

# Package ‘glca’

May 30, 2020

**Type** Package

**Title** Latent Class Analysis with Grouped Data

**Date** 2020-05-29

**Version** 0.1.1

**Author** Youngsun Kim,  
Hwan Chung

**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 (Bandein-Roche et al. (1997) <DOI:10.1080/01621459.1997.10473658>).

**License** GPL (>= 2)

**Encoding** UTF-8

**LazyData** true

**Depends** R (>= 3.5.0)

**Imports** Rcpp (>= 1.0.3), MASS

**LinkingTo** Rcpp

**RoxygenNote** 6.1.1

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2020-05-30 12:50:07 UTC

## R topics documented:

glca-package . . . . .	2
brfss . . . . .	2
coef.glca . . . . .	4
glca . . . . .	5

glca.gof . . . . .	8
gss . . . . .	10
item . . . . .	11
nhanes . . . . .	12
plot.glca . . . . .	13
summary.glca . . . . .	14

<b>Index</b>	<b>16</b>
--------------	-----------

---

glca-package	<i>Latent Class Analysis with Grouped Data</i>
--------------	--

---

### Description

Fitting latent class models for data containing group variable.  
Standard LCA (LCA), multiple-group LCA (MGLCA) and multilevel LCA (MLCA) can be fitted.

---

brfss	<i>Behavioral Risk Factor Surveillance System 2017 (BRFSS 2017)</i>
-------	---

---

### Description

This data is comprised of questions related to healthy lifestyles, such as obesity, exercise time, eating habits, and smoking and drinking behaviors from the BRFSS 2017 survey. State is a group variable indicating the respondent's residential state. The data includes both individual-level (level-1) and group-level (level-2) covariates. The level-1 covariates include respondents' gender and income level. The level-2 covariates are regions of the United States and political parties which won the 2016 presidential election in the state.

### Format

A data frame with 444023 observations on the following 10 variables.

OBESE (original : \_RFBMI5) Adults who have a body mass index greater than 25.00 (Overweight or Obese)

- (1) No
- (2) Yes

PA300 (original : \_PA300R2) Adults that participated in 300 minutes (or vigorous equivalent minutes) of physical activity per week

- (1) 301+ minutes
- (2) 1-300 minutes
- (3) 0 minutes

FRTL1A (original : \_FRTL1A) Consume fruit 1 or more times per day

- (1) Yes
- (2) No

VEGLT1A (original : \_VEGLT1A) Consume vegetables 1 or more times per day  
 (1) Yes  
 (2) No

SMOKER (original : \_SMOKER3) Four-level smoker status  
 (1) Never smoked  
 (2) Former smoker  
 (3) Occasional smoker  
 (4) Everyday smoker

DRNK30 (original : DRNKANY5) Adults who reported having had at least one drink of alcohol in the past 30 days  
 (1) No  
 (2) Yes

SEX Respondents Sex

INCOME (original : INCOME2) Income level

STATE (original : \_STATE) State FIPS code

REGION Region to which the state belongs

PARTY The party that won the 2016 presidential election by state

### Source

<https://www.cdc.gov/brfss/>  
<https://doi.org/10.7910/DVN/42MVDX>

### References

Centers for Disease Control and Prevention. (2017). 2017 Behavioral Risk Factor Surveillance System Survey Data. Retrieved from <https://www.cdc.gov/brfss/>.

MIT Election Data and Science Lab, (2017). U.S. President 1976-2016. Retrieved from doi: [10.7910/DVN/42MVDX](https://doi.org/10.7910/DVN/42MVDX).

### Examples

```
data("brfss")

brfss1000 = brfss[sample(1:nrow(brfss), 1000),]

# Model 1: LCA
lca = glca(item(OBESE, PA300, FRTL1A, VEGLT1A, SMOKER, DRNK30) ~ 1,
  data = brfss1000, nclass = 3)
summary(lca)

# Model 2: MGLCA
mglca = glca(item(OBESE, PA300, FRTL1A, VEGLT1A, SMOKER, DRNK30) ~ 1,
  group = SEX, data = brfss1000, nclass = 3)
summary(mglca)

# Model 3: MGLCA with covariate(s)
```

```

mglcr = glca(item(OBESE, PA300, FRTL1A, VEGLT1A, SMOKER, DRNK30) ~ REGION,
  group = SEX, data = brfss1000, nclass = 3)
summary(mglcr)
coef(mglcr)

# Model 4: MLCA
mlca = glca(item(OBESE, PA300, FRTL1A, VEGLT1A, SMOKER, DRNK30) ~ 1,
  group = STATE, data = brfss1000, nclass = 3, ncluster = 2)
summary(mlca)

# Model 5: MLCA with level-1 covariate(s) only
mlcr = glca(item(OBESE, PA300, FRTL1A, VEGLT1A, SMOKER, DRNK30) ~ SEX,
  group = STATE, data = brfss1000, 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(OBESE, PA300, FRTL1A, VEGLT1A, SMOKER, DRNK30) ~ SEX + PARTY,
  group = STATE, data = brfss1000, nclass = 3, ncluster = 2)
summary(mlcr2)
coef(mlcr2)

```

---

coef.glca

*Extracts glca Model Coefficients*


---

## Description

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

## Usage

```

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

```

## Arguments

object	an object of "glca"
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 = 0,
      measure.inv = TRUE, std.err = TRUE, init.param = NULL,
      n.init = 1, 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.
measure.inv	a logical value of the measurement invariance assumption across groups
std.err	a logical value for whether calculating standard errors for estimates
init.param	a list which contains user-defined initial parameter
n.init	number of random initial parameter sets
testiter	an integer for maximum number of test set iteration
maxiter	an integer for maximum number of iteration
eps	positive convergence tolerance
na.rm	a logical value for whether or not to remove observations who has at least 1 item missing
random.seed	random seed to have the equivalent solution for every trials
verbose	a logical value for whether or not to print the result of a function's execution

## 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>model</code>	a list of model description
<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 data frame 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

[gss brfss](#)

## Examples

```
##
## Example 1. GSS dataset
##
data("gss")
# LCA
lca = glca(item(ABDEFECT, ABHLTH, ABRAPE, ABPOOR, ABSINGLE) ~ 1,
           data = gss, nclass = 3)
summary(lca)

# LCA with covariate(s)
lcr = glca(item(ABDEFECT, ABHLTH, ABRAPE, ABPOOR, ABSINGLE) ~ AGE,
           data = gss, nclass = 3)
summary(lcr)
coef(lcr)

# Multiple-group LCA (MGLCA)
mglca = glca(item(ABDEFECT, ABHLTH, ABRAPE, ABPOOR, ABSINGLE) ~ 1,
            group = DEGREE, data = gss, nclass = 3)
summary(mglca)

# Multiple-group LCA with covariate(s) (MGLCR)
mglcr = glca(item(ABDEFECT, ABHLTH, ABRAPE, ABPOOR, ABSINGLE) ~ SEX,
            group = DEGREE, data = gss, nclass = 3)
summary(mglcr)
coef(mglcr)

##
## Example 2. BRFSS data
##
data("brfss")
# Multilevel LCA (MLCA)
brfss1000 = brfss[sample(1:nrow(brfss), 1000),]
mlca = glca(item(OBESE, PA300, FRTL1A, VEGL1A, SMOKER, DRNK30) ~ 1,
           group = STATE, data = brfss1000, nclass = 3, ncluster = 2)
summary(mlca)

# MLCA with covariate(s) (MLCR)
```

```
# (SEX: level-1 covariate, REGION: level-2 covariate)
mlcr = glca(item(OBESE, PA300, FRTL1A, VEGLT1A, SMOKER, DRNK30) ~ SEX + PARTY,
            group = STATE, data = brfss1000, nclass = 3, ncluster = 2)
summary(mlcr)
coef(mlcr)
```

---

glca.gof

*Goodness of Fit Tests for Fitted glca Model*


---

### Description

Provides AIC, BIC, 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, object2 = NULL, test = NULL, nboot = 25,
         random.seed = NULL, maxiter = 500, eps = 1e-05, verbose = TRUE)
```

### Arguments

<code>object</code>	an object of "glca", usually, a result of a call to <code>glca</code>
<code>object2</code>	an optional object of "glca" to be compared with <code>object</code>
<code>test</code>	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")
<code>nboot</code>	number of bootstrap samples, only used when <code>test = "boot"</code>
<code>random.seed</code>	random seed to have the equivalent solution for every bootstrap trials
<code>maxiter</code>	an integer for maximum number of iteration for bootstrap sample
<code>eps</code>	positive convergence tolerance for bootstrap sample
<code>verbose</code>	an logical value for whether or not to print the result of a function's execution

### Value

<code>criteria</code>	a table with model fit criteria
<code>dev.table</code>	a table with deviance statistic and bootstrap p-value
<code>boot</code>	a list of LRT statistics from each bootstrap sample

### 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)

## See Also

[glca.gss](#)

## Examples

```
## Example 1.
## Model selection between two LCA models with different number of latent classes.
data(gss)
class2 = glca(item(ABDEFECT, ABHLTH, ABRAPE, ABPOOR, ABSINGLE) ~ 1,
              data = gss, nclass = 2)
class3 = glca(item(ABDEFECT, ABHLTH, ABRAPE, ABPOOR, ABSINGLE) ~ 1,
              data = gss, nclass = 3)

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

## Example 2.
## Model selection between two MLCA models with different number of latent clusters.
cluster2 = glca(item(ABDEFECT, ABHLTH, ABRAPE, ABPOOR, ABSINGLE) ~ 1,
                group = REGION, data = gss, nclass = 3, ncluster = 2, na.rm = TRUE)
cluster3 = glca(item(ABDEFECT, ABHLTH, ABRAPE, ABPOOR, ABSINGLE) ~ 1,
                group = REGION, data = gss, nclass = 3, ncluster = 3, na.rm = TRUE)

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

## Example 3.
## MGLCA model selection under the measurement (invariance) assumption across groups.
measInv = glca(item(ABDEFECT, ABHLTH, ABRAPE, ABPOOR, ABSINGLE) ~ 1,
               group = SEX, data = gss, nclass = 3)
measVar = glca(item(ABDEFECT, ABHLTH, ABRAPE, ABPOOR, ABSINGLE) ~ 1,
               group = SEX, data = gss, nclass = 3, measure.inv = FALSE)

glca.gof(measInv, measVar)
## Not run: glca.gof(measInv, measVar, test = "chisq")
## Not run: glca.gof(measInv, measVar, test = "boot")
```

---

gss

*General Social Study 2018*

---

## Description

This dataset includes 6 manifest items about abortion and several covariates from 2044 respondents to the 2018 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, region, and degree of respondents.

## Format

A data frame with 2044 observations on 9 variables.

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

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

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

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

ABRAPE If she became pregnant as a result of rape?

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

AGE Respondent's age

SEX 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, 2010/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(ABDEFECT, ABHLTH, ABRAPE, ABPOOR, ABSINGLE, ABNOMORE) ~ 1,
           data = gss, nclass = 3)
summary(lca)

# Model 2: LCA with a covariate
lcr = glca(item(ABDEFECT, ABHLTH, ABRAPE, ABPOOR, ABSINGLE, ABNOMORE) ~ SEX,
```

```

        data = gss, nclass = 3)
summary(lcr)
coef(lcr)

# Model 3: MGLCA
mglca = glca(item(ABDEFECT, ABHLTH, ABRAPE, ABPOOR, ABSINGLE, ABNOMORE) ~ 1,
             group = REGION, data = gss, nclass = 3)

# Model 4: MGLCA with covariates
summary(mglca)
mglcr = glca(item(ABDEFECT, ABHLTH, ABRAPE, ABPOOR, ABSINGLE, ABNOMORE) ~ AGE,
             group = SEX, data = gss, nclass = 3)
summary(mglcr)
coef(mglcr)

# Model 5: MLCA
mlca = glca(item(ABDEFECT, ABHLTH, ABRAPE, ABPOOR, ABSINGLE, ABNOMORE) ~ 1,
            group = REGION, data = gss, nclass = 3, ncluster = 2)
summary(mlca)

```

---

item	<i>Specifies Manifest Items for glca</i>
------	--

---

## 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 <code>data.frame</code> .
starts.with	a string
ends.with	a string

## Value

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

## See Also

[glca](#)

## Examples

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

---

nhanes

*National Health and Nutrition Examination Survey 2015-2016*

---

### **Description**

There is a 9 manifest items about frequency of depression symptoms over the past 2 weeks from National Health and Nutrition Examination Survey (NHANES). All of 9 items are measured 4 point-scale, "Not at all", "Several days", "More than half the days", and "Nearly every day". Those four categories are converted to binary as "Yes" ("Nearly every day", "More than half the days", "Several days") and "No" ("Not at all"). The dataset includes gender, age, and race of respondents as covariates.

### **Format**

A data frame with 5735 observations on the following 8 variables.

DPQ010 Have little interest in doing things  
DPQ020 Feeling down, depressed, or hopeless  
DPQ030 Trouble sleeping or sleeping too much  
DPQ040 Feeling tired or having little energy  
DPQ050 Poor appetite or overeating  
DPQ060 Feeling bad about yourself  
DPQ070 Trouble concentrating on things  
DPQ080 Moving or speaking slowly or too fast  
DPQ090 Thought you would be better off dead  
AGE Age in years at screening  
GENDER Respondent's Gender  
RACE Race/Hispanic origin w/ NH Asian

### **Source**

<https://www.cdc.gov/Nchs/Nhanes/>

### **References**

Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, [2015-2016][<https://www.cdc.gov/nchs/nhanes/ContinuousNhanes/Default.aspx?BeginYear=2015>]

**Examples**

```

data("nhanes")
# Model 1: LCA
lca = glca(item(starts.with = "DPQ") ~ 1, data = nhanes, nclass = 2)
summary(lca)

# Model 2: LCA with covariate(s)
lcr = glca(item(starts.with = "DPQ") ~ AGE, data = nhanes, nclass = 3)
summary(lcr)
coef(lcr)

# Model 3: MGLCA
mglca = glca(item(starts.with = "DPQ") ~ 1, group = GENDER,
             data = nhanes, nclass = 3)
summary(mglca)

# Model 4: MGLCA with covariate(s)
mglcr = glca(item(starts.with = "DPQ") ~ AGE, group = GENDER,
            data = nhanes, nclass = 3)
summary(mglcr)
coef(mglcr)

# Model 5: MLCA
mlca = glca(item(starts.with = "DPQ") ~ 1, group = RACE,
            data = nhanes, nclass = 3, ncluster = 2)
summary(mlca)

```

---

plot.glca

*Plots the Estimated Parameters of Fitted glca Model*


---

**Description**

plot method for class "glca".

**Usage**

```
## S3 method for class 'glca'
plot(x, group.name = NULL, ...)
```

**Arguments**

x	an object of "glca", usually, a result of a call to glca
group.name	a vector of strings which indicates groups, rho-parameters of which will be printed when measure.inv = FALSE.
...	further arguments passed to or from other methods

**Value**

This function plots estimated parameters of model.

**See Also**[glca](#)**Examples**

```
## Not run:
# LCA
lca = glca(item(ABDEFECT, ABHLTH, ABRAPE, ABPOOR, ABSINGLE, ABNOMORE) ~ 1,
           data = gss, nclass = 3, na.rm = TRUE)
plot(lca)

# Multiple Group LCA (MGLCA)
mglca1 = glca(item(ABDEFECT, ABHLTH, ABRAPE, ABPOOR, ABSINGLE, ABNOMORE) ~ 1,
              group = DEGREE, data = gss, nclass = 3)
plot(mglca1)

# Multiple Group LCA (MGLCA) (measure.inv = FALSE)
mglca2 = glca(item(ABDEFECT, ABHLTH, ABRAPE, ABPOOR, ABSINGLE, ABNOMORE) ~ 1,
              group = DEGREE, data = gss, nclass = 3, measure.inv = FALSE)
plot(mglca2)
plot(mglca2, "all")

# Multilevel LCA (MLCA)
mlca = glca(item(ABDEFECT, ABHLTH, ABRAPE, ABPOOR, ABSINGLE, ABNOMORE) ~ 1,
            group = REGION, data = gss, nclass = 3, ncluster = 2)
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

## \*Topic **datasets**

brfss, [2](#)

gss, [10](#)

nhanes, [12](#)

brfss, [2](#), [7](#)

coef.glca, [4](#)

glca, [5](#), [5](#), [9](#), [11](#), [14](#), [15](#)

glca-package, [2](#)

glca.gof, [8](#)

gss, [7](#), [9](#), [10](#)

item, [11](#)

nhanes, [12](#)

plot.glca, [13](#)

summary.glca, [14](#)