Package ‘swaprinc’

April 17, 2023

Title Swap Principal Components into Regression Models
Version 1.0.1
Description Obtaining accurate and stable estimates of regression coefficients can be challenging when the suggested statistical model has issues related to multicollinearity, convergence, or overfitting. One solution is to use principal component analysis (PCA) results in the regression, as discussed in Chan and Park (2005) <doi:10.1080/01446190500039812>. The swaprinc() package streamlines comparisons between a raw regression model with the full set of raw independent variables and a principal component regression model where principal components are estimated on a subset of the independent variables, then swapped into the regression model in place of those variables. The swaprinc() function compares one raw regression model to one principal component regression model, while the compswap() function compares one raw regression model to many principal component regression models. Package functions include parameters to center, scale, and undo centering and scaling, as described by Harvey and Hansen (2022) <https://cran.r-project.org/package=LearnPCA/vignettes/Vig_03_Step_By_Step_PCA.pdf>. Additionally, the package supports using Gifi methods to extract principal components from categorical variables, as outlined by Rossiter (2021) <https://www.css.cornell.edu/faculty/dgr2/_static/files/R_html/NonlinearPCA.html#2_Package>.

License MIT + file LICENSE
Encoding UTF-8
RoxygenNote 7.2.3
URL https://github.com/mncube/swaprinc
BugReports https://github.com/mncube/swaprinc/issues
Imports broom, broom.mixed, dplyr, Gifi, lme4, magrittr, rlang, tidyselect
Suggests testthat (>= 3.0.0)
Config/testthat/edition 3
NeedsCompilation no
Description

The `swaprinc` function compares a regression model using raw variables to a model with principal components swapped in. The `compswap` function compares a regression model with raw variables to multiple models with principal components swapped in. Parameter lists are recycled to ensure they are the same length as the longest parameter list.

Usage

```r
compswap(
data,
formula,
engine = "stats",
.prc_eng_list = list("stats"),
.pca_varlist = list(c(NULL)),
.n_pca_list = list(NULL),
.lpca_center_list = list("none"),
.lpca_scale_list = list("none"),
.lpca_undo_list = list(FALSE),
.gifi_transform_list = list("none"),
.gifi_trans_vars_list = list(c(NULL)),
.gifi_trans_dims_list = list(NULL),
.no_tresp_list = list(FALSE),
.miss_handler_list = list("none"),
.model_options_list = list("noaddpars"),
.prcomp_options_list = list("noaddpars"),
.gifi_princals_options_list = list("noaddpars"),
.gifi_trans_options_list = list("noaddpars")
)
```
**Arguments**

- **data**
  A dataframe

- **formula**
  A quoted model formula

- **engine**
  The engine for fitting the model. Options are "stats" or "lme4".

- **.prc_eng_list**
  A list of prc_eng values (see swaprinc documentation)

- **.pca_varlist**
  A list of pca_vars (see swaprinc documentation)

- **.n_pca_list**
  A list of n_pca_components (see swaprinc documentation)

- **.lpca_center_list**
  A list of lpca_center values (see swaprinc documentation)

- **.lpca_scale_list**
  A list of lpca_scale values (see swaprinc documentation)

- **.lpca_undo_list**
  A list of lpca_undo values (see swaprinc documentation)

- **.gifi_transform_list**
  A list of gifi_transform values (see swaprinc documentation)

- **.gifi_trans_vars_list**
  A list of gifi_trans_vars values (see swaprinc documentation)

- **.gifi_trans_dims_list**
  A list of gifi_trans_dims values (see swaprinc documentation)

- **.no_tresp_list**
  A list of no_tresp values (see swaprinc documentation)

- **.miss_handler_list**
  A list of miss_handler values (see swaprinc documentation)

- **.model_options_list**
  A list of model_options (see swaprinc documentation)

- **.prcomp_options_list**
  A list of prcomp_options (see swaprinc documentation)

- **.gifi_princals_options_list**
  A list of gifi_princals_options (see swaprinc documentation)

- **.gifi_trans_options_list**
  A list of gifi_trans_options (see swaprinc documentation)

**Value**

A list containing a list of fitted models and a comparison metrics data frame.

**Examples**

```r
# Load the iris dataset
data(iris)

# Define the formula

# Define the pca_varlist
pca_varlist <- list(c("Sepal.Width", "Petal.Length"),)
```
SWAPRINC

Description

Compare a regression model using raw variables with another model where principal components are extracted from a subset of the raw independent variables, and a user-defined number of these principal components are then used to replace the original subset of variables in the regression model.

Usage

swaprinc(
  data,
  formula,
  engine = "stats",
  prc_eng = "stats",
  pca_vars,
  n_pca_components, norun_raw = FALSE,
  lpca_center = "none",
  lpca_scale = "none",
  lpca_undo = FALSE,
  gifi_transform = "none",
  gifi_trans_vars,
  gifi_trans_dims,
  no_tresp = FALSE,
  miss_handler = "none",
model_options = "noaddpars",
prcomp_options = "noaddpars",
gifi_princals_options = "noaddpars",
gifi_trans_options = "noaddpars"
)

Arguments

data A dataframe
formula A quoted model formula
engine The engine for fitting the model. Options are 'stats' or 'lme4'.
prc_eng Then engine or extracting principal components. Options are 'stats', 'Gifi', and 'stats_Gifi'. The stats_Gifi engine uses tidyselect::where(is.numeric) to select the pca_vars for stats::prcomp and tidyselect::where(is.numeric) to select the pca_vars for Gifi::princals. Read Rossiter (2021) for more on princals.

pca_vars Variables to include in the principal component analysis. These variables will be swapped out for principal components

n_pca_components The number of principal components to include in the model. If using a complex prc_eng (i.e., stats_Gifi) then provide a named vector (i.e., n_pca_components = c("stats" = 2, "Gifi" = 3)).
norun_raw Include regression on raw variables if TRUE, exclude if FALSE.
lpca_center Center data as in the Step-by-Step PCA vignette (Harvey & Hanson, 2022). Only numeric variables will be included in the centering. Parameter takes values 'all' to center raw and pca variables, 'raw' to only center variables for the raw variable model fitting, 'pca' to only center pca_vars before pca regression model fitting, and 'none' to skip lpca centering.
lpca_scale Scale data as in the Step-by-Step PCA vignette. Only numeric variables will be included in the scaling. Parameter takes values 'all' to scale raw and pca variables, 'raw' to only scale variables for the raw variable model fitting, 'pca' to only scale pca_vars before pca regression model fitting, and 'none' to skip lpca scaling.
lpca_undo Undo centering and scaling of pca_vars as in the Step-by-Step PCA vignette.
gifi_transform Use Gifi optimal scaling to transform a set of variables. Parameter takes values 'none', 'all', 'raw', and 'pca'
gifi_trans_vars A vector of variables to include in the Gifi optimal scaling transformation
gifi_transdims Number of dimensions to extract in the Gifi optimal scaling transformation algorithm

no_tresp When set to TRUE, no_tresp (No transform response) will exclude the response variable from from pre-modeling and pre-pca transformations. Specifically, setting no_tresp to TRUE will exclude the response variable from the transformation specified in lpca_center and lpca_scale.
Choose how swaprinc handles missing data on the input data. Default is 'none'. Use 'omit' for complete case analysis.

**model_options**

Pass additional arguments to statistical modeling functions (i.e., stats::lm, stats::glm, lme4::lmer, lme4::glmer) Default is 'noaddpars' (no additional parameters)

**prcomp_options**

Pass additional arguments to stats::prcomp for prc_eng = 'stats' and prc_eng = 'stats_Gifi' call. Default is 'noaddpars' (no additional parameters)

**gifi_princals_options**

Pass additional arguments to Gifi::princals for prc_eng = 'Gifi' and prc_eng = 'stats_Gifi' call. Default is 'noaddpars' (no additional parameters)

**gifi_trans_options**

Pass additional arguments to Gifi::princals for gifi_transform. Default is 'noaddpars' (no additional parameters)

**Value**

A list with fitted models

**References**


2. Harvey, D. T., & Hanson, B. A. Step-by-Step PCA. (2022) [https://cran.r-project.org/package=LearnPCA/vignettes/Vig_03_Step_By_Step_PCA.pdf](https://cran.r-project.org/package=LearnPCA/vignettes/Vig_03_Step_By_Step_PCA.pdf)

**Examples**

data(iris)
res <- swaprinc(iris,
pca_vars = c("Sepal.Width", "Petal.Length", "Petal.Width"),
n_pca_components = 2)
Index

compswap, 2
swapinc, 4