Package 'responsePatterns'

May 9, 2026

Title Screening for Careless Responding Patterns

Version 0.1.1

Description
Some survey participants tend to respond carelessly which complicates data analysis. This package provides functions that make it easier to explore responses and identify those that may be problematic. See Gottfried et al. (2022) <doi:10.7275/vyxb-gt24> for more information.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

Depends R (>= 3.5.0)

Imports graphics, grDevices, methods, stats, utils

Suggests magrittr

NeedsCompilation no

Author Tomas Rihacek [aut, cre] (ORCID:
<https://orcid.org/0000-0001-5893-9289>),
Jaroslav Gottfried [aut] (ORCID:
<https://orcid.org/0000-0002-6076-1632>)

Maintainer Tomas Rihacek <rihacek@fss.muni.cz>

Repository CRAN

Date/Publication 2023-08-15 10:20:02 UTC

Contents

`responsePatterns` *responsePatterns: A package for finding instances of careless responding*

Description

Some survey participants tend to respond carelessly, which complicates data analysis. This package provides functions that make it easier to find repeated patterns in data and identify responses that may be problematic. This package implements two approaches to the problem of careless responses detection: one based on the auto-correlation approach and one based on a mechanistic approach. Both approaches yield scores that serve as estimates of how problematic the observations potentially are ("suspicion" scores). However, no conclusions should be made without a closer inspection of the problematic responses. Any decision about removing or downweighing an observation should be based on visual inspection of the responses, the specifics of the instrument used to collect the data, researchers' familiarity with the whole data set and the context of the data collection process.

Auto-correlation approach

The `rp.acors` function allows for a probabilistic detection of repetitive patterns in data. This function calculates auto-correlation coefficients for all lags up to a value defined by the max.lag parameter for each observation (respondent). Subsequently, it assigns a percentile value to each observation (respondent) based either on the highest absolute auto-correlation or the sum of absolute auto-correlations.

Mechanistic approach

The `rp.patterns` function searches for repetitive patterns in the data using an iterative algorithm. Patterns are defined based on the data themselves: if a sequence of values occurs more than once within an observation, it is considered a repetition. The algorithm counts the number of repetitions for different lengths of patterns and then weighs this sum by the length of the pattern (longer patterns are assigned higher weight). The total score for each respondent is determined as the sum of scores achieved for each pattern length and is standardized to a value between 0 and 1.

Auxiliary functions

The package provides auxiliary functions to summarize the responsePatterns object (`rp.summary`), extract indices (`rp.indices`, `rp.hist`, `rp.save2csv`) and to visually inspect individual responses (`rp.plot`, `rp.plots2pdf`).

References

Gottfried, J., Jezek, S., & Kralova, M. (2022). Autocorrelation screening: A potentially efficient method for detecting repetitive response patterns in questionnaire data. *Practical Assessment, Research, and Evaluation, 27*, Article 2. https://doi.org/10.7275/vyxb-gt24

`ResponsePatterns-class` *An S4 class to represent the results of response patterns analysis.*

Description

An S4 class to represent the results of response patterns analysis.

Slots

`id` A vector. Contains the ID variable (if declared by the user) or NAs (if not).

`n.obs` An integer. Number of observations (responses) in the data set.

`n.vars` An integer. Number of variables (excluding the ID variable, if declared).

`options` A list. Contains diverse options set by the user.

`percentile` An integer. If the rp.select() function is used to select a subsample, this keeps the information about the chosen percentile. Defaults to zero.

`data` A data frame. Stores the data.

`coefficients` A data frame. Stores the intermediate products of the analysis.

`indices` A data frame. Stores the final products of the analysis.

`rp.acors` *Auto-correlation screening*

Description

Auto-correlations of survey data allow for a probabilistic detection of repetitive patterns. This function calculates auto-correlation coefficients for all lags up to the value defined by the max.lag parameter for each observation (respondent). Subsequently, it assigns a percentile value to each observation (respondent) based either on the highest absolute auto-correlation or the sum of absolute auto-correlations. It is essential to keep the variables in the order in which they were presented to respondents.

Usage

```
rp.acors(
  data,
  max.lag = NULL,
  min.lag = 1,
  id.var = NULL,
  na.rm = FALSE,
  cor.method = c("pearson", "spearman", "kendall"),
  percentile.method = c("max", "sum"),
  na.top = FALSE,
  store.data = TRUE
)
```

Arguments

data	A data frame. A data set containing variables to analyze and, optionally, an ID variable.
max.lag	An integer. Define the maximum lag for which auto-correlations should be computed (defaults to the number of items minus 3).
min.lag	An integer. Define the minimum lag for which auto-correlations should be computed (defaults to 1).
id.var	A string. If the data set contains an ID variable, specify it's name.
na.rm	A logical scalar. Should missing values be removed from the computation of auto-correlations?
cor.method	A string. Defines the method used to compute auto-correlations (defaults to "pearson").
percentile.method	A string. Should the percentiles be based on the maximum absolute auto-correlation or on the sum of the absolute values of all auto-correlations (defaults to "max").
na.top	A logical scalar. Should NA indices (i.e., those that could not be computed due to data missingness) be ranked at the top? Defaults to FALSE.
store.data	A logical scalar. Should the data be stored within the object? Set to TRUE if you want to use the rp.plot or rp.save2csv functions.

Details

A response pattern yields perfect positive autocorrelation coefficient (r = 1) when the lag is equal to the length of the pattern, provided the pattern itself is uninterrupted over the whole vector of responses. There are two reasons for which the computation of auto-correlation computation can fail, both of which are associated with possible threat to data validity: (1) the pattern is composed of a vector of identical values (e.g., 2,2,2,2,2,2,2). In such cases, an auto-correlation coefficient cannot be computed due to a zero variance but we arbitrarily set the value to r = 1 because it meets the definition of a perfectly repetitive pattern; (2) the sequence contains too many missing values. In such cases we set the value to NA.

Choosing a suitable maximum lag value, i.e. the maximum number of positions for the data to be shifted in auto-correlation analysis, is very important for a reliable screening. Maximum lag value translates into the maximum length of a sequence within a repetitive response pattern that can be efficiently detected. A too low maximum lag value hinders auto-correlation screening ability to detect longer repetitive response patterns, thus potentially lowering the method's sensitivity (i.e., the ability to correctly detect careless responses). On the other hand, maximum lag value set too high generally lowers the reliability, because it makes the instrumental data matrix smaller and it, by calculating higher numbers of auto-correlation coefficients, allows for a higher frequency of occasionally strong auto-correlations that would inflate respondent's final auto-correlation score (determined as the highest absolute autocorrelation coefficient found for the respondent), thus lowering the method's specificity (i.e., the ability to correctly not detect attentive respondents). If not specified by the user, the max.lag value is set to the number of items minus 3.

In order to prevent bias, only questions with the same answer scales should be analyzed at one time, ideally. Analyzing responses on two scales with different number ranges together (e.g., answers on scale 1-5 and answers on scale 1-100) can bias the results to a great extent. See GitHub for an example of how to analyze data from several questionnaires simultaneously. Questions with unique scales or answer options where repetitive response patterns are unlikely or even impossible to emerge, like questions about gender or education, should be excluded prior to screening.

Value

Returns an S4 object of class "ResponsePatterns".

References

Gottfried, J., Jezek, S., & Kralova, M. (2021). *Autocorrelation screening: A potentially efficient method for detecting repetitive response patterns in questionnaire data.* Manuscript submitted for review.

See Also

`rp.patterns`, `rp.indices`, `rp.select`, `rp.hist`, `rp.plot`, `rp.save2csv`

Examples

```
rp.acors(rp.simdata, max.lag=10, id.var="optional_ID")
```

---

rp.hist *Plots a histogram of the main index*

Description

This function plots a histogram of the main "suspicion" index. The choice of the index depends on the type and setting of the analysis: it is either the maximum absolute auto-correlation or the sum of absolute auto-correlations if analyzed via the `rp.acors` function and the total score of analyzed via the `rp.patterns` function.

Usage

```
rp.hist(rp.object)
```

Arguments

rp.object	A ResponsePatterns object.

Value

Returns a plot.

See Also

rp.acors, rp.patterns

Examples

```
rp <- rp.acors(rp.simdata, id.var="optional_ID")
rp.hist(rp)
```

---

rp.indices *Extract indices from a ResponsePatterns object*

Description

This function extracts indices from a ResponsePatterns object.

Usage

```
rp.indices(rp.object, round = 2, include.coefs = TRUE)
```

Arguments

rp.object	A ResponsePatterns object.
round	An integer. The number of decimal places to which the indices should be rounded.
include.coefs	A logical scalar. Should the returned data frame include also the coefficients?

Value

Returns a data frame.

See Also

rp.acors, rp.patterns

Examples

```
rp <- rp.acors(rp.simdata, id.var="optional_ID")
rp.indices(rp)
```

`rp.patterns` *Repetitive pattern analysis*

Description

This function searches mechanically for repetitive patterns in the data. It searches for patterns of a given length (all values between min.length and max.length) using an iterative algorithm. The patterns are defined based on the data: if a sequence of values occurs more than once within an observation, it is considered a repetition. The algorithm counts the number of repetitions for each pattern length and then weighs this sum by the length of the pattern (longer patterns are assigned higher weight). The total score for each respondent is determined as the sum of scores achieved for each pattern length and is standardized to a value between 0 and 1. It is essential to keep the variables in the order in which they were presented to respondents.

Usage

```
rp.patterns(
  data,
  max.length = NULL,
  min.length = 2,
  id.var = NULL,
  na.rm = FALSE,
  std.patterns = TRUE,
  na.top = FALSE,
  store.data = TRUE
)
```

Arguments

`data`	A data frame. A data set containing variables to analyze and, optionally, an ID variable.
`max.length`	An integer. Define the maximum length of a pattern (cannot be longer than the number of variables/2).
`min.length`	An integer. Define the minimum length of a pattern (defaults to 2).
`id.var`	A string. If the data set contains an ID variable, specify it's name.
`na.rm`	A logical scalar. Should missing values be ignored when comparing sequences of data?
`std.patterns`	A logical scalar. If set to true, patterns are "standardized" by subtracting the minimum value from all elements in the sequence. As a result, patterns are compared in terms of their relative relationships (i.e., "1-2-3" and "3-4-5" are considered identical patterns). If set to FALSE, patterns are compared in terms of their absolute values (i.e., "1-2-3" and "3-4-5" are considered distinct patterns).

na.top	A logical scalar. Should NA indices (i.e., those that could not be computed due to data missingness) be ranked at the top? Defaults to FALSE.
store.data	A logical scalar. Should the data be stored within the object? Set to TRUE if you want to use the rp.plot or rp.save2csv functions.

Details

#' In order to prevent bias, only questions with the same answer scales should be analyzed at one time, ideally. Analyzing responses on two scales with different number ranges together (e.g., answers on scale 1–5 and answers on scale 1–100) can bias the results to a great extent. See GitHub for an example of how to analyze data from several questionnaires simultaneously. Questions with unique scales or answer options where repetitive response patterns are unlikely or even impossible to emerge, like questions about gender or education, should be excluded prior to screening.

Value

Returns an S4 object of class "ResponsePatterns".

See Also

rp.acors, rp.indices, rp.select, rp.hist, rp.plot, rp.save2csv

Examples

```
rp.patterns(rp.simdata, id.var="optional_ID")
```

---

rp.plot *Plot an individual response*

Description

This function plots an individual response for easier visual inspection. The observation can be identified by one of the following methods: observation number (obs), row name (rowname), or the value of the ID variable (id, if defined in the rp.object). Only one of these identifiers should be specified. Using this function requires that the data are stored in the ResponsePatterns object.

Usage

```
rp.plot(
  rp.object,
  obs = NULL,
  rowname = NULL,
  id = NULL,
  plot = TRUE,
  text.output = FALSE,
  groups = NULL,
  page.breaks = NULL,
  plot.lags = 10,
```

```
  bw = FALSE
)
```

Arguments

rp.object	A ResponsePatterns object.
obs	An integer. The number of observation to plot.
rowname	A string. The row name of the observation to plot.
id	A string. The value of the ID variable (if defined in the ResponsePatterns object).
plot	A logical scalar. Should the responses be plotted?
text.output	A logical scalar. Should the responses be printed to the console?
groups	A list of vectors. Defines groups of items that should be plotted using the same color.
page.breaks	A vector. Draws a vertical line after the specified items (useful if you want to display the pagination of the questionnaire in the plot).
plot.lags	How many lags should be displayed under the plot?
bw	A logical scalar. Should the plot be printed in black and white?

Value

Plots a graph.

See Also

rp.acors, rp.patterns, rp.plots2pdf

Examples

```
rp <- rp.acors(rp.simdata, id.var="optional_ID")
rp.plot(rp, obs=1)
rp.plot(rp, rowname="12", groups=list(c(1:10),c(11:20)))
rp.plot(rp, id="Natalya", page.breaks=c(5,10,15))
```

---

rp.plots2pdf *Export multiple plots to PDF*

Description

This function exports individual plots of all observations to a PDF file. Limit the number of observation via rp.select.

Usage

```
rp.plots2pdf(
  rp.object,
  file = "rp_plots.pdf",
  groups = NULL,
  page.breaks = NULL,
  bw = FALSE
)
```

Arguments

rp.object	A ResponsePatterns object.
file	A string. A filename of the PDF file.
groups	A list of vectors. Defines groups of items that should be plotted using the same color.
page.breaks	A vector. Draws a vertical line after the items (useful if you want to display the pagination of the questionnaire in the plot).
bw	A logical scalar. Should the plot be printed in black and white?

Details

If you have trouble exporting the PDF file, close all active graphical devices by running dev.off several times.

Value

Creates a PDF file.

See Also

rp.acors, rp.patterns, rp.plot

Examples

```
rp <- rp.acors(rp.simdata, id.var="optional_ID")
## Not run: rp.plots2pdf(rp)
```

---

rp.save2csv *Export indices into a CSV file*

Description

This functions exports the ResponsePatterns object indices and, optionally, coefficients and data.

Usage

```
rp.save2csv(
  rp.object,
  file = "rp_results.csv",
  csv = c("csv", "csv2"),
  include.coefs = TRUE,
  include.data = TRUE
)
```

Arguments

rp.object	A ResponsePatterns object.
file	A string. A filename or a path.
csv	A string. Specify the CSV file format.
include.coefs	A logical scalar. Should the exported file include the coefficients?
include.data	A logical scalar. Should the exported file include the data?

Value

Exports a CSV file.

See Also

rp.acors, rp.patterns, rp.indices

Examples

```
rp <- rp.acors(rp.simdata, id.var="optional_ID")
## Not run: rp.save2csv(rp)
## Not run: rp.save2csv(rp, include.coefs=FALSE, include.data=FALSE)
```

---

rp.select *Select observations*

Description

This function reorders observations and selects those equal of above a defined percentile.

Usage

```
rp.select(rp.object, percentile = 90)
```

Arguments

rp.object	A ResponsePatterns object.
percentile	An integer. Defines a percentile cutoff. Setting the value to zero keeps all observations but the data are ordered based on the percentile.

Value

A ResponsePatterns object.

See Also

rp.acors, rp.patterns

Examples

```
rp <- rp.acors(rp.simdata, id.var="optional_ID")
rp <- rp.select(rp, percentile=80)
```

---

rp.simdata *A simulated data set of survey responses.*

Description

A simulated data set of survey responses.

Usage

```
rp.simdata
```

Format

A data frame with 100 rows and 21 variables:

optional_ID fictive participants' names

Q_01 a survey item on a Likert-type scale from 1 to 5

Q_02 a survey item on a Likert-type scale from 1 to 5

Q_03 a survey item on a Likert-type scale from 1 to 5

Q_04 a survey item on a Likert-type scale from 1 to 5

Q_05 a survey item on a Likert-type scale from 1 to 5

Q_06 a survey item on a Likert-type scale from 1 to 5

Q_07 a survey item on a Likert-type scale from 1 to 5

Q_08 a survey item on a Likert-type scale from 1 to 5

Q_09 a survey item on a Likert-type scale from 1 to 5

Q_10 a survey item on a Likert-type scale from 1 to 5

Q_11 a survey item on a Likert-type scale from 1 to 5

Q_12 a survey item on a Likert-type scale from 1 to 5

Q_13 a survey item on a Likert-type scale from 1 to 5

Q_14 a survey item on a Likert-type scale from 1 to 5

Q_15 a survey item on a Likert-type scale from 1 to 5

Q_16 a survey item on a Likert-type scale from 1 to 5

Q_17 a survey item on a Likert-type scale from 1 to 5

Q_18 a survey item on a Likert-type scale from 1 to 5

Q_19 a survey item on a Likert-type scale from 1 to 5

Q_20 a survey item on a Likert-type scale from 1 to 5

Source

A simulated data set.

`rp.summary` *SUmmary of an ResponsePatterns object*

Description

SUmmary of an ResponsePatterns object

Usage

```
rp.summary(rp.object)
```

Arguments

`rp.object`	A ResponsePatterns object.

Value

Prints a summary of a ResponsePatterns object.

Examples

```
rp <- rp.acors(rp.simdata, id.var="optional_ID")
rp.summary(rp)
summary(rp)
```

Index