Package: gscaLCA (via r-universe)

November 10, 2026

Type Package

Title Generalized Structure Component Analysis- Latent Class Analysis & Latent Class Regression

Version 0.0.5

Description Execute Latent Class Analysis (LCA) and Latent Class Regression (LCR) by using Generalized Structured Component Analysis (GSCA). This is explained in Ryoo, Park, and Kim (2019) <doi:10.1007/s41237-019-00084-6>. It estimates the parameters of latent class prevalence and item response probability in LCA with a single line comment. It also provides graphs of item response probabilities. In addition, the package enables to estimate the relationship between the prevalence and covariates.

License GPL-3

Encoding UTF-8

LazyData true

URL https://github.com/hee6904/gscaLCA

Depends R (>= 2.10)

Imports gridExtra, ggplot2, stringr, progress, psych, fastDummies, fclust, MASS, devtools, foreach, doSNOW, nnet

Suggests knitr, rmarkdown

RoxygenNote 7.1.0

NeedsCompilation no

Author Jihoon Ryoo [aut], Seohee Park [aut, cre], Seoungeun Kim [aut], heungsun Hwaung [aut]

Maintainer Seohee Park <hee6904@gmail.com>

Config/pak/sysreqs cmake libfontconfig1-dev libfreetype6-dev libfribidi-dev git make libharfbuzz-dev libgit2-dev libicu-dev libjpeg-dev libpng-dev libtiff-dev libuv1-dev libwebp-dev libxml2-dev libssl-dev libx11-dev zlib1g-dev

Repository https://hee6904.r-universe.dev

Date/Publication 2020-06-08 20:30:02 UTC

RemoteUrl https://github.com/cran/gscaLCA

RemoteRef HEAD

RemoteSha c6edb398a11d48ef15fc7786a9341a1d50d451ba

Contents

AddHealth 2
gscaLCA 3
gscaLCR 5
summary.gscaLCA 7
TALIS 7

Index **9**

`AddHealth` *Add Health data about substance use*

Description

Add Health data about substance use

Usage

```
data(AddHealth)
```

Format

A data frame with 5114 observations on the following 8 variables.

AID A numeric vector of observations' ID.

Smoking A factor with levels "Yes" or "No"; Have you ever smoked an entire cigarette?

Alcohol A factor with levels "Yes" or "No";Have you had a drink of beer, wine, or liquor more than two or three times? Do not include sips or tastes from someone else's drink.

Drug A factor with levels "Yes" or "No"; Have you ever used any of the following drugs? Other types of illegal drugs, such as LSD, PCP, ecstasy, heroin, or mushrooms; or inhalants.

Marijuana A factor with levels "Yes" or "No"; Have you ever used any of the following drugs? Marijuana (hash, bhang, ganja)

Cocaine A factor with levels "Yes" or "No"; Have you ever used any of the following drugs? Cocaine (crack, coca leaves)

Gender A factor with levels "M" or "F"

Edu An integer vector from 1 to 8. It refers to the education level.

Details

This AddHealth data consist of 5,144 participants' responses with a randomly generated ID variable and five item variables, such as Smoking, Alcohol, Other Types of Illegal Drug, Marijuana, and Cocaine. The responses of the five items are dichotomous as either "Yes" or "No" and are treated the other missing codes as systematic missing. Along with the dichotomous responses, participants' gender and education level are also included in the sample data. This data can be obtained from the National Longitudinal Study of Adolescent to Adult Health (Add Health; Harris et al., 2009) where the study has mainly focused on the investigation of how health factors in childhood affect adult outcomes. In terms of data collection, there have been four additional waves since 1994. In this package, the data of a specific section of substance use at the wave IV is pre-installed.

Source

ICPSR Add Health

References

Harris, Kathleen Mullan, and Udry, J. Richard. National Longitudinal Study of Adolescent to Adult Health (Add Health), 1994-2008 [Public Use]. Ann Arbor, MI: Carolina Population Center, University of North Carolina-Chapel Hill [distributor], Inter-university Consortium for Political and Social Research [distributor], 2018-08-06. https://doi.org/10.3886/ICPSR21600.v21

Examples

```
data(AddHealth)
str(AddHealth)
head(AddHealth)
```

`gscaLCA` *Main function of gscaLCA by using fuzzy clustering GSCA*

Description

Fitting a component-based LCA by utilizing fuzzy clustering GSCA algorithm.

Usage

```
gscaLCA(
  dat,
  varnames = NULL,
  ID.var = NULL,
  num.class = 2,
  num.factor = "EACH",
  Boot.num = 20,
  multiple.Core = FALSE,
  covnames = NULL,
```

```
  cov.model = NULL,
  multinomial.ref = "MAX"
)
```

Arguments

dat	Data that you want to fit the gscaLCA function into.
varnames	A character vector. The names of columns to be used in the gscaLCA function.
ID.var	A character element. The name of ID variable. If ID variable is not specified, the gscaLCA function will search an ID variable in the given data. The ID of observations will be automatically generated as a numeric variable if the data set does not include any ID variable. The default is NULL.
num.class	A numeric element. The number of classes to be identified The default is 2.
num.factor	Either "EACH" or "ALLin1"."EACH" specifies the sitatuion that each indicator is assumed to be its phantom latent variable. "ALLin1" indicates that all variables are assumed to be explained by a common latent variable. The default is "EACH".
Boot.num	The number of bootstraps. The standard errors of parameters are computed from the bootstrap within the gscaLCA algorithm. The default is 20.
multiple.Core	A logical element. TRUE enables to use multiple cores for the bootstrap wehn they are available. The default is FASLE.
covnames	A character vector of covariates. The covariates are used when latent class regression (LCR) is fitted.
cov.model	A numeric vector. The indicator function of latent class regression (LCR) that covariates are involved in fitting the fuzzy clustering GSCA. 1 if gscaLCA is for LCR and otherwise 0.
multinomial.ref	A character element. Options of MAX, MIX, FIRST, and LAST are available for setting a reference group. The default is MAX.

Value

A list of the sample size (N), the number of cluster (C), the number of bootstraps (Boot.num/Boot.num.im), the model fit indices (model.fit), the latent class prevalence (LCprevalence), the item response probability (RespProb), the posterior membership & the predicted class membership (membership), and the graphs of item response probability (plot). When it include covariates, the regression results are also provided.

References

Ryoo, J. H., Park, S., & Kim, S. (2019). Categorical latent variable modeling utilizing fuzzy clustering generalized structured component analysis as an alternative to latent class analysis. Behaviormetrika, 47, 291-306. https://doi.org/10.1007/s41237-019-00084-6

Examples

```
#AddHealth data with 3 clusters with 500 samples
AH.sample= AddHealth[1:500,]
R3 = gscaLCA (dat = AH.sample,
               varnames = names(AddHealth)[2:6],
               ID.var = "AID",
               num.class = 3,
               num.factor = "EACH",
               Boot.num = 0)
summary(R3)
R3$model.fit      # Model fit
R3$LCprevalence   # Latent Class Prevalence
R3$RespProb       # Item Response Probability
head(R3$membership)     # Membership for all observations

# AddHealth data with 3 clusters with 500 samples with two covariates
R3_2C = gscaLCA (dat = AH.sample,
                 varnames = names(AddHealth)[2:6],
                 ID.var = "AID",
                 num.class = 3,
                 num.factor = "EACH",
                 Boot.num = 0,
                 multiple.Core = FALSE,
                 covnames = names(AddHealth)[7:8], # Gender and Edu
                 cov.model = c(1, 0),   # Only Gender varaible is added to the gscaLCR.
                 multinomial.ref = "MAX")

# To print with the results of multinomial regression with hard partitioning of the gscaLCR,
# use the option of "multinomial.hard".
summary(R3_2C, "multinomial.hard")


# AddHealth data with 2 clusters with 20 bootstraps
R2 = gscaLCA(AddHealth,
             varnames = names(AddHealth)[2:6],
             num.class = 2,
             Boot.num = 20,
             multiple.Core = FALSE) # "multiple.Core = TRUE" is recommended.
# TALIS data with 3 clusters with 20 bootstraps and the "ALLin1" option
T3 = gscaLCA(TALIS,
             varnames = names(TALIS)[2:6],
             num.class = 3,
             num.factor = "ALLin1",
             Boot.num = 20,
             multiple.Core = FALSE) # "multiple.Core = TRUE" is recommended.
```

gscaLCR *The 2nd and 3rd step of gscaLCA, which are the partitioning and fitting regression*

Description

The 2nd and 3rd step of gscaLCA, which are the partitioning and fitting regression in the latent class regression.

Usage

```
gscaLCR(results.obj, covnames, multinomial.ref = "MAX")
```

Arguments

results.obj	the results of gscaLCA.
covnames	A character vector of covariates. The covariates are used when latent class regression (LCR) is fitted.
multinomial.ref	A character element. Options of MAX, MIX, FIRST, and LAST are available for setting a reference group. The default is MAX.

Value

Results of the gscaLCR, fitting regression after partioning in addtion to gscaLCA results.

Examples

```
R2 = gscaLCA (dat = AddHealth[1:500, ], # Data has to include the possible covarite to run gscaLCR
              varnames = names(AddHealth)[2:6],
              ID.var = "AID",
              num.class = 3,
              num.factor = "EACH",
              Boot.num = 0,
              multiple.Core = F)

R2.gender = gscaLCR (R2, covnames = "Gender")
summary(R2.gender,  "multinomial.hard") # hard partitioning with multinomial regression
summary(R2.gender,  "multinomial.soft") # soft partitioning with multinomial regression
summary(R2.gender,  "binomial.hard")    # hard partitioning with binomial regression
summary(R2.gender,  "binomial.soft")    # soft partitioning with binomial regression
```

summary.gscaLCA *Summary of gscaLCA output or gscaLCR output*

Description

Summary of gscaLCA output or gscaLCR output

Usage

```
## S3 method for class 'gscaLCA'
summary(object, print.cov.output = NULL, ...)
```

Arguments

object	the object of gscaLCA or gscaLCR
print.cov.output	a character of what type partitioning and regression. Four possible option are possible "multinomial.hard", "multinomial.soft", "binomial.hard", and "binomial.soft".
...	Additional arguments affecting the summary produced.

Value

print model fit, prevalence, item probabilities, and regression results

Examples

```
# summary(R2)
```

TALIS *Teaching and Learning International Survey*

Description

Teaching and Learning International Survey

Usage

```
data(TALIS)
```

Format

A data frame with 2560 observations on the following 6 variables.

IDTEACH a numeric vector of teachers' ID.

Mtv_1 Integers with levels from 1 to 3 (1: not/low important, 2: moderate important, 3: high important); Motivation item 1: To become a teacher, teaching offered a steady career path.

Mtv_2 Integers with levels from 1 to 3 (1: not/low important, 2: moderate important, 3: high important); Motivation item 2: To become a teacher, teaching schedule fit with responsibilities in my personal life.

Pdgg_1 Integers with levels from 1 to 3 (1: not at all/to some extent, 2: quite a bit 3: a lot); Pedagogy item 1: What extend you can do help my students value learning.

Pdgg_2 Integers with levels from 1 to 3 (1: not at all/to some extent, 2: quite a bit 3: a lot); Pedagogy item 2: What extend you can do control disruptive behavior in the classroom.

Stsf Integers with levels from 1 to 3 (1: strongly disagree/disagree, 2: agree, 3: strongly agree); Satisfaction item: Feeling I enjoy working at this school.

Details

The Teaching and Learning International Survey (TALIS) 2018 focusing on teachers, school leaders, and the learning environment in schools was conducted by the Organization for Economic Cooperation and Development (OECD). There have been three cycles, TALIS 2008, TALIS 2013, and TALIS 2018. In this study, we utilize publicly available TALIS 2018 U.S. Data, 2,560 teachers' responses. The sample data include five items: two items are on motivation, two items are on pedagogy, and the last item is on satisfaction. Items' responses are originally four ordered categorical data of (1) Not at all, (2) To some extent, (3) Quite a bit, and (4) A lot. Due to too small frequencies in the first category, we modified them into three ordered categories.

Source

TALIS 2018 data

References

OECD (2019), TALIS 2018 Results (Volume I): Teachers and School Leaders as Lifelong Learners, TALIS, OECD Publishing, Paris, https://doi.org/10.1787/1d0bc92a-en.

Examples

```
str(TALIS)
head(TALIS)
```

Index

∗ **datasets**
AddHealth, 2
TALIS, 7

AddHealth, 2

gscaLCA, 3
gscaLCR, 5

summary.gscaLCA, 7

TALIS, 7