model, women's mobility data

Items	$\hat{\alpha}_{i0}$	s.e.	$\hat{\alpha}_{i1}$	s.e.	$\hat{\alpha}_{i2}$	s.e.	st$\hat{\alpha}_{i1}$	st$\hat{\alpha}_{i2}$	$\hat{\pi}_i(\mathbf{0})$
1	2.66	(0.18)	2.46	(0.28)	0.98	(0.17)	0.87	0.34	0.94
2	−1.58	(0.09)	2.48	(0.21)	1.32	(0.15)	0.83	0.44	0.17
3	1.56	(0.05)	1.25	(0.08)	0.86	(0.10)	0.69	0.47	0.83
4	−1.17	(0.06)	1.97	(0.16)	2.26	(0.17)	0.62	0.72	0.24
5	−6.58	(0.30)	1.98	(0.23)	3.57	(0.22)	0.47	0.85	0.00
6	−5.11	(0.27)	1.32	(0.23)	3.60	(0.24)	0.33	0.91	0.01
7	−17.24	(94.82)	2.20	(0.43)	10.01	(58.02)	0.21	0.97	0.00
8	−4.94	(0.17)	1.51	(0.17)	2.80	(0.15)	0.45	0.84	0.01

Polytomous items, nominal - Multinomial logistic regression

$$x_{i(s)} = \begin{cases} 1, & \text{if the response falls in category } s, \\ & s = 1, \ldots, c_i \\ 0, & \text{otherwise} \end{cases}$$

where c_i denotes the number of categories of variable i

$$g(x_{i(s)} \mid \mathbf{y}) = \prod_{s=1}^{c_i} (\pi_{i(s)}(\mathbf{y}))^{x_{i(s)}}$$

$$\pi_{i(s)}(\mathbf{y}) = P(x_{i(s)} = 1 \mid \mathbf{y})$$

$$\text{logit}\pi_{i(s)}(\mathbf{y}) = \alpha_{i0(s)} + \sum_{j=1}^{q} \alpha_{ij(s)} y_j$$

Ordinal observed variables - Proportional odds model

To take into account the ordinality property of the items we model the cumulative probabilities, $\gamma_{i,s}(\mathbf{y}) = P(x_i \leq s \mid \mathbf{y})$.

The response category probabilities are denoted by

$$\pi_{i,s}(\mathbf{y}) = \gamma_{i,s}(\mathbf{y}) - \gamma_{i,s-1}(\mathbf{y}), \quad s = 1, \ldots, m_i$$

m_i the number of categories for the ith item.

The model used is the proportional odds model:

$$\ln\left[\frac{\gamma_{i,s}(\mathbf{y})}{1-\gamma_{i,s}(\mathbf{y})}\right] = \alpha_{is} - \sum_{j=1}^{k} \beta_{ij} y_j$$

$$\gamma_{i,s}(\mathbf{y}) = P(x_i \leq s) = \pi_{i1}(\mathbf{y}) + \pi_{i2}(\mathbf{y}) + \cdots + \pi_{is}(\mathbf{y})$$

The α_{is}: threshold parameters.

$$\alpha_{i1} < \alpha_{i2} \cdots < \alpha_{im_{i-1}} < \alpha_{im_i} = \infty$$

The β_{ij}: factor loadings.

Under the assumption of conditional independence

$$g(\mathbf{x} \mid \mathbf{y}) = \prod_{i=1}^{p} g(x_i \mid \mathbf{y})$$

The conditional distribution of $x_i \mid \mathbf{y}$ is multinomial:

$$\begin{aligned} g(x_i \mid \mathbf{y}) &= \prod_{s=1}^{m_i} \pi_{is}(\mathbf{y})^{x_{i,s}} \\ &= \prod_{s=1}^{m_i} (\gamma_{i,s} - \gamma_{i,s-1})^{x_{i,s}} \end{aligned}$$

where $x_{i,s}$ takes the value 1 or 0.

The latent variables are assumed to have independent standard normal distributions.

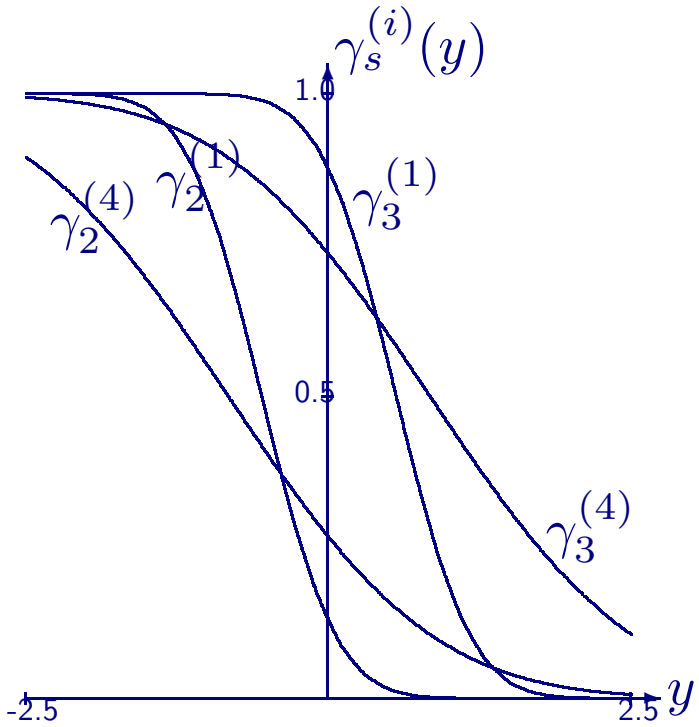

Figure 4: Probit: Four Cumulative Response Functions $\gamma_s^{(i)}(y)$

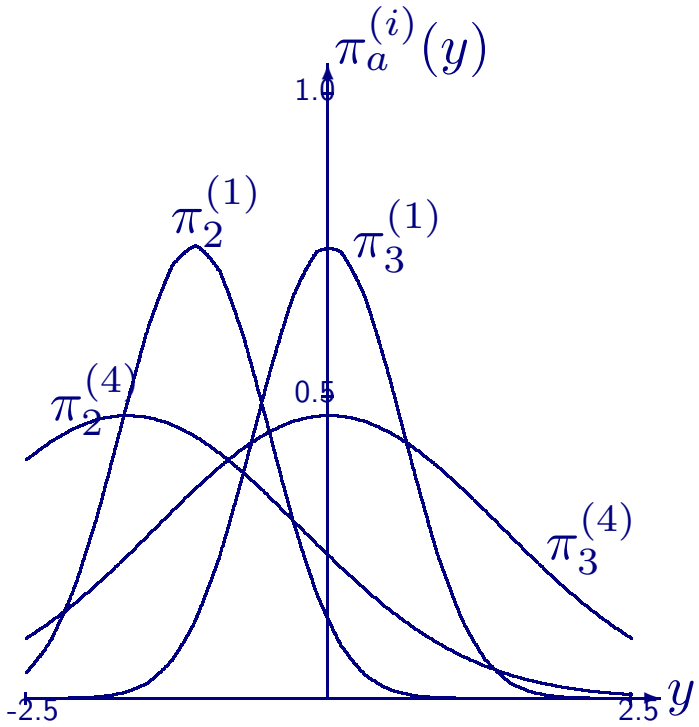

Figure 5: Probit: Four Category Response Functions $\pi_a^{(i)}(y)$

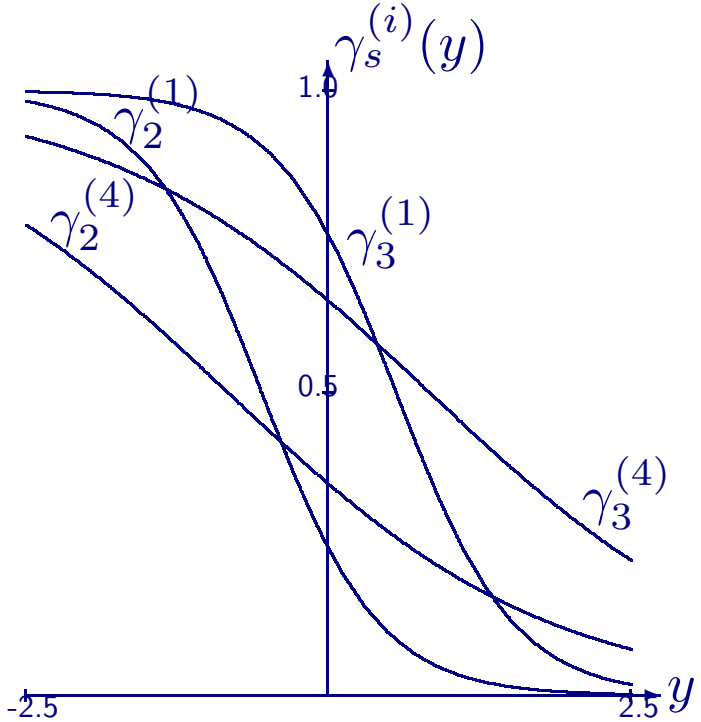

Figure 6: Logit: Four Cumulative Response Functions $\gamma_s^{(i)}$

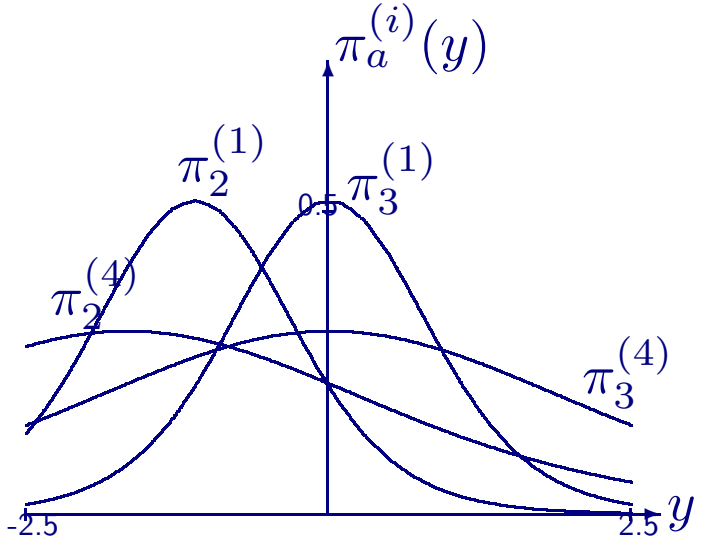

Figure 7: Logit: Four Category Response Functions $\pi_a^{(i)}$

Scoring methods

A. Component scores for different type of items:

Items	$c_j(\mathbf{x})$, $j = 1, \ldots, q$
Binary	$\sum_i^p \alpha_{ij} x_i$
Polytomous	$\sum_i^p \alpha_{ij(s)} x_{i(s)}$
Normal	$\sum_i^p \frac{\lambda_{ij}}{\Psi_{ii}} x_i$
Ordinal	It does not exist

B. Posterior mean

$$E(y_j \mid \mathbf{x}_h), \quad j = 1, \ldots, q$$

For the one factor model, both scoring methods give the same ranking to the individuals.

Latent class model for binary items: examples

1. *Educational assessment.*

2. *Medical diagnosis.* Many symptoms can be easily observed, some of which may point towards one cause and some to another. It would be useful if we could use observations of an individual's symptoms to estimate the probability that the patient has any of the possible conditions. A latent class model may help us to do this.

3. *Selection methods.* Aptitude for performing a complex task, like flying an aircraft, can only be inferred in advance by testing the candidate's performance on a variety of tests designed to give an indication of the required skills.

Objectives of Latent Class Analysis

i) To reduce the complexity of a data set by explaining the associations between the observed variables in terms of membership of a small number of unobservable latent classes, and hence to gain understanding of the interrelationships between the observed variables

ii) To be able to allocate an object to one of these classes.

Assumptions and other characteristics

- Conditional Independence: conditional on an object belonging to a given class, the observable variables are independent.

- The difference between latent class models and the factor analysis: FA assumes that the latent variables are metrical, and possibly normally distributed, whereas in LCA the $single$ latent variable is categorical.

- In a model with J latent classes, the latent variable, y, can be defined to take the value 1 for an object in class 1, 2 for an object in class 2, . . . , and J for an object in class J. The precise labelling is irrelevant.

Notation

Let $\pi_{ij} = \Pr(x_i = 1 \mid j)$ be the probability that a randomly selected object from class j will answer positively to item i, for $(i = 1, \ldots, p;\ j = 1, \ldots, J)$. Thus, π_{ij} is the conditional probability of a positive response to item i, given (or conditional on) membership of class j.

Let η_j be the proportion of the population in latent class j or equivalently the probability that a randomly selected object from the population belongs to latent class j, for $(j = 1, \ldots, J)$.

Model estimation

The joint distribution of the observed responses is written as:

$$f(\mathbf{x}) = \sum_{j=0}^{J-1} \eta_j g(\mathbf{x} \mid j)$$

where under the assumption of conditional independence:

$$g(\mathbf{x} \mid j) = \prod_{i=1}^{p} g(x_i \mid j)$$

Since the responses are binary:

$$g(x_i \mid j) = \pi_{ij}^{x_i}(1 - \pi_{ij})^{1-x_i}$$

The log-likelihood for a random sample of size n is:

$$L = \log f(\mathbf{x}) = \sum_{h=1}^{n} \log \sum_{j=1}^{J} \eta_j \prod_{i=1}^{p} g(x_{ih} \mid j)$$

The log-likelihood function can be maximized using standard optimization routines.

The above log-likelihood can be maximized using an EM algorithm under the constraint that: $\sum_{j=0}^{k-1} \eta_j = 1$. Therefore the function to be maximized becomes:

$$\phi = L + \theta \sum_{j=0}^{k-1} \eta_j$$

where θ is an undetermined multiplier.

Finding partial derivatives:

$$\begin{aligned} \frac{\partial \phi}{\partial \eta_j} &= \sum_{h=1}^{n} \left[\prod_{i=1}^{p} \pi_{ij}^{x_{ih}} (1 - \pi_{ij})^{1 - x_{ih}} / f(\mathbf{x}_h) \right] + \theta \\ &= \sum_{h=1}^{n} \left[g(\mathbf{x} \mid j) / f(\mathbf{x}_h) \right] + \theta, \quad j = 0, \ldots, J-1 \end{aligned}$$

$$\frac{\partial \phi}{\partial \pi_{ij}} = \sum_{h=1}^{n} \eta_j \frac{\partial}{\partial \pi_{ij}} g(\mathbf{x}_h \mid j)/f(\mathbf{x}_h), \quad i = 1, \ldots, p; j = 0, \ldots, J-1$$

Now,

$$\begin{aligned}
\frac{\partial g(\mathbf{x}_h \mid j)}{\partial \pi_{ij}} &= \frac{\partial}{\partial \pi_{ij}} \exp \sum_{i=1}^{p} \left[x_{ih} \ln \pi_{ij} + (1 - x_{ih}) \ln(1 - \pi_{ij}) \right] \\
&= g(\mathbf{x}_h \mid j) \left[\frac{x_{ih}}{\pi_{ij}} - \frac{1 - x_{ih}}{1 - \pi_{ij}} \right] \\
&= \frac{(x_{ih} - \pi_{ij})}{\pi_{ij}(1 - \pi_{ij})} g(\mathbf{x}_h \mid j)
\end{aligned}$$

Therefore,

$$\frac{\partial \phi}{\partial \pi_{ij}} = \frac{\eta_j}{\pi_{ij}(1-\pi_{ij})} \sum_{h=1}^{n} (x_{ih} - \pi_{ij}) g(\mathbf{x}_h \mid j) / f(\mathbf{x}_h)$$

The derivatives can be simplified by expressing them in terms of the posterior probabilities $phi(j \mid \mathbf{x})$:

$$\phi(j \mid \mathbf{x}) = \eta_j g(\mathbf{x}_h \mid j) / f(\mathbf{x}_h)$$

Substituting that into the partial derivatives equations and setting them equal to zero we get:

$$\sum_{h=1}^{n} \phi(j \mid \mathbf{x}_h) = -\theta \eta_j$$

Summing both sides over j and using $\sum_j \eta_j = 1$ we get that

$$\theta = -n$$

and hence the first and second estimating equations are:

$$\hat{\eta}_j = \sum_{h=1}^{n} \phi(j \mid \mathbf{x}_h)/n \tag{5}$$

$$\hat{\pi}_{ij} = \sum_{h=1}^{n} x_{ih}\phi(j \mid \mathbf{x}_h)/(n\hat{\eta}_j) \tag{6}$$

where, the posterior probability than an individual with response pattern $\mathbf{x}_h$ will be in class j, is given by:

$$\phi(j \mid \mathbf{x}_h) = \eta_j g(\mathbf{x}_h \mid j)/f(\mathbf{x}_h) \tag{7}$$

The EM algorithm works as follow:

i. Choose initial values for the posterior probabilities $\phi(j \mid \mathbf{x}_h)$.

ii. Obtain a first approximation for $\hat{\eta}_j$, $\hat{\pi}_{ij}$ from the equations (5), (6).

iii. Substitute these in (7) to obtain a new estimate for $\phi(j \mid \mathbf{x}_h)$.

iv. Return to (ii) and continue until convergence is attained.

The initial allocation of individuals into classes is based on their total score.

General remarks

- The solution reached will be a local maximum.
- Latent class models are known for multiple maxima.
- Use different starting values.

i) The n objects are a random sample from some population and every object in that population belongs to just one of the J latent classes

ii) The probability of giving a positive response to a particular item is the same for all objects in the same class but may be different for objects in different classes

Allocation to classes

We solve the problem by estimating the probability that an object with a particular response pattern falls into a particular class. This probability, sometimes called the $posterior$ probability, is:

$$\Pr(\text{object is in class } j \mid x_1, \ldots, x_p) \quad (j = 1, \ldots, J).$$

Example: Macready and Dayton data

Sample size = 142

Results from a test on four items selected at random from a domain of items each involving the multiplication of a two-digit number by a three- or four-digit number. Respondents are expected to be divided into two groups: Masters and Non-Masters.

Table 8: Observed and predicted frequencies and estimated class probabilities for the two-class model, Macready and Dayton data

Observed frequency	Expected frequency	$\hat{Pr}(\text{master} \mid \mathbf{x})$	Class	Response pattern
15	14.96	1.00	2	1111
23	19.72	1.00	2	1101
7	6.19	1.00	2	1110
4	4.90	1.00	2	0111
1	4.22	1.00	2	1011
7	8.92	0.91	2	1100
6	6.13	0.90	2	1001
5	6.61	0.98	2	0101
3	1.93	0.90	2	1010
2	2.08	0.97	2	0110
4	1.42	0.97	2	0011
13	12.91	0.18	1	1000
6	5.62	0.47	1	0100
4	4.04	0.45	1	0001
1	1.31	0.44	1	0010
41	41.04	0.02	1	0000

The $X^2 = 9.5$ and the $G^2 = 9.0$ on six degrees of freedom indicate a near perfect fit to the data. The percentage of G^2 explained is 91%.

Table 9: Estimated conditional probabilities, $\hat{\pi}_{ij}$, and prior probabilities, $\hat{\eta}_j$, with standard errors in brackets for the two-class model, Macready and Dayton data

Item (i)	$\hat{\pi}_{i1}$	$\hat{\pi}_{i2}$
1	0.21 (0.06)	0.75 (0.06)
2	0.07 (0.06)	0.78 (0.06)
3	0.02 (0.03)	0.43 (0.06)
4	0.05 (0.05)	0.71 (0.06)
$\hat{\eta}_j$	0.41 (0.06)	0.59 (0.06)

Members of the first class have small estimated probabilities of answering items correctly. This class is clearly the "non-master" one. Members in the second class have for all items much higher probabilities of answering correctly. This class is the "master" class.

Example: Abortion data

1. The woman decides on her own that she does not. [WomanDecide]
2. The couple agree that they do not wish to have the child. [CoupleDecide]
3. The woman is not married and does not wish to marry the man. [NotMarried]
4. The couple cannot afford any more children. [CannotAfford]

Item (i)	$\hat{\pi}_{i1} = \hat{\Pr}(x_i = 1 \mid 1)$	$\hat{\pi}_{i2} = \hat{\Pr}(x_i = 1 \mid 2)$
WomanDecide	0.01 (0.01)	0.71 (0.03)
CoupleDecide	0.09 (0.03)	0.91 (0.02)
NotMarried	0.12 (0.04)	0.96 (0.02)
CannotAfford	0.15 (0.04)	0.91 (0.02)
$\hat{\eta}_j$	0.39 (0.03)	0.61 (0.03)

	Items	O	E	$O-E$	$(O-E)^2/E$
Response (0,0)	2, 1	147	137.79	9.21	0.62
	3, 1	131	130.16	0.84	0.05
	3, 2	117	117.58	−0.58	0.00
	4, 1	129	129.12	−0.12	0.00
	4, 2	114	114.61	−0.61	0.00
	4, 3	116	109.97	6.03	0.33
Response (0,1)	1, 2	66	75.21	−9.21	1.13
	1, 3	82	82.84	−0.84	0.01
	1, 4	84	83.88	0.12	0.00
	2, 1	7	16.21	−9.21	5.24
	2, 3	37	36.42	0.58	0.01
	2, 4	40	39.39	0.61	0.01
	3, 1	7	7.84	−0.84	0.09
	3, 2	21	20.42	0.58	0.02
	3, 4	22	28.03	−6.03	1.30
	4, 1	16	15.88	0.12	0.00
	4, 2	31	30.39	0.61	0.01
	4, 3	29	35.03	−6.03	1.04
Response (1,1)	2, 1	159	149.79	9.21	0.57
	3, 1	159	158.16	0.84	0.01
	3, 2	204	204.58	−0.58	0.00
	4, 1	150	150.12	−0.12	0.00
	4, 2	194	194.61	−0.61	0.00
	4, 3	212	205.97	6.03	0.18

Workplace Industrial Relations Survey, WIRS

The six items measure the amount of consultation that takes place in firms at different levels of the firm structure.

1. Informal discussion with individual workers.
2. Meetings with groups of workers.
3. Discussions in established joint consultative committee.
4. Discussions in specially constituted committee to consider the change.
5. Discussions with union representatives at the establishment.
6. Discussions with paid union officials from outside.

Items 1 to 6 cover a range of informal to formal types of consultation. The first two items are less formal practices, and items 3 to 6 are more formal.

• The latent class analysis aims to group the firms with respect to the patterns of consultation they are adopting.

• The two-class model fitted to the six items is rejected not only by the overall goodness-of-fit measures ($X^2 = 350.28, G^2 = 299.12$ on 21 degrees of freedom) but also by the large chi-squared residuals for some of the two and three-way margins. All the chi-squared residuals with values greater than 3 include item 1.

• The three-class model is still rejected ($X^2 = 64.89, G^2 = 67.78$ on 14 degrees of freedom).

• However, the fit to the two- and three-way margins is very good.

Items	$\hat{\pi}_{i1}$	$\hat{\pi}_{i2}$	$\hat{\pi}_{i3}$
1	0.21	0.95	0.06
2	0.59	0.27	1.00
3	0.08	0.43	0.68
4	0.14	0.19	0.62
5	0.11	0.53	0.85
6	0.02	0.25	0.37
$\hat{\eta}_j$	0.55	0.26	0.19

Class 1 represents those firms that mainly use informal policies (items 1 and 2).

Class 3 includes those firms that use all the methods but not the first informal one.

Firms in Class 2 use all methods including that under item 1 (with lower probabilities than in Class 3 for items 2 to 6)

Applications in Archaeometry

The metric variables, 25 in total, measure the chemical composition of the ceramic, obtained with the latest methodologies available such as Neutron Activation Analysis (NAA).

The categorical variables aim to derive information regarding the provenance of the objects. Recently, a system of 19 categorical variables has been derived in order to objectively describe the thin sections of the ceramics and use this for reproducible statistical applications. The levels of each of the 19 variables give information about the amount (if any) of different rock types, minerals and structure. More specifically the categorical variables are: optical activity, inclusion orientation, void orientation, texture, special components, plutonic rocks, metamorphic rocks, sedimentary rocks, quartz, feldspar, plagioclase, pyroxenes, amphiboles, volcanic rocks, micas, phyllosillicates, carbonates, packing and other constituents.

Teraccota data set

The 73 sample objects are maiolica vases and floor tiles, manufactured between the XVI-XVIII centuries. The data set consists of 19 (binary) variables and 21 metric variables.

The metric variables measure the chemical composition of the ceramic. The categorical variables aim to derive information regarding the provenance of the objects (petrological analysis).

The groups are:

- Group 1: ceramics from Napoli (n)
- Group 2: ceramics from Caltagirone (c)
- Group 3: ceramics from Palermo (p)

Results

- The AIC and BIC suggested a three-class solution.
- Class I comprises 33% of the objects, class II 37% and class III 30%.
- The estimated posterior class-membership probabilities was for each object greater than 0.99.
- The obtained classification was the same as the grouping with respect to location.
- When the analysis was done separately on the binary and the metric variables both using hierarchical clustering techniques and latent class analysis the groups were not as clearly defined suggesting that the method of analysis using both binary and metric variables is preferable.

Can Sora Data set

The Can Sora data set comes from a ceramic assemblage found in a cistern at the Punic and Roman site of Ses Paises de Cala d'Hort in Eivissa.

Variables: 15 binary variables, 3 ordinal and 25 metric. The natural logarithms of the metric variables were taken first and they were standardized afterwards.

The AIC and BIC suggested a 6-class solution.

Table 10: Residuals for the second order margins, 5-class model, Can Sora

Response	Variable i	Variable j	Observed frequency (O)	Expected frequency (E)	$O-E$	$(O-E)^2/E$
(0,0)	8	6	4	1.62	2.37	3.46
(0,1)	15.1	5	0	2.37	-2.37	2.37
	15.2	5	4	1.62	2.37	3.46
(1,0)	5	15.1	0	2.37	-2.37	2.37
	5	15.2	4	1.62	2.37	3.46
(1,1)	15.1	5	4	1.62	2.37	3.46
	15.2	5	0	2.37	-2.37	2.37
	15.2	15.1	0	2.37	-2.37	2.37

Table 11: Residuals for the third order margins, 5-class model, response (1,1,1) to variables (i, j, k), Can Sora

Variable i	Variable j	Variable k	$O - E$	$(O - E)^2/E$
2	5	15.1	2.37	3.47
2	5	15.2	-2.37	2.37
5	6	18.1	0.75	2.25
5	7	9	2.57	4.67
5	7	10	2.57	4.67
5	7	15.1	3.07	10.27
5	9	15.1	3.07	10.27
5	10	15.1	3.07	10.27
5	13	15.1	2.37	3.47
6	8	15.2	2.79	3.55

Table 12: Classification of the 22 objects, six-class model, Can Sora

Group	Objects
Plutonic	CS2, CS3, CS4, CS5, CS6, CS14, CS23
Volcanic	CS10, CS11, CS15, CS16, CS17
Muscovite	CS18, CS19, CS20
Phyllite	CS21, CS22
Pantellerian	CS26, CS27
Outliers	CS7, CS24, CS25

Free software for latent class analysis

• The program LATCLASS in LAMI: Bartholomew, Knott, Tzamourani and deMenezes available in GENLAT.

• LEM: Vermunt, J.K.

Non-free software

• Mplus: Muthen, L. and Muthen, B.O.

• LatentGold: Vermunt, J.K. and Magison, J.

• WinLTA: Collins, L.M. and Flaherty, B.P. and Hyatt, S.L. and Schafer, J.L.