Package ‘gscaLCA’

October 9, 2019

Type Package
Title Generalized Structure Component Analysis- Latent Class Analysis
Version 0.0.2
Description Execute Latent Class Analysis (LCA) by using Generalized Structured Component Analysis (GSCA). This is explained in Ryoo, Park, and Kim (2009) <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 provide graphs of item response probabilities.
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
Suggests knitr, rmarkdown, R.rsp
RoxygenNote 6.1.1.9000
NeedsCompilation no
Author Jihoon Ryoo [aut], Seohee Park [aut, cre], Seoungeun Kim [aut], heungsun Hwaung [aut]
Maintainer Seohee Park <hee6904@gmail.com>
Repository CRAN
Date/Publication 2019-10-09 09:00:02 UTC

R topics documented:

AddHealth ....................................................... 2
gscaLCA .......................................................... 3
TALIS ............................................................... 4
Add Health data about substance use and abuse

Description

Add Health data about substance use and abuse

Usage

data(AddHealth)

Format

A data frame with 5114 observations on the following 6 variables.

AID  a numeric vector of observations' ID.
Smoking  a factor with levels "Yes" or "No"; H4TO1: Have you ever smoked an entire cigarette?
Alcohol  a factor with levels "Yes" or "No"; H4TO33: 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"; H4TO65E: Have you every used any of the following drugs? (65E) Other types of illegal drugs, such as LSD, PCP, ecstasy, heroin, or mushrooms; or inhalants.
Marijuana  a factor with levels "Yes" or "No"; H4TO65B: Have you every used any of the following drugs? (65B) Marijuana (hash, bhang, ganja)
Cocaine  a factor with levels "Yes" or "No"; H4TO65C: Have you every used any of the following drugs? (65C) Cocaine (crack, coca leaves)

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. This data can be obtained from the National Longitudinal Study of Adolescent to Adult Health (Add Health; Harris et al., 2009) where the study have 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


## gscALCA

**Main function of LCA by using fuzzy clustering GSCA**

### Description

This function enables to run LCA based on GSCA algorithm.

### Usage

```r
gscaLCA(dat, varnames = NULL, ID.var = NULL, num.cluster = 2,
         num.factor = "EACH", Boot.num = 20, multiple.Core = FALSE,
         verbose = TRUE, graphs_print = TRUE)
```

### Arguments

- **dat**: Data that you want to implement GSCA.
- **varnames**: A character vector. The names of columns to be used for gscALCA.
- **ID.var**: A character element. The name of ID variable. If ID variable is not specified, gscALCA will find a ID variable in a given data. The ID of observation are automatically created when data set does not have any ID variable. The default is NULL.
- **num.cluster**: A numeric element. Number of cluster to be analyzed. The default is 2.
- **num.factor**: Either "EACH" or "ALLin1"."EACH" indicates that each variable assumes to have latent variable. "ALLin1" indicates that all variables assumes to share one latent variable. The default is "EACH".
- **Boot.num**: Number of bootstrap. The standard errors of parameters are obtained by bootstrap in GSCA algorithm. The default is 20.
- **multiple.Core**: A logical element. TRUE enables to use multiple cores for the bootstrap. The default is FALSE.
- **verbose**: Logical. It indicates whether gscALCA prints output to the console. The default is TRUE.
- **graphs_print**: Logical. It indicates whether gscALCA prints the graphs. The default is TRUE.

### Examples

```r
data(AddHealth)
str(AddHealth)
head(AddHealth)
```
Value

A list of the used sample size (N), the number of cluster (C), the number of Bootstrap actually used (Boot.num.im), the model fit indices(model.fit), the latent class prevalence (LCprevalence), the item response probability (RespRrob), the posterior membership & the predicted class membership (membership), and the graphs of item response probability (plot).

References


Examples

```r
# ddHealth data with 2 clusters with 1000 samples
AH.sample = AddHealth[1:1000,]
R2 = gscaLCA(AH.sample, varnames = names(AddHealth)[2:6], num.cluster = 2, Boot.num=0)

R2$model.fit # Model fit
R2$LCprevalence # Latent Class Prevalence
R2$RespProb # Item Reponse Probability
R2$membership # Membership for all observations

# AddHealth data with 2 clusters
R2 = gscaLCA(AddHealth, varnames = names(AddHealth)[2:6], num.cluster = 2, Boot.num=0)

# TALIS data with 3 clusters
T3 = gscaLCA(TALIS, names(TALIS)[2:6], num.factor = "ALLin1", num.cluster = 3, Boot.num=0)
```

---

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); TT3G07A: 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); TT3G07D: To become a teacher, The 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); TT3G34B: 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); TT3G34D: What extent 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); TT3G53E: 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. We focused on 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. Due to too small frequencies, We modified them into three ordered categories.

Source

TALIS 2018 data

References


Examples

str(TALIS)
head(TALIS)
Index

*Topic datasets
  AddHealth, 2
  TALIS, 4

AddHealth, 2

gscaLCA, 3

TALIS, 4