

# Package ‘glca’

October 23, 2020

**Type** Package

**Title** An R Package for Estimating a Group-Effect in Latent Class Analysis

**Date** 2020-10-23

**Version** 1.1.0

**Author** Youngsun Kim [aut, cre],  
Hwan Chung [aut]

**Maintainer** Youngsun Kim <kim0sun@korea.ac.kr>

**Description** Fits latent class analysis (LCA) including group variable and covariates. The group variable can be handled either by multilevel LCA described in Vermunt (2003) <DOI:10.1111/j.0081-1750.2003.t01-1-00131.x> or standard LCA at each level of group variable. The covariates can be incorporated in the form of logistic regression (Bandeeno-Roche et al. (1997) <DOI:10.1080/01621459.1997.10473658>).

**License** GPL-3

**BugReports** <https://github.com/kim0sun/glca/issues/>

**Encoding** UTF-8

**LazyData** true

**Depends** R (>= 3.5.0)

**RoxygenNote** 7.1.1

**Imports** MASS, Rcpp

**LinkingTo** Rcpp

**URL** <https://kim0sun.github.io/glca/>

**Suggests** covr, testthat

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2020-10-23 10:20:02 UTC

## R topics documented:

glca-package	2
coef.glca	2
glca	3
glca.gof	7
gss08	9
item	10
nyts18	11
plot.glca	12
summary.glca	13
<b>Index</b>	<b>15</b>

---

glca-package	<i>Latent Class Analysis with Group Variable</i>
--------------	--

---

### Description

Fits latent class analysis (LCA) including group variable and covariates. The group variable can be handled either by multilevel LCA described in Vermunt (2003) <DOI:10.1111/j.0081-1750.2003.t01-1-00131.x> or standard LCA at each level of group variable. The covariates can be incorporated in the form of logistic regression (Bandein-Roche et al. (1997) <DOI:10.1080/01621459.1997.10473658>).

---

coef.glca	<i>Extracts glca Model Coefficients</i>
-----------	---

---

### Description

Extracts regression coefficients of glca model if the model includes covariates.

### Usage

```
## S3 method for class 'glca'
coef(
  object,
  intercept = FALSE,
  digits = max(3, getOption("digits") - 3),
  show.signif.stars = getOption("show.signif.stars"),
  ...
)
```

**Arguments**

object            an object of "glca".  
 intercept        a logical value for whether to print intercept".  
 digits            number of significant digits to use when printing.  
 show.signif.stars        logical. If TRUE, 'significance stars' are printed for each coefficient.  
 ...                further arguments passed to or from other methods.

**Value**

Coefficient matrix from the glca model

If the model has calculated standard errors, coefficient matrix contains standard errors, t-statistic, and its p-value.

**See Also**

[glca](#)

**Examples**

```
## For examples see example(glca)
```

---

glca	<i>Fits Latent Class Models for Data Containing Group Variable and Covariates</i>
------	---

---

**Description**

Function for fitting latent class models with multiple groups, which may or may not include latent class structure for group variable.

**Usage**

```
glca(
  formula,
  group = NULL,
  data = NULL,
  nclass = 3,
  ncluster = NULL,
  std.err = TRUE,
  measure.inv = TRUE,
  coeff.inv = TRUE,
  init.param = NULL,
  n.init = 10,
  testiter = 50,
```

```

maxiter = 1000,
eps = 1e-06,
na.rm = FALSE,
random.seed = NULL,
verbose = TRUE
)

```

### Arguments

formula	a formula for specifying manifest items and covariates using the "item" function.
group	an optional vector specifying a group of observations. Given group variable, group covariates can be incorporated.
data	a data frame containing the manifest item, covariates and group variable.
nclass	number of level-1 (individual-level) latent classes.
ncluster	number of level-2 (group-level) latent classes. When group and ncluster (>1) are given the multilevel latent class models will be fitted.
std.err	a logical value for whether calculating standard errors for estimates.
measure.inv	a logical value of the measurement invariance assumption across groups.
coeff.inv	a logical value of the coefficient invariance assumption across groups (random intercept model).
init.param	A set of model parameters to be used as the user-defined initial values for the EM algorithm. It should be list with the named parameters and have same structure of param of the glca output. In default, initial parameters are randomly generated.
n.init	number of randomly generated initial parameter sets to be used for avoiding the problem of local maxima.
testiter	number of iterations in the EM algorithm for each initial parameter set. The initial parameter set that provides the largest log-likelihood will be selected for estimating the model.
maxiter	maximum number of iterations for the EM algorithm.
eps	a convergence tolerance value. When the largest absolute difference between former estimates and current estimates is less than eps, the algorithm will stop updating and consider the convergence to be reached.
na.rm	a logical value for deleting the lines that have at least one missing manifest item. If na.rm = FALSE, MAR procedure will be conducted.
random.seed	In default, the set of initial parameters is drawn randomly. As the same value for random.seed guarantees the same initial parameters to be drawn, this argument can be used to generate an identical solution.
verbose	a logical value indicating whether glca should print the estimation procedure onto the screen.

## Details

The `glca` is the function for implementing LCA consist of two-type latent categorical variables (i.e., level-1 and level-2 latent class). The level-1 (individual-level) latent class is identified by the association among the individuals' responses to multiple manifest items, but level-2 (group-level) latent class is categorized by the prevalence of level-1 latent class for group variable. The function `glca` can handle two types of covariates: level-1 and level-2 covariates. If covariates vary across individuals, they are considered as level-1 covariates. When `group` and `ncluster (>1)` are given, covariates which are varying across groups are considered as level-2 covariates. Both types of covariates have effect on level-1 class prevalence.

The formula should consist of an `~` operator between two sides. Manifest items should be indicated in LHS of formula using `item` function and covariates should be specified in RHS of formula. For example,

```
item(y1,y2,y3) ~ 1
```

```
item(y1,y2,y3) ~ x1 + x2
```

where the first formula indicates LCA with three manifest variables (`y1`, `y2`, and `y3`) and no covariate, and the second formula includes two covariates (`x1` and `x2`). Two types of covariates (i.e., level-1 and level-2 covariates) will be automatically detected by `glca`.

The estimated parameters in `glca` are `rho`, `gamma`, `delta`, and `beta`. The set of item response probabilities for each level-1 class is `rho`. The sets of prevalences for level-1 and level-2 class are `gamma` and `delta`, respectively. The prevalence for level-1 class (i.e., `gamma`) can be modeled as logistic regression using level-1 and/or level-2 covariates. The set of logistic regression coefficients is `beta` in `glca` output.

## Value

`glca` returns an object of class "glca".

The function `summary` prints estimates for parameters and `glca.gof` function gives goodness of fit measures for the model.

An object of class "glca" is a list containing the following components:

<code>call</code>	the matched call.
<code>terms</code>	the <code>terms</code> object used.
<code>model</code>	a list of model description.
<code>var.names</code>	a list of names of data.
<code>datalist</code>	a list of data used for fitting.
<code>param</code>	a list of parameter estimates.
<code>std.err</code>	a list of standard errors for estimates.
<code>coefficient</code>	a list of logistic regression coefficients for prevalence of level-1 class.
<code>posterior</code>	a <code>data.frame</code> or a list of posterior probabilities of each individual for latent classes and each group for latent clusters.
<code>gof</code>	a list of goodness of fit measures.
<code>convergence</code>	a list containing information about convergence.

**Author(s)**

Youngsun Kim

**References**

Vermunt, J.K. (2003) Multilevel latent class models. *Sociological Methodology*, **33**, 213–239. doi: [10.1111/j.00811750.2003.t01100131.x](https://doi.org/10.1111/j.00811750.2003.t01100131.x)

Collins, L.M. and Lanza, S.T. (2009) *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. John Wiley & Sons Inc.

**See Also**

[gss08 nyts18](#)

**Examples**

```
##
## Example 1. GSS dataset
##
data("gss08")
# LCA
lca = glca(item(DEFECT, HLTH, RAPE, POOR, SINGLE, NOMORE) ~ 1,
           data = gss08, nclass = 3)
summary(lca)

# LCA with covariate(s)
lcr = glca(item(DEFECT, HLTH, RAPE, POOR, SINGLE, NOMORE) ~ AGE,
           data = gss08, nclass = 3)
summary(lcr)
coef(lcr)

# Multiple-group LCA (MGLCA)
mglca = glca(item(DEFECT, HLTH, RAPE, POOR, SINGLE, NOMORE) ~ 1,
            group = DEGREE, data = gss08, nclass = 3)
summary(mglca)

# Multiple-group LCA with covariate(s) (MGLCR)
mglcr = glca(item(DEFECT, HLTH, RAPE, POOR, SINGLE, NOMORE) ~ SEX,
            group = DEGREE, data = gss08, nclass = 3)
summary(mglcr)
coef(mglcr)

##
## Example 2. NYTS dataset
##
data("nyts18")
# Multilevel LCA (MLCA)
mlca = glca(item(ECIGT, ECIGAR, ESLT, EELCIGT, EHOOKAH) ~ 1,
            group = SCH_ID, data = nyts18, nclass = 3, ncluster = 2)
```

```
summary(mlca)

# MLCA with covariate(s) (MLCR)
# (SEX: level-1 covariate, SCH_LEV: level-2 covariate)
mlcr = glca(item(ECIGT, ECIGAR, ESLT, EELCIGT, EHOOKAH) ~ SEX + SCH_LEV,
            group = SCH_ID, data = nyts18, nclass = 3, ncluster = 2)
coef(mlcr)
```

---

glca.gof

*Goodness of Fit Tests for Fitted glca Model*


---

### Description

Provides AIC, CAIC, BIC, entropy and deviance statistic for goodness of fit test for the fitted model. Given object2, the function computes the log-likelihood ratio (LRT) statistic for comparing the goodness of fit for two models. The bootstrap p-value can be obtained from the empirical distribution of LRT statistic by choosing test = "boot".

### Usage

```
glca.gof(
  object,
  ...,
  test = NULL,
  nboot = 50,
  random.seed = NULL,
  criteria = c("logLik", "AIC", "CAIC", "BIC", "entropy"),
  maxiter = 500,
  eps = 1e-04,
  verbose = FALSE
)
```

### Arguments

object	an object of "glca", usually, a result of a call to glca.
...	an optional object of "glca" to be compared with object.
test	a character string indicating type of test (chi-square test or bootstrap) to obtain the p-value for goodness of fit test ("chisq" or "boot").
nboot	number of bootstrap samples, only used when test = "boot".
random.seed	random seed to have the equivalent solution for every bootstrap trials.
criteria	a character vector indicating criteria to be printed.
maxiter	an integer for maximum number of iteration for bootstrap sample.
eps	positive convergence tolerance for bootstrap sample.
verbose	an logical value for whether or not to print the result of a function's execution.

**Value**

gtable            a matrix with model goodness-of-fit criteria  
 dtable           a matrix with deviance statistic and bootstrap p-value  
 boot             a list of LRT statistics from each bootstrap sample

gtable, which is always included in output of this function, includes goodness-of-fit criteria which are indicated criteria arguments for the object(s). dtable are contained when the objects are competing models. (when used items of the models are identical) dtable prints deviance and p-value. (bootstrap or chi-square) Lastly, when the bootstrap sample is used, the  $G^2$ -statistics for each bootstrap samples will be included in return object..

**Author(s)**

Youngsun Kim

**References**

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723. doi: [10.1109/tac.1974.1100705](https://doi.org/10.1109/tac.1974.1100705)

Schwarz, G. (1978) Estimating the dimensions of a model. *The Annals of Statistics*, **6**, 461–464. doi: [10.1214/aos/1176344136](https://doi.org/10.1214/aos/1176344136)

Langeheine, R., Pannekoek, J., and van de Pol, F. (1996) Bootstrapping goodness-of-fit measures in categorical data analysis. *Sociological Methods and Research*. **24**. 492-516. doi: [10.1177/0049124196024004004](https://doi.org/10.1177/0049124196024004004)

Ramaswamy, V., Desarbo, W., Reibstein, D., & Robinson, W. (1993). An Empirical Pooling Approach for Estimating Marketing Mix Elasticities with PIMS Data. *Marketing Science*, 12(1), 103-124. doi: [10.1287/mksc.12.1.103](https://doi.org/10.1287/mksc.12.1.103)

**See Also**

[glca gss08 nyts18](#)

**Examples**

```
## Example 1.
## Model selection between two LCA models with different number of latent classes.
data(gss08)
class2 = glca(item(DEFECT, HLTH, RAPE, POOR, SINGLE, NOMORE) ~ 1,
              data = gss08, nclass = 2)
class3 = glca(item(DEFECT, HLTH, RAPE, POOR, SINGLE, NOMORE) ~ 1,
              data = gss08, nclass = 3)
class4 = glca(item(DEFECT, HLTH, RAPE, POOR, SINGLE, NOMORE) ~ 1,
              data = gss08, nclass = 4)

glca.gof(class2, class3, class4)
## Not run: glca.gof(class2, class3, class4, test = "boot")

## Example 2.
## Model selection between two MLCA models with different number of latent clusters.
```



```

cluster2 = glca(item(ECIGT, ECIGAR, ESLT, EELCIGT, EHOOKAH) ~ 1,
                group = SCH_ID, data = nyts18, nclass = 2, ncluster = 2)
cluster3 = glca(item(ECIGT, ECIGAR, ESLT, EELCIGT, EHOOKAH) ~ 1,
                group = SCH_ID, data = nyts18, nclass = 2, ncluster = 3)

glca.gof(cluster2, cluster3)
## Not run: glca.gof(cluster2, cluster3, test = "boot")

## Example 3.
## MGLCA model selection under the measurement (invariance) assumption across groups.
measInv = glca(item(DEFECT, HLTH, RAPE, POOR, SINGLE, NOMORE) ~ 1,
               group = DEGREE, data = gss08, nclass = 3)
measVar = glca(item(DEFECT, HLTH, RAPE, POOR, SINGLE, NOMORE) ~ 1,
               group = DEGREE, data = gss08, nclass = 3, measure.inv = FALSE)

glca.gof(measInv, measVar)
glca.gof(measInv, measVar, test = "chisq")

```

---

## Description

This dataset includes 6 manifest items about abortion and several covariates from 355 respondents to the 2008 General Social Survey. Respondents answer the questions whether or not think it should be possible for a pregnant woman to obtain a legal abortion. The covariates include age, sex, race, region, and degree of respondents.

## Format

A data frame with 355 observations on 11 variables.

DEFECT If there is a strong chance of serious defect in the baby?

HLTH If the womans own health is seriously endangered by the pregnancy?

RAPE If she became pregnant as a result of rape?

POOR If the family has a very low income and cannot afford any more children?

SINGLE If she is not married and does not want to marry the man?

NOMORE If she is married and does not want any more children?

AGE Respondent's age

SEX Respondent's race

RACE Respondent's sex

REGION Region of interview

DEGREE Respondent's degree

**Source**

<http://gss.norc.org>

**References**

Smith, Tom W, Peter Marsden, Michael Hout, and Jibum Kim. General Social Surveys, 2008/Principal Investigator, Tom W. Smith; Co-Principal Investigator, Peter V. Marsden; Co-Principal Investigator, Michael Hout; Sponsored by National Science Foundation. -NORC ed.- Chicago: NORC at the University of Chicago

**Examples**

```
data("gss")
# Model 1: LCA
lca = glca(item(DEFECT, HLTH, RAPE, POOR, SINGLE, NOMORE) ~ 1,
           data = gss08, nclass = 3)
summary(lca)

# Model 2: LCA with a covariate
lcr = glca(item(DEFECT, HLTH, RAPE, POOR, SINGLE, NOMORE) ~ SEX,
           data = gss08, nclass = 3)
summary(lcr)
coef(lcr)

# Model 3: MGLCA
mglca = glca(item(DEFECT, HLTH, RAPE, POOR, SINGLE, NOMORE) ~ 1,
            group = REGION, data = gss08, nclass = 3)

# Model 4: MGLCA with covariates
summary(mglca)
mglcr = glca(item(DEFECT, HLTH, RAPE, POOR, SINGLE, NOMORE) ~ AGE,
            group = SEX, data = gss08, nclass = 3)
summary(mglcr)
coef(mglcr)
```

---

item

*Specifies Manifest Items for glca*

---

**Description**

Specifying manifest items in formula of `glca` function.

**Usage**

```
item(..., starts.with = NULL, ends.with = NULL)
```

**Arguments**

... vectors of manifest items. These can be given as named arguments which is colnames of data.frame.

starts.with a string for prefix of variable names to be selected.

ends.with a string for suffix of variable names to be selected.

**Value**

a matrix of specified variables, which contains names and levels of manifest items.

**See Also**

[glca](#)

**Examples**

```
## For examples see example(glca)
```

---

nyts18

*National Youth Tobacco Survey (NYTS) 2018*


---

**Description**

This dataset includes 5 manifest items about abortion and several covariates. From the original 2018 National Youth Tobacco Survey data, the Non Hispanic, white students are selected and schools with 30-50 students were selected. Thus, the dataset has 1743 respondents. The covariates include the sex of the respondents and the school ID to which the respondents belong, and the level of the corresponding school.

**Format**

A data frame with 1734 observations on the following 8 variables.

ECIGT Whether to have tried cigarette smoking, even one or two puffs

ECIGAR Whether to have ever tried cigar smoking, even one or two puffs

ESLT Whether to have used chewing tobacco, snuff, or dip

EELCIGT Whether to have used electronic cigarettes or e-cigarettes

EHOOKAH Whether to have tried smoking tobacco from a hookah or a waterpipe

SEX Respondent's Sex

SCH\_ID School ID to which the respondent belongs

SCH\_LEV Level of the corresponding school

**Source**

[https://www.cdc.gov/tobacco/data\\_statistics/surveys/nyts/index.htm](https://www.cdc.gov/tobacco/data_statistics/surveys/nyts/index.htm)

**Examples**

```

data("nyts18")

# Model 1: LCA
lca = glca(item(ECIGT, ECIGAR, ESLT, EELCIGT, EHOOKAH) ~ 1,
           data = nyts18, nclass = 3)
summary(lca)

# Model 2: LCR
lca = glca(item(ECIGT, ECIGAR, ESLT, EELCIGT, EHOOKAH) ~ SEX,
           data = nyts18, nclass = 3)
summary(lca)
coef(lca)

# Model 3: MGLCA
mglca = glca(item(ECIGT, ECIGAR, ESLT, EELCIGT, EHOOKAH) ~ 1,
             group = SEX, data = nyts18, nclass = 3)
summary(mglca)

# Model 4: MLCA
mlca = glca(item(ECIGT, ECIGAR, ESLT, EELCIGT, EHOOKAH) ~ 1,
            group = SCH_ID, data = nyts18, nclass = 3, ncluster = 2)
summary(mlca)

# Model 5: MLCA with level-1 covariate(s) only
mlcr = glca(item(ECIGT, ECIGAR, ESLT, EELCIGT, EHOOKAH) ~ SEX,
            group = SCH_ID, data = nyts18, nclass = 3, ncluster = 2)
summary(mlcr)
coef(mlcr)

# Model 6: MLCA with level-1 and level-2 covariate(s)
# (SEX: level-1 covariate, PARTY: level-2 covariate)
mlcr2 = glca(item(ECIGT, ECIGAR, ESLT, EELCIGT, EHOOKAH) ~ SEX + SCH_LEV,
             group = SCH_ID, data = nyts18, nclass = 3, ncluster = 2)
summary(mlcr2)
coef(mlcr2)

```

---

plot.glca

*Plots the Estimated Parameters of Fitted glca Model*


---

**Description**

plot method for class "glca".

**Usage**

```

## S3 method for class 'glca'
plot(x, ask = TRUE, ...)

```

**Arguments**

x                    an object of "glca", usually, a result of a call to glca.  
 ask                  a logical value whether to be asked before printing each plot.  
 ...                  further arguments passed to or from other methods.

**Value**

This function plots estimated parameters of model.

**See Also**

[glca gss08 nyts18](#)

**Examples**

```
## Not run:
# LCA
lca = glca(item(DEFECT, HLTH, RAPE, POOR, SINGLE, NOMORE) ~ 1,
           data = gss08, nclass = 3, na.rm = TRUE)
plot(lca)

# Multitple Group LCA (MGLCA)
mglca1 = glca(item(DEFECT, HLTH, RAPE, POOR, SINGLE, NOMORE) ~ 1,
              group = DEGREE, data = gss08, nclass = 3)
plot(mglca1)

# Multitple Group LCA (MGLCA) (measure.inv = FALSE)
mglca2 = glca(item(DEFECT, HLTH, RAPE, POOR, SINGLE, NOMORE) ~ 1,
              group = DEGREE, data = gss08, nclass = 3, measure.inv = FALSE)
plot(mglca2)
plot(mglca2, "all")

# Multilvel LCA (MLCA)
mlca = glca(item(ECIGT, ECIGAR, ESLT, EELCIGT, EHOOKAH) ~ 1,
            group = SCH_ID, data = nyts18, nclass = 3, ncluster = 3)
plot(mlca)

## End(Not run)
```

---

summary.glca

*Summarizes the Estimated Parameters of Fitted glca Model*

---

**Description**

summary method for class "glca".

**Usage**

```
## S3 method for class 'glca'  
summary(object, digits = max(3, getOption("digits") - 3), ...)
```

**Arguments**

object	an object of "glca", usually, a result of a call to glca
digits	the number of digits to be printed
...	further arguments passed to or from other methods

**Value**

This function prints descriptions of model and its more detailed estimated parameters but returns NULL.

**See Also**

[glca](#)

**Examples**

```
## For examples see example(glca)
```

# Index

## \* datasets

gss08, 9

nyts18, 11

coef.glca, 2

glca, 3, 3, 8, 11, 13, 14

glca-package, 2

glca.gof, 7

gss08, 6, 8, 9, 13

item, 10

nyts18, 6, 8, 11, 13

plot.glca, 12

summary.glca, 13

terms, 5