

Package ‘tidyLPA’

February 18, 2018

Type Package

Title Easily Carry Out Latent Profile Analysis

Version 0.1.2

Description An interface to the 'mclust' package to easily carry out latent profile analysis (`LPA"). Provides functionality to estimate commonly-specified models. Follows a tidy approach, in that output is in the form of a data frame that can subsequently be computed on. Also has functions to interface to the commercial 'MPlus' software via the 'MplusAutomation' package.

License MIT + file LICENSE

Depends R (>= 2.10)

Imports dplyr, forcats, ggplot2, magrittr, mclust, purrr, readr, rlang, stringr, tibble, tidyr

Suggests covr, devtools, knitr, MplusAutomation, rmarkdown, roxygen2, testthat,

VignetteBuilder knitr

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

NeedsCompilation no

Author Joshua M Rosenberg [aut, cre],
Jennifer A Schmidt [ctb],
Patrick N Beymer [ctb],
Rebecca R Steingut [ctb]

Maintainer Joshua M Rosenberg <jmichaelrosenberg@gmail.com>

Repository CRAN

Date/Publication 2018-02-18 15:52:06 UTC

R topics documented:

bootstrap_lrt 2

compare_solutions	3
compare_solutions_mplus	4
estimate_profiles	5
estimate_profiles_mplus	6
extract_mplus_summary	8
pisaUSA15	9
plot_profiles	9
plot_profiles_mplus	10

Index 11

bootstrap_lrt	<i>Bootstrap the likelihood-ratio test statistic for mixture components</i>
---------------	---

Description

Bootstrap the likelihood-ratio test statistic for mixture components

Usage

```
bootstrap_lrt(df, ..., n_profiles, model = 1)
```

Arguments

df	data.frame with two or more columns with continuous variables
...	unquoted variable names separated by commas
n_profiles	the number of profiles (or mixture components) to be estimated
model	the mclust model to explore: 1 (varying means, equal variances, and residual covariances fixed to 0); 2 (varying means, equal variances and covariances; 3 (varying means and variances, covariances fixed to 0), 4 (varying means and covariances, equal variances; can only be specified in Mplus); 5 (varying means, equal variances, varying covariances); and 6 (varying means, variances, and covariances), in order least to most freely-estimated; see the introductory vignette for more information

Details

Bootstrap the p-values for the likelihood-ratio test statistic for the number of mixture components for an mclust model.

Examples

```
## Not run:
d <- pisaUSA15
d <- dplyr::sample_n(d, 200)
bootstrap_lrt(d,
             broad_interest, enjoyment, self_efficacy)

## End(Not run)
```

compare_solutions *Explore BIC for various models and numbers of profiles*

Description

Explore BIC for various models and numbers of profiles

Usage

```
compare_solutions(df, ..., n_profiles_range = 1:9, model = c(1, 2, 3, 6),
  center_raw_data = FALSE, scale_raw_data = FALSE, statistic = "BIC",
  return_table = FALSE, prior_control = F)
```

Arguments

df	data.frame with two or more columns with continuous variables
...	unquoted variable names separated by commas
n_profiles_range	a vector with the range of the number of mixture components to explore; defaults to 1 through 9 (1:9)
model	the mclust model to explore: 1 (varying means, equal variances, and residual covariances fixed to 0); 2 (varying means, equal variances and covariances; 3 (varying means and variances, covariances fixed to 0), 4 (varying means and covariances, equal variances; can only be specified in Mplus); 5 (varying means, equal variances, varying covariances); and 6 (varying means, variances, and covariances), in order least to most freely-estimated; see the introductory vignette for more information
center_raw_data	logical for whether to center ($M = 1$) the raw data (before clustering); defaults to FALSE
scale_raw_data	logical for whether to scale ($SD = 1$) the raw data (before clustering); defaults to FALSE
statistic	what statistic to plot; BIC or ICL are presently available as options
return_table	logical (TRUE or FALSE) for whether to return a table of the output instead of a plot; defaults to FALSE
prior_control	whether to include a regularizing prior; defaults to false

Details

Explore the BIC values of a range of models in terms of a) the structure of the residual covariance matrix and b) the number of mixture components (or profiles)

Value

a ggplot2 plot of the BIC values for the explored models

Examples

```
compare_solutions(iris, Sepal.Length, Sepal.Width, Petal.Length, Petal.Width)
```

```
compare_solutions_mplus
```

Explore BIC for various models and numbers of profiles using MPlus (requires purchasing and installing MPlus to use)

Description

Explore BIC for various models and numbers of profiles using MPlus (requires purchasing and installing MPlus to use)

Usage

```
compare_solutions_mplus(df, ..., n_profiles_max = 10, model = 1:6,
  starts = c(20, 4), m_iterations = 500, st_iterations = 10,
  convergence_criterion = 1e-06, save_models = NULL, return_table = FALSE,
  n_processors = 1)
```

Arguments

df	data.frame with two or more columns with continuous variables
...	unquoted variable names separated by commas
n_profiles_max	a vector with the range of the number of mixture components to explore; defaults to 2 through 10 (2:10)
model	which models to include; defaults to 1:6 (see https://jrosen48.github.io/tidyLPA/articles/Introduction_to_t)
starts	number of initial stage starts and number of final stage optimizations; defaults to c(20, 4); can be set to be more conservative to c(500, 50)
m_iterations	number of iterations for the EM algorithm; defaults to 500
st_iterations	the number of initial stage iterations; defaults to 10; can be set more to be more conservative to 50
convergence_criterion	convergence criterion for the Quasi-Newton algorithm for continuous outcomes; defaults to 1E-6 (.000001); can be set more conservatively to 1E-7 (.0000001)
save_models	whether to save the models as an rds file (i.e., set to "output.rds" to save the models with this filename)
return_table	logical (TRUE or FALSE) for whether to return a table of the output instead of a plot; defaults to FALSE
n_processors	= 1

Details

Explore the BIC values of a range of Mplus models in terms of a) the structure of the residual covariance matrix and b) the number of mixture components (or profiles)

Value

a list with a data.frame with the BIC values and a list with all of the model output; if save_models is the name of an rds file (i.e., "out.rds"), then the model output will be written with that filename and only the data.frame will be returned

Examples

```
## Not run:
compare_solutions_mplus(iris, Sepal.Length, Sepal.Width, Petal.Length, Petal.Width,
n_profiles_max = 4)

## End(Not run)
```

estimate_profiles	<i>Estimate parameters for profiles for a specific solution</i>
-------------------	---

Description

Estimate parameters for profiles for a specific solution

Usage

```
estimate_profiles(df, ..., n_profiles, model = 1, to_return = "tibble",
  center_raw_data = FALSE, scale_raw_data = FALSE,
  return_posterior_probs = TRUE, return_orig_df = FALSE,
  prior_control = FALSE, print_which_stats = "some")
```

Arguments

df	data.frame with two or more columns with continuous variables
...	unquoted variable names separated by commas
n_profiles	the number of profiles (or mixture components) to be estimated
model	the mclust model to explore: 1 (varying means, equal variances, and residual covariances fixed to 0); 2 (varying means, equal variances and covariances; 3 (varying means and variances, covariances fixed to 0), 4 (varying means and covariances, equal variances; can only be specified in Mplus); 5 (varying means, equal variances, varying covariances); and 6 (varying means, variances, and covariances), in order least to most freely-estimated; see the introductory vignette for more information
to_return	character string for either "tibble" (or "data.frame") or "mclust" if "tibble" is selected, then data with a column for profiles is returned; if "mclust" is selected, then output of class mclust is returned
center_raw_data	logical for whether to center ($M = 1$) the raw data (before clustering); defaults to FALSE

`scale_raw_data` logical for whether to scale (SD = 1) the raw data (before clustering); defaults to FALSE
`return_posterior_probs` TRUE or FALSE (only applicable if `to_return == "tibble"`); whether to include posterior probabilities in addition to the posterior profile classification; defaults to TRUE
`return_orig_df` TRUE or FALSE (if TRUE, then the entire data.frame is returned; if FALSE, then only the variables used in the model are returned)
`prior_control` whether to include a regularizing prior; defaults to false
`print_which_stats` if set to "some", prints (as a message) the log-likelihood, BIC, and entropy; if set to "all", prints (as a message) all information criteria and other statistics about the model; if set to any other values, then nothing is printed

Details

Creates profiles (or estimates of the mixture components) for a specific mclust model in terms of the specific number of mixture components and the structure of the residual covariance matrix

Value

either a tibble or a ggplot2 plot of the BIC values for the explored models

Examples

```
estimate_profiles(iris,
  Sepal.Length, Sepal.Width, Petal.Length, Petal.Width,
  model = 1,
  n_profiles = 3)
```

estimate_profiles_mplus

Estimate parameters for profiles for a specific solution (requires purchasing and installing MPlus to use)

Description

Estimate parameters for profiles for a specific solution (requires purchasing and installing MPlus to use)

Usage

```
estimate_profiles_mplus(df, ..., n_profiles, the_title = "test",
  data_filename = "d.dat", script_filename = "i.inp",
  output_filename = "i.out", savedata_filename = "d-mod.dat", model = 1,
  starts = c(20, 4), m_iterations = 500, st_iterations = 10,
  convergence_criterion = 1e-06, remove_tmp_files = TRUE,
  print_input_file = FALSE, return_save_data = TRUE, optseed = NULL,
  n_processors = 1)
```

Arguments

df	data.frame with two or more columns with continuous variables
...	unquoted variable names separated by commas
n_profiles	the number of profiles (or mixture components) to be estimated
the_title	title of the model; defaults to test
data_filename	name of data file to prepare; defaults to d.dat
script_filename	name of script to prepare; defaults to i.inp
output_filename	name of the output; defaults to o.out
savedata_filename	name of the output for the save data (with the original data conditional probabilities); defaults to o-mod.out
model	the mclust model to explore: 1 (varying means, equal variances, and residual covariances fixed to 0); 2 (varying means, equal variances and covariances; 3 (varying means and variances, covariances fixed to 0), 4 (varying means and covariances, equal variances; can only be specified in Mplus); 5 (varying means, equal variances, varying covariances); and 6 (varying means, variances, and covariances), in order least to most freely-estimated; see the introductory vignette for more information
starts	number of initial stage starts and number of final stage optimizations; defaults to c(20, 4); can be set to be more conservative to c(500, 50)
m_iterations	number of iterations for the EM algorithm; defaults to 500
st_iterations	the number of initial stage iterations; defaults to 10; can be set more to be more conservative to 50
convergence_criterion	convergence criterion for the Quasi-Newton algorithm for continuous outcomes; defaults to 1E-6 (.000001); can be set more conservatively to 1E-7 (.0000001)
remove_tmp_files	whether to remove data, script, and output files; defaults to TRUE
print_input_file	whether to print the input file to the console
return_save_data	whether to return the save data (with the original data and the posterior probabilities for the classes and the class assignment) as a data.frame along with the MPlus output; defaults to TRUE
optseed	random seed for analysis
n_processors	= 1

Details

Creates an mplus model (.inp) and associated data file (.dat)

Value

either a tibble or a ggplot2 plot of the BIC values for the explored models

Examples

```
## Not run:  
m1 <- estimate_profiles_mplus(iris,  
                             Sepal.Length, Sepal.Width, Petal.Length, Petal.Width,  
                             n_profiles = 2,  
                             model = 1)  
  
## End(Not run)
```

extract_mplus_summary *Extract summary statistics from an Mplus model*

Description

Extract summary statistics from an Mplus model

Usage

```
extract_mplus_summary(x)
```

Arguments

x an mplus model

Details

Extract log likelihood, BIC, and entropy statistics from an Mplus model

Value

a tibble with summary statistics

pisaUSA15	<i>student questionnaire data with four variables from the 2015 PISA for students in the United States</i>
-----------	--

Description

student questionnaire data with four variables from the 2015 PISA for students in the United States

Usage

```
pisaUSA15
```

Format

Data frame with columns #'

CNTSTUID international student ID

SCHID international school ID ...

Source

<http://www.oecd.org/pisa/data/>

plot_profiles	<i>Plot variable means and variances by profile for mclust output</i>
---------------	---

Description

Plot variable means and variances by profile for mclust output

Usage

```
plot_profiles(x, to_center = F, to_scale = F, plot_what = "tibble",
  plot_errorBars = TRUE)
```

Arguments

x	output from create_profiles_mclust()
to_center	whether to center the data before plotting
to_scale	whether to scale the data before plotting
plot_what	whether to plot tibble or mclust output from estimate_profiles(); defaults to tibble
plot_errorBars	whether to plot error bars (representing the 95 percent confidence interval for the mean of each variable)

Details

Plot the variable means and variances for data frame output from estimate_profiles()

Examples

```
m3 <- estimate_profiles(iris,
  Sepal.Length, Sepal.Width, Petal.Length, Petal.Width,
  model = 1,
  n_profiles = 3)
plot_profiles(m3)
```

plot_profiles_mplus *Plot variable means and variances by profile for MPlus output (requires purchasing and installing MPlus to use)*

Description

Plot variable means and variances by profile for MPlus output (requires purchasing and installing MPlus to use)

Usage

```
plot_profiles_mplus(mplus_data, to_center = T, to_scale = T)
```

Arguments

mplus_data	output from estimate_profiles_mclust() with return_savedata = T specified
to_center	whether to center the data before plotting
to_scale	whether to scale the data before plotting

Details

Plot the variable means and variances for data frame output from estimate_profiles_mclust()

Index

*Topic **datasets**

pisaUSA15, [9](#)

bootstrap_lrt, [2](#)

compare_solutions, [3](#)

compare_solutions_mplus, [4](#)

estimate_profiles, [5](#)

estimate_profiles_mplus, [6](#)

extract_mplus_summary, [8](#)

pisaUSA15, [9](#)

plot_profiles, [9](#)

plot_profiles_mplus, [10](#)