# Package 'multilevLCA'

February 5, 2025

Type Package	
Version 2.0.0	
<b>Date</b> 2025-02-04	
<b>Title</b> Estimates and Plots Single-Level and Multilevel Latent Class Models	
<b>Description</b> Efficiently estimates single- and multilevel latent class models with covariates, allowing for output visualization in all specifications. For more technical details, see Lyrvall et al (2023) <doi:10.48550 arxiv.2305.07276="">.</doi:10.48550>	
<b>Depends</b> R (>= 3.5.0)	
<b>Imports</b> Rcpp (>= 1.0.9), magrittr, MASS, dplyr, tidyr, klaR, foreach, parallel, clustMixType, pracma, tictoc	
LinkingTo Rcpp, RcppArmadillo	
License GPL (>= 2)	
LazyData true	
NeedsCompilation yes	
Author Roberto Di Mari [aut, cre], Johan Lyrvall [aut], Zsuzsa Bakk [ctb], Jennifer Oser [ctb], Jouni Kuha [ctb]	
Maintainer Roberto Di Mari <roberto.dimari@unict.it></roberto.dimari@unict.it>	
Repository CRAN	
<b>Date/Publication</b> 2025-02-04 23:50:02 UTC	
Contents	
multilevLCA-package . dataIEA . dataTOY . multiLCA . plot.multiLCA .	2 3 4 5
Index	16

multilevLCA-package

Estimates and Plots Single-Level and Multilevel Latent Class Models

## **Description**

Efficiently estimates single- and multilevel latent class models with covariates, allowing for output visualization in all specifications. For more technical details, see Lyrvall et al (2023) <doi:10.48550/arXiv.2305.07276>.

## **Details**

For estimating latent class models, see multiLCA.

For plotting latent class models, see plot.multiLCA

#### Author(s)

Roberto Di Mari and Johan Lyrvall.

Maintainer: Roberto Di Mari <roberto.dimari@unict.it>

#### References

Bakk, Z., & Kuha, J. (2018). Two-step estimation of models between latent classes and external variables. *Psychometrika*, 83, 871-892.

Bakk, Z., Di Mari, R., Oser, J., & Kuha, J. (2022). Two-stage multilevel latent class analysis with covariates in the presence of direct effects. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(2), 267-277.

Di Mari, Bakk, Z., R., Oser, J., & Kuha, J. (2023). A two-step estimator for multilevel latent class analysis with covariates. Psychometrika.

Lukociene, O., Varriale, R., & Vermunt, J. K. (2010). The simultaneous decision(s) about the number of lower-and higher-level classes in multilevel latent class analysis. Sociological Methodology, 40(1), 247-283.

## **Examples**

```
data = dataIEA
Y = colnames(dataIEA)[4+1:12]

out = multiLCA(data = data, Y = Y, iT = 2)
out
plot(out, horiz = FALSE)
```

dataIEA 3

dataIEA

Data for understanding of good citizenship behaviour

## **Description**

Data set from the International Civic and Citizenship Education Study 2016 (Schulz et al., 2018). As part of a comprehensive evaluation of education systems, the IEA conducted surveys in 1999, 2009 and 2016 in school classes of 14-year olds to investigate civic education with the same scientific rigor as the evaluation of more traditional educational skills of language and mathematics. The present study focuses on the third wave of the survey that was conducted in 2016.

Questions regarding citizenship norms in all three waves asked respondents to explain their understanding of what a good adult citizen is or does. The survey then lists a variety of activities for respondents to rate in terms of how important these activities are in order to be considered a good adult citizen. The twelve items range from obeying the law and voting in elections, to protecting the environment and defending human rights.

Covariates included are customary determinants of citizenship norms from the literature at the individual-level of socio-economic measures and country-level measure of gross domestic product (GDP) per capita.

## Usage

data("dataIEA")

#### **Format**

A data frame with 90221 observations on the following 28 variables.

ICCS\_year Year of survey

**COUNTRY Country** 

IDSTUD Study ID

TOTWGTS Study weight

obey Always obeying the law

rights Taking part in activities promoting human rights

local Participating in activities to benefit people in the local community

work Working hard

envir Taking part in activities to protect the environment

vote Voting in every national election

history Learning about the country's history

respect Showing respect for government representatives

news Following political issues in the newspaper, on the radio, on TV, or on the Internet

protest Participating in peaceful protests against laws believed to be unjust

discuss Engaging in political discussions

4 dataTOY

```
party Joining a political party
female Female
books Number of books at home
edexp Educational expectations
ed_mom Mother education
ed_dad Father education
nonnat_born Non-native born
immigrantfam Immigrant family
nonnat_lang Non-native language level
gdp_constant GDP
log_gdp_constant Log GDP
gdp_currentusd GDP in USD
log_gdp_currentusd Log GDP in USD
```

#### References

Schulz, W., Ainley, J., Fraillon, J., Losito, B., Agrusti, G., & Friedman, T. (2018). Becoming citizens in a changing world: IEA International Civic and Citizenship Education Study 2016 international report. Springer.

dataT0Y

Artificial data set

## **Description**

Artificial multilevel data set.

## Usage

data("dataTOY")

## **Format**

A data frame with 3000 observations on the following 13 variables.

id\_high High-level id

Y\_1 Indicator n.1

Y\_2 Indicator n.2

Y\_3 Indicator n.3

Y\_4 Indicator n.4

Y\_5 Indicator n.5

Y\_6 Indicator n.6

```
Y_7 Indicator n.7
Y_8 Indicator n.8
Y_9 Indicator n.9
Y_10 Indicator n.10
Z_1ow Continuous low-level covariate
Z_high Continuous high-level covariate
```

#### References

Di Mari, Bakk, Z., R., Oser, J., & Kuha, J. (2023). A two-step estimator for multilevel latent class analysis with covariates. Under review. Available from https://arxiv.org/abs/2303.06091.

multiLCA

Estimates and plots single- and multilevel latent class models

## **Description**

The multilCA function in the multilevLCA package estimates single- and multilevel measurement and structural latent class models. Moreover, the function performs two different strategies for model selection. Methodological details can be found in Bakk et al. (2022), Bakk and Kuha (2018), and Di Mari et al. (2023).

Different output visualization tools are available for all model specifications. See, e.g., plot.multiLCA.

## Usage

```
multiLCA(
data,
Υ,
iT,
id_high = NULL,
iM = NULL,
Z = NULL,
Zh = NULL,
incomplete = FALSE,
fixedslopes = FALSE,
startval = NULL,
kmea = TRUE,
extout = FALSE,
dataout = TRUE,
sequential = TRUE,
numFreeCores = 2,
maxIter = 1e3,
tol = 1e-8,
reord = TRUE,
fixedpars = 1,
```

```
NRmaxit = 100,
NRtol = 1e-6,
verbose = TRUE
)
```

## Arguments

data	Input matrix or dataframe.
Υ	Names of data columns with indicators.
iT	Number of lower-level latent classes.
id_high	Name of data column with higher-level id. Default: NULL.
iM	Number of higher-level latent classes. Default: NULL.
Z	Names of data columns with lower-level covariates (non-numeric covariates are treated as nominal). Default: NULL.
Zh	Names of data columns with higher-level covariates (non-numeric covariates are treated as nominal). Default: NULL.
incomplete	Whether to estimate the model with missing values included by means of full-information maximum-likelihood estimation (TRUE) or perform row-wise deletion of missing values (FALSE). Default: FALSE.
fixedslopes	Whether to estimate multilevel models with covariates with fixed lower-level slope parameters across the higher-level classes by means of log-linear parametrization. Default: FALSE.
startval	Name of data column with starting values for lower-level latent classes. Default: NULL.
kmea	Whether to compute starting values for single-level model using $K$ -means (TRUE), which is recommended for algorithmic stability, or $K$ -modes (FALSE). Default: TRUE.
extout	Whether to output extensive model and estimation information. Default: FALSE.
dataout	Whether to match class predictions to the observed data. Default: TRUE.
sequential	Whether to perform sequential model selection (TRUE) or parallelized model selection (FALSE). Default: TRUE.
numFreeCores	If performing parallelized model selection, the number of CPU cores to keep free. Default: 2.
maxIter	Maximum number of iterations for EM algorithm. Default: 1e3.
tol	Tolerance for EM algorithm. Default: 1e-8.
reord	Whether to (re)order classes in decreasing order according to probability of scoring yes on all items. Default: TRUE.
fixedpars	One-step estimator $(0)$ , two-step estimator $(1)$ or two-stage estimator $(2)$ . Default: 1.
NRmaxit	Maximum number of iterations for Newton-Raphson algorithm. Default: 100.
NRtol	Tolerance for Newton-Raphson algorithm. Default: 1e-6.
verbose	Whether to print estimation progress. Default: TRUE.

#### **Details**

The indicator columns may be coded as as consecutive sequence of integers from 0, or as characters.

To directly estimate a latent class model, iT and (optionally) iM should be specified as a single positive integer. To perform model selection over range of consecutive positive integers as the number of latent classes, iT and/or iM may be specified in the form iT\_min:iT\_max and/or iM\_min:iM\_max. It is possible to specify iT = iT\_min:iT\_max with either iM = NULL or iM equal to a single positive integer, iM = iM\_min:iM\_max with iT equal to a single positive integer, or iT = iT\_min:iT\_max with iM = iM\_min:iM\_max. All model selection procedures return the output of the optimal model based on the BIC.

In the case where both iT and iM are defined as a range of consecutive positive integers, model selection can be performed using the sequential three-stage approach (Lukociene et al., 2010) or a simultaneous approach. The sequential approach involves (first step) estimating iT\_min:iT\_max single-level models and identifying the optimal alternative iT\_opt1 based on the BIC, (second step) estimating iM\_min:iM\_max|iT = iT\_opt1 multilevel models and identifying the optimal alternative iM\_opt2 based on the higher-level BIC, and (third step) estimating iT\_min:iT\_max|iM = iM\_opt2 multilevel models and identifying the optimal alternative iT\_opt3 based on the lower-level BIC. The simultaneous approach involves devoting multiple CPU cores on the local machine to estimate all combinations in iT = iT\_min:iT\_max, iM = iM\_min:iM\_max and identifying the optimal alternative based on the lower-level BIC.

#### Value

Single-level model estimation returns (if extout = FALSE, a subset):

vPi Class proportions

mPhi Response probabilities given the latent classes

mU Matrix of posterior class assignment (proportional assignment)

mU\_modal Matrix of posterior class assignment (modal assignment)
vU\_modal Vector of posterior class assignment (modal assignment)

mClassErr Expected number of classification errors
mClassErrProb Expected proportion of classification errors

AvgClassErrProb

Average of mClassErrProb

R2entr Entropy-based R<sup>2</sup>

BIC Bayesian Information Criterion (BIC)

AIC Akaike Information Criterion (AIC)

vGamma Intercepts in logistic parametrization for class proportions

mBeta Intercepts in logistic parametrization for response probabilities

parvec Vector of logistic parameters

SEs Standard errors

Variance-covariance matrix

iter Number of iterations for EM algorithm

eps Difference between last two elements of log-likelihood sequence for EM algo-

rithm

LLKSeries Full log-likelihood series for EM algorithm

mScore Contributions to log-likelihood score

spec Model specification

missing\_values Strategy for handling of eventual missing values

sample\_size Final sample size for model estimation

Single-level model estimation with covariates returns (if extout = FALSE, a subset):

mPi Class proportions given the covariates

vPi\_avg Sample average of mPi

mPhi Response probabilities given the latent classes

mU Matrix of posterior class assignment (proportional assignment)

mClassErr Expected number of classification errors
mClassErrProb Expected proportion of classification errors

AvgClassErrProb

Average of mClassErrProb

R2entr Entropy-based R<sup>2</sup>

BIC Bayesian Information Criterion (BIC)
AIC Akaike Information Criterion (AIC)

mGamma Intercept and slope parameters in logistic models for conditional class member-

ship

mBeta Intercepts in logistic parametrization for response probabilities

parvec Vector of logistic parameters
SEs\_unc Uncorrected standard errors

SEs\_cor Corrected standard errors (see Bakk & Kuha, 2018; Di Mari et al., 2023)

SEs\_cor\_gamma Corrected standard errors only for the gammas (see Bakk & Kuha, 2018; Di

Mari et al., 2023)

mQ Cross-derivatives for asymptotic standard error correction in two-step estimation

(see Bakk & Kuha, 2018; Di Mari et al., 2023)

Varmat\_unc Uncorrected variance-covariance matrix

Varmat\_cor Corrected variance-covariance matrix (see Bakk & Kuha, 2018; Di Mari et al.,

2023)

mV2 Inverse of information matrix for structural model

iter Number of iterations for EM algorithm

eps Difference between last two elements of log-likelihood sequence for EM algo-

rithm

LLKSeries Full log-likelihood series for EM algorithm

spec Model specification

estimator Estimation approach for structural model

missing\_values Strategy for handling of eventual missing values

sample\_size Final sample size for model estimation

Multilevel model estimation returns (if extout = FALSE, a subset):

vOmega Higher-level class proportions

mPi Lower-level class proportions given the higher-level latent classes

mPhi Response probabilities given the lower-level latent classes cPMX Posterior joint class assignment (proportional assignment)

cLogPMX Log of cPMX

cPX Posterior lower-level class assignment given high-level class membership (pro-

portional assignment)

cLogPX Log of cPX

mSumPX Posterior higher-level class assignment for lower-level units after marginaliza-

tion over the lower-level classes (proportional assignment)

mPW Posterior higher-level class assignment for higher-level units (proportional as-

signment)

mlogPW Log of mPW

mPW\_N Posterior higher-level class assignment for lower-level units (proportional as-

signment)

mPMsumX Posterior lower-level class assignment for lower-level units after marginalization

over the higher-level classes (proportional assignment)

R2entr\_low Lower-level entropy-based R<sup>2</sup>
R2entr\_high Higher-level entropy-based R<sup>2</sup>

BIClow Lower-level Bayesian Information Criterion (BIC)

BIChigh Higher-level Bayesian Information Criterion (BIC)

ICL\_BIClow Lower-level BIC-type approximation the integrated complete likelihood ICL\_BIChigh Higher-level BIC-type approximation the integrated complete likelihood

AIC Akaike Information Criterion (AIC)

vAlpha Intercepts in logistic parametrization for higher-level class proportions

mGamma Intercepts in logistic parametrization for conditional lower-level class propor-

tions

mBeta Intercepts in logistic parametrization for response probabilities

parvec Vector of logistic parameters

SEs Standard errors

Variance-covariance matrix
Infomat Expected information matrix

iter Number of iterations for EM algorithm

eps Difference between last two elements of log-likelihood sequence for EM algo-

rithm

LLKSeries Full log-likelihood series for EM algorithm vLLK Current log-likelihood for higher-level units

mScore Contributions to log-likelihood score

spec Model specification

missing\_values Strategy for handling of eventual missing values

sample\_size Final sample size for model estimation

Multilevel model estimation with lower-level covariates returns (if extout = FALSE, a subset):

vOmega Higher-level class proportions

mPi Lower-level class proportions given the higher-level latent classes and the co-

variates

mPi\_avg Sample average of mPi

mPhi Response probabilities given the lower-level latent classes cPMX Posterior joint class assignment (proportional assignment)

cLogPMX Log of cPMX

cPX Posterior lower-level class assignment given high-level class membership (pro-

portional assignment)

cLogPX Log of cPX

mSumPX Posterior higher-level class assignment for lower-level units after marginaliza-

tion over the lower-level classes (proportional assignment)

mPW Posterior higher-level class assignment for higher-level units (proportional as-

signment)

mlogPW Log of mPW

mPW\_N Posterior higher-level class assignment for lower-level units (proportional as-

signment)

mPMsumX Posterior lower-level class assignment for lower-level units after marginalization

over the higher-level classes (proportional assignment)

R2entr\_low Lower-level entropy-based R<sup>2</sup>
R2entr\_high Higher-level entropy-based R<sup>2</sup>

BIClow Lower-level Bayesian Information Criterion (BIC)

BIChigh Higher-level Bayesian Information Criterion (BIC)

ICL\_BIClow Lower-level BIC-type approximation the integrated complete likelihood ICL\_BIChigh Higher-level BIC-type approximation the integrated complete likelihood

AIC Akaike Information Criterion (AIC)

vAlpha Intercepts in logistic parametrization for higher-level class proportions

cGamma Intercept and slope parameters in logistic models for conditional lower-level

class membership

mBeta Intercepts in logistic parametrization for response probabilities

parvec Vector of logistic parameters

SEs\_unc Uncorrected standard errors

SEs\_cor Corrected standard errors (see Bakk & Kuha, 2018; Di Mari et al., 2023)

SEs\_cor\_gamma Corrected standard errors only for the gammas (see Bakk & Kuha, 2018; Di

Mari et al., 2023)

mQ Cross-derivatives for asymptotic standard error correction in two-step estimation

(see Bakk & Kuha, 2018; Di Mari et al., 2023)

Varmat\_unc Uncorrected variance-covariance matrix

Varmat\_cor Corrected variance-covariance matrix (see Bakk & Kuha, 2018; Di Mari et al.,

2023)

Infomat Expected information matrix

cGamma\_Info Expected information matrix only for the gammas mV2 Inverse of information matrix for structural model

iter Number of iterations for EM algorithm

eps Difference between last two elements of log-likelihood sequence for EM algo-

rithm

LLKSeries Full log-likelihood series for EM algorithm vLLK Current log-likelihood for higher-level units

mScore Contributions to log-likelihood score

mGamma\_Score Contributions to log-likelihood score only for the gammas

spec Model specification

estimator Estimation approach for structural model

missing\_values Strategy for handling of eventual missing values

sample\_size Final sample size for model estimation

Multilevel model estimation with lower- and higher-level covariates returns (if extout = FALSE, a subset):

mOmega Higher-level class proportions given the covariates

vOmega\_avg Higher-level class proportions averaged over higher-level units

mPi Lower-level class proportions given the higher-level latent classes and the co-

variates

mPi\_avg Sample average of mPi

mPhi Response probabilities given the lower-level latent classes cPMX Posterior joint class assignment (proportional assignment)

cLogPMX Log of cPMX

cPX Posterior lower-level class assignment given high-level class membership (pro-

portional assignment)

cLogPX Log of cPX

mSumPX Posterior higher-level class assignment for lower-level units after marginaliza-

tion over the lower-level classes (proportional assignment)

mPW Posterior higher-level class assignment for higher-level units (proportional as-

signment)

mlogPW Log of mPW

mPW\_N Posterior higher-level class assignment for lower-level units (proportional as-

signment)

mPMsumX Posterior lower-level class assignment for lower-level units after marginalization

over the higher-level classes (proportional assignment)

R2entr\_low Lower-level entropy-based R<sup>2</sup>
R2entr\_high Higher-level entropy-based R<sup>2</sup>

BIClow Lower-level Bayesian Information Criterion (BIC)

BIChigh Higher-level Bayesian Information Criterion (BIC)

ICL\_BIClow Lower-level BIC-type approximation the integrated complete likelihood ICL\_BIChigh Higher-level BIC-type approximation the integrated complete likelihood

AIC Akaike Information Criterion (AIC)

mAlpha Intercept and slope parameters in logistic models for conditional higher-level

class membership

cGamma Intercept and slope parameters in logistic models for conditional lower-level

class membership

mBeta Intercepts in logistic parametrization for response probabilities

parvec Vector of logistic parameters
SEs\_unc Uncorrected standard errors

SEs\_cor Corrected standard errors (see Bakk & Kuha, 2018; Di Mari et al., 2023)

SEs\_cor\_alpha Corrected standard errors only for the alphas (see Bakk & Kuha, 2018; Di Mari

et al., 2023)

SEs\_cor\_gamma Corrected standard errors only for the gammas (see Bakk & Kuha, 2018; Di

Mari et al., 2023)

mQ Cross-derivatives for asymptotic standard error correction in two-step estimation

(see Bakk & Kuha, 2018; Di Mari et al., 2023)

Varmat\_unc Uncorrected variance-covariance matrix

Varmat\_cor Corrected variance-covariance matrix (see Bakk & Kuha, 2018; Di Mari et al.,

2023)

Infomat Expected information matrix

cAlpha\_Info Expected information matrix only for the alphas cGamma\_Info Expected information matrix only for the gammas mV2 Inverse of information matrix for structural model

iter Number of iterations for EM algorithm

eps Difference between last two elements of log-likelihood sequence for EM algo-

rithm

LLKSeries Full log-likelihood series for EM algorithm
vLLK Current log-likelihood for higher-level units

mScore Contributions to log-likelihood score

mAlpha\_Score Contributions to log-likelihood score only for the alphas mGamma\_Score Contributions to log-likelihood score only for the gammas

spec Model specification

estimator Estimation approach for structural model

missing\_values Strategy for handling of eventual missing values

sample\_size Final sample size for model estimation

#### References

Bakk, Z., & Kuha, J. (2018). Two-step estimation of models between latent classes and external variables. *Psychometrika*, 83, 871-892.

Bakk, Z., Di Mari, R., Oser, J., & Kuha, J. (2022). Two-stage multilevel latent class analysis with covariates in the presence of direct effects. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(2), 267-277.

Di Mari, Bakk, Z., R., Oser, J., & Kuha, J. (2023). A two-step estimator for multilevel latent class analysis with covariates. Psychometrika.

Lukociene, O., Varriale, R., & Vermunt, J. K. (2010). The simultaneous decision(s) about the number of lower-and higher-level classes in multilevel latent class analysis. Sociological Methodology, 40(1), 247-283.

## **Examples**

```
# Use the artificial data set
data = dataTOY
# Define vector with names of columns with items
Y = colnames(data)[1+1:10]
# Define name of column with higher-level id
id_high = "id_high"
# Define vector with names of columns with lower-level covariates
Z = c("Z_low")
# Define vector with names of columns with higher-level covariates
Zh = c("Z_high")
# Single-level 3-class LC model with covariates
out = multiLCA(data, Y, 3, Z = Z, verbose = FALSE)
# Multilevel LC model
out = multiLCA(data, Y, 3, id_high, 2, verbose = FALSE)
# Multilevel LC model lower-level covariates
out = multiLCA(data, Y, 3, id_high, 2, Z, verbose = FALSE)
```

14 plot.multiLCA

```
out
# Multilevel LC model lower- and higher-level covariates
out = multiLCA(data, Y, 3, id_high, 2, Z, Zh, verbose = FALSE)
out
# Model selection over single-level models with 1-3 classes
out = multiLCA(data, Y, 1:3, verbose = FALSE)
out
# Model selection over multilevel models with 1-3 lower-level classes and
# 2 higher-level classes
out = multiLCA(data, Y, 1:3, id_high, 2, verbose = FALSE)
# Model selection over multilevel models with 3 lower-level classes and
# 1-2 higher-level classes
out = multiLCA(data, Y, 3, id_high, 1:2, verbose = FALSE)
out
# Model selection over multilevel models with 1-3 lower-level classes and
# 1-2 higher-level classes using the default sequential approach
out = multiLCA(data, Y, 1:3, id_high, 1:2, verbose = FALSE)
out
```

plot.multiLCA

Plots conditional response probabilities

## **Description**

Visualizes conditional response probabilities estimated by the multiLCA function. The method works for both single- and multilevel models.

Let out denote the list object returned by the multiLCA function. Executing plot(out) visualizes the conditional response probabilities given by the mPhi matrix in out.

## Usage

```
## S3 method for class 'multiLCA'
plot(x, horiz = FALSE, clab = NULL, ...)
```

## **Arguments**

X	The object returned by the multiLCA function
horiz	Whether item labels should be oriented horizontally (TRUE) or vertically (FALSE). Default FALSE
clab	A character vector with user-specified class labels, if available, in the order "Class 1", "Class 2", under the default settings, i.e. top-to-bottom. Default NULL
	Additional plotting arguments

plot.multiLCA 15

## Value

No return value

## **Examples**

```
# Use IEA data
data = dataIEA

# Define vector with names of columns with items
Y = colnames(data)[4+1:12]

# Define number of (low-level) classes
iT = 3

# Estimate single-level measurement model
out = multiLCA(data = data, Y = Y, iT = iT)
out

# Plot conditional response probabilities with default settings
plot(out)

# Plot with vertical item labels and custom class labels
plot(out, horiz = FALSE, clab = c("Maximal", "Engaged", "Subject"))
```

## **Index**

```
* datasets
dataIEA, 3
dataTOY, 4

* package
multilevLCA-package, 2

dataIEA, 3
dataTOY, 4

multiLCA, 2, 5, 14
multilevLCA-package, 2

plot.multiLCA, 2, 5, 14
```