Package: ISCA (via r-universe)

March 2, 2025

Version 0.1.0

Author Lucas Drouhot [aut, cre], Marion Späth [aut]

Maintainer Lucas Drouhot <l.g.m.drouhot@uu.nl>

License GPL (>= 3)

Title Compare Heterogeneous Social Groups

Description The Inductive Subgroup Comparison Approach ('ISCA') offers a way to compare groups that are internally differentiated and heterogeneous. It starts by identifying the social structure of a reference group against which a minority or another group is to be compared, yielding empirical subgroups to which minority members are then matched based on how similar they are. The modelling of specific outcomes then occurs within specific subgroups in which majority and minority members are matched. ISCA is characterized by its data-driven, probabilistic, and iterative approach and combines fuzzy clustering, Monte Carlo simulation, and regression Analysis. ISCA_random_assignments() assigns subjects probabilistically to subgroups. ISCA_clustertable() provides summary statistics of each cluster across iterations. ISCA_modeling provides OLS regression results for each cluster across iterations.

Encoding UTF-8

RoxygenNote 7.3.1

Depends R (>= 4.3)

Imports data.table (>= 1.15.4), stringr (>= 1.5.1), e1071 (>= 1.7-14), Hmisc (>= 5.1-3), broom (>= 1.0.6), dplyr (>= 1.1.4), tidyselect (>= 1.2.1), stats (>= 4.3.1), tibble (>= 3.2.1), plyr (>= 1.8.9), magrittr (>= 2.0.3)

Suggests knitr, rmarkdown, tidyr (>= 1.3.0), testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

LazyData true

NeedsCompilation no

Config/pak/sysreqs make libicu-dev

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

RemoteUrl https://github.com/ldrouhot/isca

RemoteRef HEAD

RemoteSha 19057b903c48b6e00d99653e0719a97dca9d1d77

Contents

ISCA_clustertable *ISCA Cluster Tables*

Description

Function to create a cluster or descriptive table across iterations.

Usage

```
ISCA_clustertable(data, cluster_vars, draws = 500)
```

Arguments

data	The dataset including all relevant variables and the random assignments from the first ISCA_random_assignments()-function.
cluster_vars	A vector specifying the variables of interest.
draws	Specification of the number of probabilistic draws. The number of draws should be equal to the number of draws specified in the first step. If not specified, the default is 500.

Value

The output is a table containing the grand mean, grand standard deviation, and cluster error for each variable and cluster. No cluster error is calculated for dichotomous variables.

Examples

```
data(sim_data)
ISCA_step1 <- ISCA_random_assignments(data=sim_data, filter=native,
majority_group=1, minority_group=c(0), fuzzifier = 1.5, n_clusters=4,
draws=5, cluster_vars= c("female", "age", "education", "income"))
result_ISCA_clustertable <- ISCA_clustertable(data = ISCA_step1,
cluster_vars = c("native", "education", "age", "female",
"discrimination", "religiosity"), draws = 5);
```

ISCA_modeling *ISCA Modeling*

Description

Function to compute an OLS regression across all clusters and iterations.

Usage

```
ISCA_modeling(data, model_spec, weights = NULL, n_clusters, draws = 500)
```

Arguments

data	The dataset including all relevant variables and the random assignments from the first ISCA_random_assignments()-function.
model_spec	A model specification similar to the lm()-function.
weights	A vector specifying the variable in which the weights are stored. The default is NONE.
n_clusters	Specification of the number of clusters. This value should be equal to the number of clusters specified in the first and second step.
draws	Specification of the number of probabilistic draws. The number of draws should be equal to the number of draws specified in the first and second step. If not specified, the default is 500.

Value

The output is a table containing the regression coefficients, standard error and p-value for each regression term and cluster across all iterations. It also contains the regression coefficient, standard error and p-value for a pooled model, that is a model with all clusters combined.

Examples

```
data(sim_data)
ISCA_step1 <- ISCA_random_assignments(data=sim_data, filter=native,
majority_group=1, minority_group=c(0), fuzzifier = 1.5, n_clusters=4,
draws=5, cluster_vars= c("female", "age", "education", "income"))
ISCA_modeling_res <- ISCA_modeling(data= ISCA_step1,
model_spec="religiosity ~ native + female + age + education + discrimination",
draws = 5, n_clusters = 4);
```

ISCA_random_assignments

ISCA Random Assignments per Subgroup

Description

Function that calculates membership scores for each subgroup and assigns a cluster for a number of random draws.

Usage

```
ISCA_random_assignments(
  data,
  filter,
  majority_group,
  minority_group,
  cluster_vars,
  fuzzifier = 1.5,
  n_clusters,
  draws = 500
)
```

Arguments

data	A dataset containing all relevant variables
filter	Specification of the variable name that contains information on majority / minority status.
majority_group	Specification of the value within the variable specified in the previous filter-argument indicating majority status. This could be either a numeric value or a character string.
minority_group	specification of the value(s) indicating minority status in the filter variable. This could be either a numeric value or a character string. It can be one single minority group or a vector of several minority groups.
cluster_vars	Vector specifying the variables that should be used to create the clusters.
fuzzifier	The fuzzifier is a value larger than 1 determining the extent of overlap between clusters. A value of 1 effectively makes fuzzy c-means equivalent to hard k-means. The default is 1.5.
n_clusters	Specification of the number of clusters to be created.
draws	Specification of the number of probabilistic draws. If not specified, the default is 500.

Value

The output is a dataframe with all original variables and a new column for every draw, each containing one random assignment. This dataframe is the foundation of the subsequent functions in the ISCA package.

Examples

```
data(sim_data)
ISCA_step1 <- ISCA_random_assignments(data=sim_data,
filter=native, majority_group=1, minority_group=c(0),
fuzzifier = 1.5, n_clusters=4, draws=5,
cluster_vars= c("female", "age", "education", "income"));
```

---

sim_data   *Cross-sectional, artificial data on 1000 individuals*

Description

Small, artificially created dataset in a cross-sectional format. Provides information on 1000 individuals to illustrate the use of the package.

Usage

```
data(sim_data)
```

Format

A data frame with 1,000 rows and 7 columns:

female Dichotomous variable (0/1) indicating a person's sex

age value indicating a person's age, range 18-80

education value indicating a person's level of education, range 1-9

income value indicating a person's income

religiosity value indicating a person's level of religiosity, range 1-10

discrimination value indicating a person's level of experience discrimination, range 0-8

native Dichotomous variable (0/1) indicating whether a person is a native (1) or an immigrant (0)

References

The data was artificially created for the ISCA package.

Examples

```
data(sim_data)
head(sim_data)
```

Index