Package 'plasmode'

September 18, 2024

Type Package

Title plasmode: R package for plasmode simulations in the IRT context

Version 1.0

Date 2024-09-18

Author Lena Monsch

Maintainer Lena Monsch <lena.monsch@gmx.de>

Imports mirt, stats

Description Package containing functions around plasmode simulations. It helps creating plasmode datasets as well as investigating their behaviour.

License GPL (>=2)

NeedsCompilation no

R topics documented:

`plasmode-package` *plasmode: R package for plasmode simulations in the IRT context*

Description

Package that contains functions around plasmode simulations in the IRT context. That includes the function `plasmodeData` in order to generate plasmode datasets as well as to investigate the behaviour of plasmode simulations by means of the help function `fnr` as comparison criterion. The latter is always implemented on both, person level with the function `plasmodePerson` and item level with the function `plasmodeItem`.

Details

Package:	plasmode
Type:	Package
Title:	plasmode: R package for plasmode simulations in the IRT context
Version:	1.0
Date:	2024-09-18
Author:	Lena Monsch
Maintainer:	Lena Monsch <lena.monsch@gmx.de>
Imports:	mirt, stats
Description:	Package containing functions around plasmode simulations. It helps creating plasmode datasets as well as
License:	GPL (>=2)

Index of help topics:

Author(s)

Lena Monsch (lena.monsch@gmx.de)

References

Bentler, Peter M. (1990). Comparative fit indixes in structural models. Psychological Bulletin 107.2. doi: https://doi.org/10.1037/0033-2909.107.2.238.

Chalmers, R. Philip (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. Journal of Statistical Software 48.6, pp. 1-29. doi: https://doi.org/10.18637/jss.v048.i06.

Chernick, Michael R. (2008). Bootstrap Methods: A Guide for Practitioners and Researchers. 2nd ed. John Wiley & Sons, Inc. doi: https://doi.org/10.1002/9780470192573.

James, Gareth et al. (2021). An Introduction to Statistical Learning: with Applications in R. 2nd ed. Springer Texts in Statistics. doi: https://doi.org/10.1007/978-1-4614-7138-7.

Maydeu-Olivares, Albert (2013). Goodness-of-Fit Assessment of Item Response Theory Models. Measurement Interdisciplinary Research and Perspectives. doi: http://dx.doi.org/10.1080/15366367.2013.831680.

Maydeu-Olivares, Albert and Joe, Harry (2005). Limited- and Full-Information Estimation and Goodness-of-Fit Testing in 2n Contingency Tables. Journal of the American Statistical Association 100, pp. 1009-1020. doi: https://doi.org/10.1198/016214504000002069.

Nagy, George K. (1997). False negative rate. A misnomer, misunderstood and misused. Acta Cytologica 41.3. doi: https://doi.org/10.1159/000332703.

Orlando, Maria and Thissen, David (2000). Likelihood-Based Item-Fit Indices for Dichotomous Item Response Theory Models. Applied Psychological Measurement 24.1, pp. 50-64. doi: https://doi.org/10.1177/01466216000241003.

van der Linden, Wim J. (2016). Handbook of Item Response Theory. Models. Vol. 1. CRC Press. doi: https://doi.org/10.1201/9781315119144. - 2018. Handbook of Item Response Theory. Applications. Vol. 3. CRC Press. doi: https://doi.org/10.1201/9781315119144.

Wright, Benjamin D. and Masters, Geofferey N. (1982). Rating Scale Analysis. Rasch Measurement. MESA Press. url: https://research.acer.edu.au/measurement/2/.

See Also

fnr plasmodeItem plasmodePerson plasmodeData

`fnr` *Function to Compute the False-Negative-Rate (FNR)*

Description

This function computes the false-negative-rate

$$FNR = \frac{f_n}{f_n + t_p}$$

with values in $[0, 1]$. Therefor, it needs the number of true positives t_p and false negatives f_n. The higher the FNR is, the worse the classification of the truely positive observations proceeds.

Usage

```
fnr(tp, fn)
```

Arguments

`tp`	number of true positives
`fn`	number of false negatives

Details

This function computes the false-negative-rate

$$FNR = \frac{f_n}{f_n + t_p}$$

with values in $[0, 1]$. Therefor, it needs the number of true positives t_p and false negatives f_n. True positives are observations, where true and considered class are consistently positive, while false negatives are observations classified as negative although the actual class is positive. The higher the FNR is, the worse the classification of the truely positive observations proceeds. Thus, it measures the sensitivity of a method. The function serves as comparison criterion for the `plasmodePerson` and `plasmodeItem` functions of the `plasmode-package`.

Value

returns false-negative-rate

$$FNR = \frac{f_n}{f_n + t_p}$$

in $[0, 1]$ of entered input as numeric value. The higher the FNR is, the worse the classification of the truely positive observations proceeds.

Author(s)

Lena Monsch

References

Nagy, George K. (1997). False negative rate. A misnomer, misunderstood and misused. Acta Cytologica 41.3. doi: https://doi.org/10.1159/000332703.

See Also

plasmode-package plasmodeItem plasmodePerson

Examples

```
fnr(tp = 2, fn = 5)
fnr(tp = 0, fn = 5)
fnr(tp = 2, fn = 0)
```

`plasmodeData` *Function That Creates Plasmode Dataset on Person and Item Level in the IRT Context*

Description

This function creates dichotomous plasmode datasets in the IRT context on both, person and item level. It needs an entered dataset so that a 2PL model can be fitted and mismatching observations can consequently be determined. One part of the data is then simulated of the model and the other part sampled of the mismatching observations. The merged plasmode dataset is then returned. The number of needed datasets and the proportion of real to simulated data can be varied.

Usage

```
plasmodeData(data, number = 1, realdata = 0, type = "person")
```

Arguments

`data`	Complete dataframe or matrix of dichotomous data that the analysis is based on. Every column must represent an item and every row needs to contain responses of one participant, respectively.
`number`	number of needed plasmode datasets per setting of real data
`realdata`	Vector that contains the needed proportions of real to simulated data. If `type = "person"`, each vector component corresponds to a percentage of real data in $[0, 1)$. If `type = "item"`, an integer number of real item needs to be entered. If one component is set to 0, a pure simualtion study is performed in both cases.

`type` Either `"person"` or `"item"` in order to merge simulated and real data on person or item level, respectively.

Details

If a plasmode dataset on person level is created, the mismatching values are determined by l_z-values. While the artificial data is simulated of the initial 2PL model, the preparation of the real data is slightly more complex: Therefor, the needed amount of real data is at first sampled with replacement out of the mismatching response patterns. In order to avoid dependence in the data, the response patterns are then sampled in four blocks of the same size of the responses. If the number of items is not a multiple of four, the fourth block has the length of the modulo.

If a plasmode dataset on item level is created, the mismatching items are determined by the S-χ^2 statistic. While the artificial data is simulated of a sampled composition of matching items, the preparation of the real data is slightly more complex: Therefor, the needed amount of real items is at first sampled with replacement out of the mismatching items. In order to avoid dependence in the data, each item value is then sampled of the responses to that item of the five nearest neighbours to the corresponding response pattern.

Value

A list is returned, where each component consists of one dataset. If multiple settings of real data are entered, the needed number of datasets with the first setting is returned in the first `number` components, followed by the second `number` datasets with the second setting and so forth.

Note

The actual number of real observations in the function is the product of entered percentage multiplied to the sample size of the data, which is rounded up to the next integer if necessary. Note as well that the entered data must not contain missing data and needs to be dichotomous. The function is implemented by means of the `mirt`-package. The behaviour of plasmode simulations can be investigated with the functions `plasmodeItem` and `plasmodePerson` on item and person level, respectively.

Author(s)

Lena Monsch

References

Chalmers, R. Philip (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. Journal of Statistical Software 48.6, pp. 1-29. doi: https://doi.org/10.18637/jss.v048.i06.

Chernick, Michael R. (2008). Bootstrap Methods: A Guide for Practitioners and Researchers. 2nd ed. John Wiley & Sons, Inc. doi: https://doi.org/10.1002/9780470192573.

James, Gareth et al. (2021). An Introduction to Statistical Learning: with Applications in R. 2nd ed. Springer Texts in Statistics. doi: https://doi.org/10.1007/978-1-4614-7138-7.

Orlando, Maria and Thissen, David (2000). Likelihood-Based Item-Fit Indices for Dichotomous Item Response Theory Models. Applied Psychological Measurement 24.1, pp. 50-64. doi: https://doi.org/10.1177/01466216000241003.

van der Linden, Wim J. (2016). Handbook of Item Response Theory. Models. Vol. 1. CRC Press. doi: https://doi.org/10.1201/9781315119144. - 2018. Handbook of Item Response Theory. Applications. Vol. 3. CRC Press. doi: https://doi.org/10.1201/9781315119144.

See Also

plasmode-package plasmodeItem plasmodePerson

Examples

```
set.seed(1609)
data <- sample(c(0,1), 9000, replace = TRUE)
dset <- matrix(data, nrow = 1000)
dset <- cbind(dset, matrix(c(rep(1,99),0),nrow=1000))
colnames(dset) <- 1:10

set.seed(1609)
resPers <- plasmodeData(data = dset, number = 1, realdata = 0, type = "person")
set.seed(1609)
resIt <- plasmodeData(data = dset, number = 2, realdata = 2, type = "item")
```

`plasmodeItem` *Function That Computes Plasmode Simulations on Item Level in the IRT Context*

Description

A function that investigates the behaviour of plasmode simulations on item level in the IRT context. Therefor, a 2PL model is fitted of the entered dataset so that the S-χ^2 statistic can classify the items as matching or mismatching to the model. In order to create the plasmode dataset in each iteration, one part of the items is simulated of the matching items and another part is sampled out of the real mismatching items. Since the item fit is then determined by three item fit statistics, the FNR of the classification based on the plasmode dataset is eventually returned. Additionally, the model fit is determined by three model fit statistics in each iteration, which is returned as well together with some fit statistics of the initial dataset.

Usage

```
plasmodeItem(data, iter = 1000, number_realitems, saving = FALSE, comparison = f
```

Arguments

`data`	Complete dataframe or matrix of dichotomous data that the analysis is based on. Every column must represent an item and every row needs to contain responses of one participant, respectively.
`iter`	Number of iterations that is conducted with default of 1000 iterations.
`number_realitems`	Vector with numbers of real items as integer components. If one component is set to 0, a pure simualtion study is performed in this case.
`saving`	If set to `TRUE`, the fit statistics are saved in each iteration in a .RData object. Since that increases the runtime noticeably, it is set to `FALSE` by default.
`comparison`	Function of comparison criterion of the analysis. By default, the false-negative-rate `fnr` is entered, which is also included in this package.

Details

This function investigates the behaviour of plasmode simulations on item level in the IRT context. The function proceeds in the following steps:

(0a) At first, a 2PL model is fitted of the entered dataset.

(0b) Additionally to the model fit statistics mentioned below, the S-χ^2 statistic is calculated based on the initial dataset next. This fit statistic is then chosen to classify the items as matching or mismatching to the model.

(1a) In order to create the plasmode dataset in each iteration, one part of the items is simulated of a sampled composition of matching items.

(1b) The second part of the data consists of real mismatching data. Therefor, the needed amount of real items is at first sampled with replacement out of the mismatching items. In order to avoid dependence in the data, each item value is then sampled of the responses to that item of the five nearest neighbours to the corresponding response pattern.

(1c) The simulated data, considered as truely matching, and the real data, considered as truely mismatching to the model, are then merged to a plasmode dataset.

(2) Another 2PL model is then fitted based on the plasmode dataset.

(3) Since the item and model fit is subsequently determined once more, additionally containing the item-specific infit and outfit statistics, the items can be classified as matching or mismatching again.

(4) The false-negative-rate of the classification based on the plasmode dataset is eventually computed.

(5) It is returned together with the initial fit statistics, the p-value of the model fit statistic M_2, RMSEA and CFI, which are equally determined in each iteration.

Note that only the steps (0a) and (0b) are once performed at the beginning. The other steps need to be performed in every iteration of the loop.

Value

The function returns a list with seven components. The first one is another list containing the S-χ^2 and model fit statistics of the initial dataset. Then, each component incorporates a matrix with `iter` columns and `length(number_realitems)` rows with the results of one fit statistic. Each value specifically corresponds to the false-negative-rate or model fit value of the relevant iteration and percentage of real data.

Note

Note that the entered data must not contain missing data and needs to be dichotomous. The function is implemented by means of the `mirt`-package. In order to create a plasmode dataset, use `plasmodeData`.

Author(s)

Lena Monsch

References

Bentler, Peter M. (1990). Comparative fit indixes in structural models. Psychological Bulletin 107.2. doi: https://doi.org/10.1037/0033-2909.107.2.238.

Chalmers, R. Philip (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. Journal of Statistical Software 48.6, pp. 1-29. doi: https://doi.org/10.18637/jss.v048.i06.

James, Gareth et al. (2021). An Introduction to Statistical Learning: with Applications in R. 2nd ed. Springer Texts in Statistics. doi: https://doi.org/10.1007/978-1-4614-7138-7.

Maydeu-Olivares, Albert (2013). Goodness-of-Fit Assessment of Item Response Theory Models. Measurement Interdisciplinary Research and Perspectives. doi: http://dx.doi.org/10.1080/15366367.2013.831680.

Maydeu-Olivares, Albert and Joe, Harry (2005). Limited- and Full-Information Estimation and Goodness-of-Fit Testing in 2n Contingency Tables. Journal of the American Statistical Association 100, pp. 1009-1020. doi: https://doi.org/10.1198/016214504000002069.

Nagy, George K. (1997). False negative rate. A misnomer, misunderstood and misused. Acta Cytologica 41.3. doi: https://doi.org/10.1159/000332703.

Orlando, Maria and Thissen, David (2000). Likelihood-Based Item-Fit Indices for Dichotomous Item Response Theory Models. Applied Psychological Measurement 24.1, pp. 50-64. doi: https://doi.org/10.1177/01466216000241003.

van der Linden, Wim J. (2016). Handbook of Item Response Theory. Models. Vol. 1. CRC Press. doi: https://doi.org/10.1201/9781315119144.

Wright, Benjamin D. and Masters, Geofferey N. (1982). Rating Scale Analysis. Rasch Measurement. MESA Press. url: https://research.acer.edu.au/measurement/2/.

See Also

plasmode-package fnr plasmodePerson plasmodeData

Examples

```
set.seed(1609)
data <- sample(c(0,1), 9000, replace = TRUE)
dset <- matrix(data, nrow = 1000)
dset <- cbind(dset, matrix(c(rep(1,99),0),nrow=1000))
colnames(dset) <- 1:10

set.seed(1609)
res0 <- plasmodeItem(data = dset, iter = 10, number_realitems = 0)
set.seed(1609)
res2 <- plasmodeItem(data = dset, iter = 10, number_realitems = 2)

par <- par(mfrow = c(1,2))
hist(res0[[2]], freq = FALSE)
hist(res2[[2]], freq = FALSE)
par(par)
```

`plasmodePerson` *Function That Computes Plasmode Simulations on Person Level in the IRT Context*

Description

A function that investigates the behaviour of plasmode simulations on person level in the IRT context. Therefor, a 2PL model is fitted of the entered dataset so that the l_z-values can classify the observations as matching or mismatching to the model. In order to create the plasmode dataset in each iteration, one part of the data is simulated based on the contained items and another part is sampled out of the real mismatching response pattern. Since the person fit is then determined by three person fit statistics, the false-negative-rate of the classification based on the plasmode dataset is eventually returned. Additionally, the model fit is determined by three model fit statistics in each iteration, which is returned as well together with some fit statistics of the initial dataset.

Usage

```
plasmodePerson(data, iter = 1000, percentage_realdata, saving = FALSE, comparisc
```

Arguments

`data`	Complete dataframe or matrix of dichotomous data that the analysis is based on. Every column must represent an item and every row needs to contain responses of one participant, repsectively.
`iter`	Number of iterations that is conducted with default of 1000 iterations.
`percentage_realdata`	Vector with percentages of real data as components, being in $[0, 1)$, respectively. If one component is set to 0, a pure simualtion study is performed in this case.
`saving`	If set to `TRUE`, the fit statistics are saved in each iteration in a .RData object. Since that increases the runtime noticeably, it is set to `FALSE` by default.
`comparison`	Function of comparison criterion of the analysis that needs to be entered. By default, the false-negative-rate `fnr` is entered, which is also included in this package.

Details

This function investigates the behaviour of plasmode simulations on person level in the IRT context. The function proceeds in the following steps:

(0a) At first, a 2PL model is fitted of the entered dataset.

(0b) Additionally to the model fit statistics mentioned below, the l_z-values are calculated based on the initial dataset next. This fit statistic is then chosen to classify the response patterns as matching or mismatching to the model.

(1a) In order to create the plasmode dataset in each iteration, one part of the data is simulated of the contained items.

(1b) The second part of the data consists of real mismatching data. Therefor, the needed amount of real data is at first sampled with replacement out of the mismatching response patterns. In order to avoid dependence in the data, the response patterns are then sampled in four blocks of the same size of the responses. If the number of items is not a multiple of four, the fourth block has the length of the modulo.

(1c) The simulated data, considered as truely matching, and the real data, considered as truely mismatching to the model, are then merged to a plasmode dataset.

(2) Another 2PL model is then fitted based on the plasmode dataset.

(3) Since the person and model fit is subsequently determined once more, additionally including person-specific infit and outfit statistics, the responses can be classified as matching or mismatching again.

(4) The false-negative-rate of the classification based on the plasmode dataset is eventually computed.

(5) It is returned together with the initial fit statistics, the p-value of the model fit statistic M_2, RMSEA and CFI, which are equally determined in each iteration.

Note that only the steps (0a) and (0b) are once performed at the beginning. The other steps need to be performed in every iteration of the loop.

Value

The function returns a list with seven components. The first one is another list containing the l_z-values and model fit statistics of the initial dataset. Then, each component incorporates a matrix with `iter` columns and `length(percentage_realdata)` rows with the results of one fit statistic. Each value specifically corresponds to the false-negative-rate or model fit value of the relevant iteration and percentage of real data.

Note

The actual number of real observations in the function is the product of entered percentage multiplied to the sample size of the data, which is rounded up to the next integer if necessary. Note as well that the entered data must not contain missing data and needs to be dichotomous. The function is implemented by means of the `mirt`-package. In order to create a plasmode dataset, use `plasmodeData`.

Author(s)

Lena Monsch

References

Bentler, Peter M. (1990). Comparative fit indixes in structural models. Psychological Bulletin 107.2. doi: https://doi.org/10.1037/0033-2909.107.2.238.

Chalmers, R. Philip (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. Journal of Statistical Software 48.6, pp. 1-29. doi: https://doi.org/10.18637/jss.v048.i06.

Chernick, Michael R. (2008). Bootstrap Methods: A Guide for Practitioners and Researchers. 2nd ed. John Wiley & Sons, Inc. doi: https://doi.org/10.1002/9780470192573.

Maydeu-Olivares, Albert (2013). Goodness-of-Fit Assessment of Item Response Theory Models. Measurement Interdisciplinary Research and Perspectives. doi: http://dx.doi.org/10.1080/15366367.2013.831680.

Maydeu-Olivares, Albert and Joe, Harry (2005). Limited- and Full-Information Estimation and Goodness-of-Fit Testing in 2n Contingency Tables. Journal of the American Statistical Association 100, pp. 1009-1020. doi: https://doi.org/10.1198/016214504000002069.

Nagy, George K. (1997). False negative rate. A misnomer, misunderstood and misused. Acta Cytologica 41.3. doi: https://doi.org/10.1159/000332703.

Orlando, Maria and Thissen, David (2000). Likelihood-Based Item-Fit Indices for Dichotomous Item Response Theory Models. Applied Psychological Measurement 24.1, pp. 50-64. doi: https://doi.org/10.1177/01466216000241003.

van der Linden, Wim J. (2016). Handbook of Item Response Theory. Models. Vol. 1. CRC Press. doi: https://doi.org/10.1201/9781315119144. - 2018. Handbook of Item Response Theory. Applications. Vol. 3. CRC Press. doi: https://doi.org/10.1201/9781315119144.

See Also

plasmode-package fnr plasmodeItem plasmodeData

Examples

```
set.seed(1609)
data <- sample(c(0,1), 9000, replace = TRUE)
dset <- matrix(data, nrow = 1000)
dset <- cbind(dset, matrix(c(rep(1,99),0),nrow=1000))
colnames(dset) <- 1:10


set.seed(1609)
res0 <- plasmodePerson(data = dset, iter = 10, percentage_realdata = 0)
set.seed(1609)
res5 <- plasmodePerson(data = dset, iter = 10, percentage_realdata = 0.05)

par <- par(mfrow = c(1,2))
hist(res0[[2]], freq = FALSE)
hist(res5[[2]], freq = FALSE)
par(par)
```