Package 'careless'

June 19, 2018

Type Package

Title Procedures for Computing Indices of Careless Responding

Version 1.1.3

Date 2018-06-19

Maintainer Richard Yentes <rdyentes@ncsu.edu>

Description When taking online surveys, participants sometimes respond to items without regard to their content.
These types of responses, referred to as careless or insufficient effort responding, constitute significant problems for data quality, leading to distortions in data analysis and hypothesis testing, such as spurious correlations. The 'R' package 'careless' provides solutions designed to detect such careless / insufficient effort responses by allowing easy calculation of indices proposed in the literature. It currently supports the calculation of longstring, even-odd consistency, psychometric synonyms/antonyms, Mahalanobis distance, and intra-individual response variability (also termed inter-item standard deviation). For a review of these methods, see Curran (2016) <doi:10.1016/j.jesp.2015.07.006>.

License MIT + file LICENSE

URL https://github.com/ryentes/careless/

BugReports https://github.com/ryentes/careless/issues

Imports psych

Suggests testthat, knitr

Encoding UTF-8

LazyData true

VignetteBuilder knitr

RoxygenNote 6.0.1

NeedsCompilation no

Author Richard Yentes [cre, aut],
Francisco Wilhelm [aut]

Repository CRAN

Date/Publication 2018-06-19 14:08:10 UTC

R topics documented:

`careless` *careless: A package providing procedures for computing indices of careless responding*

Description

Careless or insufficient effort responding in surveys, i.e. responding to items without regard to their content, is a common occurence in surveys. These types of responses constitute significant problems for data quality leading to distortions in data analysis and hypothesis testing, such as spurious correlations. The R package careless provides solutions designed to detect such careless / insufficient effort responses by allowing easy calculation of indices proposed in the literature. It currently supports the calculation of Longstring, Even-Odd Consistency, Psychometric Synonyms/Antonyms, Mahalanobis Distance, and Intra-individual Response Variability (also termed Inter-item Standard Deviation).

Statistical outlier function

- `mahad` computes Mahalanobis Distance, which gives the distance of a data point relative to the center of a multivariate distribution.

Consistency indices

- `evenodd` computes the Even-Odd Consistency Index. It divides unidimensional scales using an even-odd split; two scores, one for the even and one for the odd subscale, are then computed as the average response across subscale items. Finally, a within-person correlation is computed based on the two sets of subscale scores for each scale.
- `psychsyn` computes the Psychometric Synonyms Index, or, alternatively, the Psychometric Antonyms Index. Psychometrical synonyms are item pairs which are correlated highly positively, whereas psychometric antonyms are item pairs which are correlated highly negatively. A within-person correlation is then computed based on these item pairs.
- `psychant` is a convenience wrapper for `psychsyn` that computes psychological antonyms.
- `psychsyn_critval` is a helper designed to set an adequate critical value (i.e. magnitude of correlation) for the psychometric synonyms/antonyms index.

Response pattern functions

- longstring computes the longest (and optionally, average) length of consecutive identical responses given.
- irv computes the Intra-individual Response Variability (IRV), the "standard deviation of responses across a set of consecutive item responses for an individual" (Dunn et al. 2018)

Datasets

- careless_dataset, a simulated dataset with 200 observations and 10 subscales of 5 items each.
- careless_dataset2, a simulated dataset with 1000 observations and 10 subscales of 10 items each.

The sample datasets differ in the types of careless responding simulated.

Author(s)

Richard Yentes <rdyentes@ncsu.edu>, Francisco Wilhelm <franciscowilhelm@gmail.com>

---

careless_dataset *Simulated dataset with insufficient effort responses.*

Description

A simulated dataset mimicking insufficient effort responding. Contains three types of responses: (a) Normal responses with answers centering around a trait/attitude value (80 percent probability per simulated observation), (b) Straightlining responses (10 percent probability per simulated observation), (c) Random responses (10 percent probability per simulated observation). Simulated are 10 subscales of 5 items each (= 50 variables).

Usage

```
careless_dataset
```

Format

A data frame with 200 observations (rows) and 50 variables (columns).

careless_dataset2 *Simulated dataset with careless responses.*

Description

A simulated dataset mimicking insufficient effort responding. Contains three types of responses: (a) Normal responses with answers mimicking a diligent respondent (b) Some number of longstring careless responders, (c) some number of generally careless responders. Simulated are 10 subscales of 10 items each (= 100 variables).

Usage

```
careless_dataset2
```

Format

A data frame with 1000 observations (rows) and 100 variables (columns).

evenodd *Calculates the even-odd consistency score*

Description

Takes a matrix of item responses and a vector of integers representing the length each factor. The even-odd consistency score is then computed as the within-person correlation between the even and odd subscales over all the factors.

Usage

```
evenodd(x, factors, diag = FALSE)
```

Arguments

x	a matrix of data (e.g. survey responses)
factors	a vector of integers specifying the length of each factor in the dataset
diag	optionally returns a column with the number of available (i.e., non-missing) even/odd pairs per observation. Useful for datasets with many missing values.

Author(s)

Richard Yentes <rdyentes@ncsu.edu>, Francisco Wilhelm <franciscowilhelm@gmail.com>

References

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality, 39*, 103-129. doi: 10.1016/j.jrp.2004.09.009

Examples

```
careless_eo <- evenodd(careless_dataset, rep(5,10))
careless_eodiag <- evenodd(careless_dataset, rep(5,10), diag = TRUE)
```

`irv` *Calculates the intra-individual response variability (IRV)*

Description

The IRV is the "standard deviation of responses across a set of consecutive item responses for an individual" (Dunn, Heggestad, Shanock, & Theilgard, 2018, p. 108). By default, the IRV is calculated across all columns of the input data. Additionally it can be applied to different subsets of the data. This can detect degraded response quality which occurs only in a certain section of the questionnaire (usually the end). Whereas Dunn et al. (2018) propose to mark persons with *low* IRV scores as outliers - reflecting straightlining responses, Marjanovic et al. (2015) propose to mark persons with *high* IRV scores - reflecting highly random responses (see References).

Usage

```
irv(x, split = FALSE, num.split = 3)
```

Arguments

`x`	a matrix of data (e.g. survey responses)
`split`	boolean indicating whether to additionally calculate the IRV on subsets of columns (of equal length).
`num.split`	the number of subsets the data is to be split in.

Author(s)

Francisco Wilhelm <franciscowilhelm@gmail.com>

References

Dunn, A. M., Heggestad, E. D., Shanock, L. R., & Theilgard, N. (2018). Intra-individual Response Variability as an Indicator of Insufficient Effort Responding: Comparison to Other Indicators and Relationships with Individual Differences. *Journal of Business and Psychology, 33(1)*, 105-121. doi: 10.1007/s1086901694790

Marjanovic, Z., Holden, R., Struthers, W., Cribbie, R., & Greenglass, E. (2015). The inter-item standard deviation (ISD): An index that discriminates between conscientious and random responders. *Personality and Individual Differences*, 84, 79-83. doi: 10.1016/j.paid.2014.08.021

Examples

```
# calculate the irv over all items
irv_total <- irv(careless_dataset)

#calculate the irv over all items + calculate the irv for each quarter of the questionnaire
irv_split <- irv(careless_dataset, split = TRUE, num.split = 4)
boxplot(irv_split$irv4) #produce a boxplot of the IRV for the fourth quarter
```

`longstring` *Identifies the longest string of identical consecutive responses for each observation*

Description

Takes a matrix of item responses and, beginning with the second column (i.e., second item) compares each column with the previous one to check for matching responses. For each observation, the length of the maximum uninterrupted string of identical responses is returned. Additionally, can return the average length of uninterrupted string of identical responses.

Usage

```
longstring(x, avg = FALSE)
```

Arguments

`x`	a matrix of data (e.g. item responses)
`avg`	a boolean indicating whether to additionally return the average length of identical consecutive responses

Author(s)

Richard Yentes <rdyentes@ncsu.edu>, Francisco Wilhelm <franciscowilhelm@gmail.com>

References

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality, 39*, 103-129. doi: 10.1016/j.jrp.2004.09.009

Examples

```
careless_long <- longstring(careless_dataset, avg = FALSE)
careless_avg <- longstring(careless_dataset, avg = TRUE)
boxplot(careless_avg$longstr) #produce a boxplot of the longstring index
boxplot(careless_avg$avgstr)
```

mahad *Find and graph Mahalanobis Distance (D) and flag potential outliers.*

Description

Takes a matrix of item responses and computes Mahalanobis D. Can additionally return a vector of binary outlier flags. Mahalanobis distance is calculated using the function `psych::outlier` of the **psych** package, an implementation which supports missing values.

Usage

```
mahad(x, plot = TRUE, flag = FALSE, confidence = 0.99, na.rm = TRUE)
```

Arguments

`x`	a matrix of data
`plot`	Plot the resulting QQ graph
`flag`	Flag potential outliers using the confidence level specified in parameter `confidence`
`confidence`	The desired confidence level of the result
`na.rm`	Should missing data be deleted

Author(s)

Richard Yentes <rdyentes@ncsu.edu>, Francisco Wilhelm <franciscowilhelm@gmail.com>

References

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17(3)*, 437-455. doi: 10.1037/a0028085

See Also

`psych::outlier` on which this function is based.

Examples

```
mahad_raw <- mahad(careless_dataset) #only the distances themselves
mahad_flags <- mahad(careless_dataset, flag = TRUE) #additionally flag outliers
mahad_flags <- mahad(careless_dataset, flag = TRUE, confidence = 0.999) #Apply a strict criterion
```

psychant *Computes the psychometric antonym score*

Description

A convenient wrapper that calls psychsyn with argument anto = TRUE to compute the psychometric antonym score.

Usage

```
psychant(x, critval = -0.6, diag = FALSE)
```

Arguments

x	is a matrix of item responses
critval	is the minimum magnitude of the correlation between two items in order for them to be considered psychometric synonyms. Defaults to -.60
diag	additionally return the number of item pairs available for each subject. Useful if dataset contains many missing values.

Author(s)

Richard Yentes <rdyentes@ncsu.edu>, Francisco Wilhelm <franciscowilhelm@gmail.com>

See Also

psychsyn for the main function, psychsyn_critval for a helper that allows to set an adequate critical value for the size of the correlation.

Examples

```
antonyms <- psychant(careless_dataset2, .50)
antonyms <- psychant(careless_dataset2, .50, diag = TRUE)
```

psychsyn *Computes the psychometric synonym/antonym score*

Description

Takes a matrix of item responses and identifies item pairs that are highly correlated within the overall dataset. What defines "highly correlated" is set by the critical value (e.g., r > .60). Each respondents' psychometric synonym score is then computed as the within-person correlation between the identified item-pairs. Alternatively computes the psychometric antonym score which is a variant that uses item pairs that are highly *negatively* correlated.

Usage

```
psychsyn(x, critval = 0.6, anto = FALSE, diag = FALSE)
```

Arguments

x	is a matrix of item responses
critval	is the minimum magnitude of the correlation between two items in order for them to be considered psychometric synonyms. Defaults to .60
anto	determines whether psychometric anonyms are returned instead of psychometric synonyms. Defaults to FALSE
diag	additionally return the number of item pairs available for each observation. Useful if dataset contains many missing values.

Author(s)

Richard Yentes <rdyentes@ncsu.edu>, Francisco Wilhelm <franciscowilhelm@gmail.com>

References

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17(3)*, 437-455. doi: 10.1037/a0028085

See Also

psychant for a more concise way to calculate the psychometric antonym score, psychsyn_critval for a helper that allows to set an adequate critical value for the size of the correlation.

Examples

```
synonyms <- psychsyn(careless_dataset, .60)
antonyms <- psychsyn(careless_dataset2, .50, anto = TRUE)
antonyms <- psychant(careless_dataset2, .50)

#with diagnostics
synonyms <- psychsyn(careless_dataset, .60, diag = TRUE)
antonyms <- psychant(careless_dataset2, .50, diag = TRUE)
```

---

psychsyn_critval *Compute the correlations between all possible item pairs and order them by the magnitude of the correlation*

Description

A function intended to help finding adequate critical values for psychsyn and psychant. Takes a matrix of item responses and returns a data frame giving the correlations of all item pairs ordered by the magnitude of the correlation.

Usage

```
psychsyn_critval(x, anto = FALSE)
```

Arguments

x	a matrix of item responses.
anto	ordered by the largest positive correlation, or, if `anto = TRUE`, the largest negative correlation.

Author(s)

Francisco Wilhelm <franciscowilhelm@gmail.com>

See Also

after determining an adequate critical value, continue with psychsyn and/or psychant

Examples

```
psychsyn_cor <- psychsyn_critval(careless_dataset)
psychsyn_cor <- psychsyn_critval(careless_dataset, anto = TRUE)
```

Index