Package: msd (via r-universe)

October 31, 2024

Type Package

Title Method of Successive Dichotomizations

Version 0.3.1

Author Chris Bradley <cbradley05@gmail.com>

Maintainer Chris Bradley <cbradley05@gmail.com>

Imports stats

Description Implements the method of successive dichotomizations by Bradley and Massof (2018) <doi:10.1371/journal.pone.0206106>, which estimates item measures, person measures and ordered rating category thresholds given ordinal rating scale data.

License GPL

Encoding UTF-8

LazyData true

NeedsCompilation no

Date/Publication 2021-03-04 01:00:05 UTC

Repository https://cbradl23.r-universe.dev

RemoteUrl https://github.com/cran/msd

RemoteRef HEAD

RemoteSha 42ea97ae1e1d7e643aa289c552c8d92387100711

Contents

`expdata` *Expected Ratings Matrix*

Description

Expected ratings matrix given item measures, person measures and ordered rating category thresholds.

Usage

```
expdata(items, persons, thresholds, minRating)
```

Arguments

`items`	a numeric vector of item measures with missing values set to NA.
`persons`	a numeric vector of person measures with missing values set to NA.
`thresholds`	a numeric vector of ordered rating category thresholds with no NA.
`minRating`	integer representing the smallest ordinal rating category (see Details).

Details

It is assumed that the set of ordinal rating categories consists of all integers from the lowest rating category specified by `minRating` to the highest rating category, which is `minRating + length(thresholds)`.

Value

A numeric matrix of expected ratings.

Note

Expected ratings are literally the expected value of the ordinal rating categories when treated as integers. Expected ratings that cannot be calculated return as NA (e.g., if either the person or item measure is NA). Intended use is for chi-squared tests or for calculating infit and outfit statistics.

Author(s)

Chris Bradley (cbradley05@gmail.com)

See Also

`misfit`

Examples

```
# Using randomly generated values with minimum rating set to zero
im <- runif(20, -2, 2)
pm <- runif(50, -2, 2)
th <- sort(runif(5, -2, 2))
m <- expdata(items = im, persons = pm, thresholds = th, minRating = 0)
```

---

ims *Item Measures*

Description

Estimates item measures assuming person measures are known and all persons use the same set of rating category thresholds.

Usage

```
ims(data, persons, thresholds, misfit = FALSE, minRating = NULL)
```

Arguments

`data`	a numeric matrix of ordinal rating scale data whose entries are integers with missing data set to NA. Rows are persons and columns are items. The ordinal rating scale is assumed to go from the smallest to largest integer in integer steps unless `minRating` is specified (see Details).
`persons`	a numeric vector of person measures with missing values set to NA. The length of `persons` must equal the number of rows in `data`.
`thresholds`	a numeric vector of ordered rating category thresholds with no NA.
`misfit`	logical for calculating infit and outfit statistics. Default is FALSE.
`minRating`	integer representing the smallest ordinal rating category. Default is NULL (see Details).

Details

`minRating` must be specified if either the smallest or largest possible rating category is not in `data` (i.e., no person used one of the extreme rating categories). If `minRating` is specified, the ordinal rating scale is assumed to go from `minRating` to `minRating + length(thresholds)` in integer steps.

Value

A list whose elements are:

`item_measures`	a vector of person measures for each person
`item_std_errors`	a vector of standard errors for the persons
`infit_items`	if `misfit` = TRUE, a vector of infit statistics for the items
`outfit_items`	if `misfit` = TRUE, a vector of outfit statistics for the items

Note

Item measures estimated with `ims` differ from those estimated with `msd` because `ims` assumes all persons use the same rating category thresholds while `msd` does not. Intended use of `ims` is with an anchored set of persons and thresholds. Item measures that cannot be estimated will return as NA (e.g., if all responses to an item consist of only the highest rating category, or of only the lowest rating category, that item's item measure cannot be estimated).

Author(s)

Chris Bradley (cbradley05@gmail.com)

See Also

`msd`

Examples

```
# Simple example with randomly generated values and lowest rating category = 0.
d <- as.numeric(sample(0:4, 500, replace = TRUE))
dm <- matrix(d, nrow = 50, ncol = 10)
pm <- runif(50, -2, 2)
th <- sort(runif(4, -2, 2))
im <- ims(data = dm, persons = pm, thresholds = th, misfit = TRUE, minRating = 0)
```

---

misfit *Infit and Outfit Statistics*

Description

Calculates infit and outfit statistics for items and persons.

Usage

```
misfit(data, items, persons, thresholds, minRating = NULL)
```

Arguments

`data`	a numeric matrix of ordinal rating scale data whose entries are integers with missing data set to NA. Rows are persons and columns are items. The ordinal rating scale is assumed to go from the smallest to largest integer in integer steps unless `minRating` is specified.
`items`	a numeric vector of item measures with missing values set to NA.
`persons`	a numeric vector of person measures with missing values set to NA.
`thresholds`	a numeric vector of ordered rating category thresholds with no NA.
`minRating`	integer representing the smallest ordinal rating category. Default is NULL (see Details).

Details

minRating must be specified if either the smallest or largest possible rating category is not in data (no person used one of the extreme rating categories). If minRating is specified, the ordinal rating scale is assumed to go from minRating to minRating + length(thresholds).

Value

A list whose elements are:

infit_items	a vector of infit statistics for the items
outfit_items	a vector of outfit statistics for the items
infit_persons	a vector of infit statistics for the persons
outfit_persons	a vector of outfit statistics for the persons

Author(s)

Chris Bradley (cbradley05@gmail.com)

Examples

```
# Using randomly generated values
d <- as.numeric(sample(0:5, 500, replace = TRUE))
dm <- matrix(d, nrow = 50, ncol = 10)
im <- runif(10, -2, 2)
pm <- runif(50, -2, 2)
th <- sort(runif(5, -2, 2))
m <- misfit(data = dm, items = im, persons = pm, thresholds = th)

# If the lowest or highest rating category is not in \code{data}, specify \code{minRating}
dm[dm == 0] <- NA
m2 <- misfit(data = dm, items = im, persons = pm, thresholds = th, minRating = 0)
```

---

msd *Method of Successive Dichotomizations*

Description

Estimates item measures, person measures, rating category thresholds and their standard errors using the method of successive dichotomizations. Option provided for anchoring certain items and persons while estimating the rest. Option also provided for estimating infit and outfit statistics.

Usage

```
msd(data, items = NULL, persons = NULL, misfit = FALSE)
```

Arguments

data	a numeric matrix of ordinal rating scale data whose entries are integers with missing data set to NA. Rows are persons and columns are items. The ordinal rating scale is assumed to go from the smallest integer to the largest integer in data in integer steps.
items	a numeric vector of anchored item measures. Item measures to be estimated are set to NA. Default is NULL (see Details).
persons	a numeric vector of anchored person measures. Person measures to be estimated are set to NA. Default is NULL (see Details).
misfit	logical for calculating infit and outfit statistics. Default is FALSE.

Details

items and persons are optional numeric vectors that specify item and person measures that are "anchored" and not estimated. The length of items must equal the number of columns in data and the length of persons must equal the number of rows in data. Only entries set to NA in items and persons are estimated. Default for both items and persons is NULL, which is equivalent to a vector of NA so that all items and persons are estimated.

Value

A list whose elements are:

item_measures	a vector of item measures for each item
person_measures	a vector of person measures for each person
thresholds	a vector of average rating category thresholds used by the persons when rating the items
item_std_errors	a vector of standard errors for the items
person_std_errors	a vector of standard errors for the persons
threshold_std_errors	a vector of standard errors for the thresholds
item_reliability	reliability of the item measures
person_reliability	reliability of the person measures
infit_items	if misfit = TRUE, a vector of infit statistics for the items
outfit_items	if misfit = TRUE, a vector of outfit statistics for the items
infit_persons	if misfit = TRUE, a vector of infit statistics for the persons
outfit_persons	if misfit = TRUE, a vector of outfit statistics for the persons

Note

The axis origin is set by convention at the mean item measure. All item measures and person measures that cannot be estimated will return as NA (e.g., if a person responds with only the highest rating category, or with only the lowest rating category, to all items, that person's person measure cannot be estimated).

The accuracy of `msd` can be tested using the `simdata` function (see Examples).

Author(s)

Chris Bradley (cbradley05@gmail.com)

References

Bradley, C. and Massof, R. W. (2018) Method of successive dichotomizations: An improved method for estimating measures of latent variables from rating scale data. *PLoS One*, **13(10)** doi:10.1371/journal.pone.0206106

See Also

`simdata`

Examples

```
# Simple example using a randomly generated ratings matrix
d <- as.numeric(sample(0:5, 200, replace = TRUE))
dm <- matrix(d, nrow = 20, ncol = 10)
m1 <- msd(dm, misfit = TRUE)

# Anchor first 5 item measures and first 10 person measures
im <- m1$item_measures
im[6:length(im)] <- NA
pm <- m1$person_measures
pm[11:length(pm)] <- NA
m2 <- msd(dm, items = im, persons = pm)

# To test the accuracy of msd using simdata, set the mean item measure to zero
# (axis origin in msd is the mean item measure) and the mean threshold to
# zero (any non-zero mean threshold is reflected in the person measures).
im <- runif(100, -2, 2)
im <- im - mean(im)
pm <- runif(100, -2, 2)
th <- sort(runif(5, -2, 2))
th <- th - mean(th)
d <- simdata(im, pm, th, missingProb = 0.15, minRating = 0)
m <- msd(d)

# Compare msd parameters to true values.  Linear regression should
# yield a slope very close to 1 and an intercept very close to 0.
lm(m$item_measures ~ im)
lm(m$person_measures ~ pm)
lm(m$thresholds ~ th)
```

`msdprob` *Rating Category Probabilities*

Description

Estimates the probability of observing each rating category given a set of ordered rating category thresholds.

Usage

```
msdprob(x, thresholds)
```

Arguments

`x`	a real number or a vector of real numbers with no NA representing a set of person minus item measures.
`thresholds`	a numeric vector of ordered rating category thresholds with no NA.

Details

It is assumed that `thresholds` partitions the real line into `length(thresholds)+1` ordered intervals that represent the rating categories.

Value

A matrix of probabilities where each of the `length(thresholds)+1` rows represents a different rating category (lowest rating category is the top row) and each of the `length(x)` columns represents a different person minus item measure.

Note

`msdprob` can be used to create probability curves, which represent the probability of rating an item with each rating category as a function of the person measure minus item measure (see Examples).

Author(s)

Chris Bradley (cbradley05@gmail.com)

Examples

```
# Simple example
p <- msdprob(c(1.4, -2.2), thresholds = c(-1.1, -0.3, 0.5, 1.7, 2.2))

# Plot probability curves — each curve represents the probability of
# rating an item with a given rating category as a function of the
# person measure minus item measure.
x <- seq(-6, 6, 0.1)
p <- msdprob(x, thresholds = c(-3.2, -1.4, 0.5, 1.7, 3.5))
```

```
plot(0, 0, xlim = c(-6, 6), ylim = c(0, 1), type = "n",
    xlab = "Person minus item measure", ylab = "Probability")
for (i in seq(1, dim(p)[1])){
  lines(x, p[i,], type = "l", lwd = "2" , col = rainbow(6)[i])
}
```

---

pms *Person Measures*

Description

Estimates person measures assuming item measures are known and all persons use the same set of rating category thresholds.

Usage

```
pms(data, items, thresholds, misfit = FALSE, minRating = NULL)
```

Arguments

data	a numeric matrix of ordinal rating scale data whose entries are integers with missing data set to NA. Rows are persons and columns are items. The ordinal rating scale is assumed to go from the smallest to largest integer in integer steps unless minRating is specified (see Details).
items	a numeric vector of item measures with missing values set to NA. The length of items must equal the number of columns in data.
thresholds	a numeric vector of ordered rating category thresholds with no NA.
misfit	logical for calculating infit and outfit statistics. Default is FALSE.
minRating	integer representing the smallest ordinal rating category. Default is NULL (see Details).

Details

minRating must be specified if either the smallest or largest possible rating category is not in data (i.e., no person used one of the extreme rating categories). If minRating is specified, the ordinal rating scale is assumed to go from minRating to minRating + length(thresholds) in integer steps.

Value

A list whose elements are:

person_measures	a vector of person measures for each person
person_std_errors	a vector of standard errors for the persons
infit_persons	if misfit = TRUE, a vector of infit statistics for the persons
outfit_persons	if misfit = TRUE, a vector of outfit statistics for the persons

Note

Person measures estimated with pms differ from those estimated with msd because pms assumes all persons use the same rating category thresholds while msd does not. Intended use of pms is with an anchored set of items and thresholds. Person measures that cannot be estimated will return as NA (e.g., if a person responds to all items with only the highest rating category, or with only the lowest rating category, that person's person measure cannot be estimated).

Author(s)

Chris Bradley (cbradley05@gmail.com)

See Also

msd

Examples

```
# Simple example with randomly generated values and lowest rating category = 0
d <- as.numeric(sample(0:4, 500, replace = TRUE))
dm <- matrix(d, nrow = 25, ncol = 20)
im <- runif(20, -2, 2)
th <- sort(runif(4, -2, 2))
pm <- pms(data = dm, items = im, thresholds = th, misfit = TRUE, minRating = 0)
```

---

rasch *Dichotomous Rasch Model*

Description

Estimates item measures, person measures and their standard errors using the dichotomous Rasch model. A special case of the function msd when the rating scale consists of only two rating categories: 0 and 1. Option provided for anchoring certain items and persons while estimating the rest. Option also provided for estimating infit and outfit statistics.

Usage

```
rasch(data, items = NULL, persons = NULL, misfit = FALSE)
```

Arguments

data	a numeric matrix of 0's and 1's with missing data set to NA. Rows are persons and columns are items.
items	a numeric vector of anchored item measures. Item measures to be estimated are set to NA. Default is NULL (see Details).
persons	a numeric vector of anchored person measures. Person measures to be estimated are set to NA. Default is NULL (see Details).
misfit	logical for calculating infit and outfit statistics. Default is FALSE.

Details

items and persons are optional numeric vectors that specify item and person measures that should be "anchored" and not estimated. The length of items must equal the number of columns in data and the length of persons must equal the number of rows in data. Only entries set to NA in items and persons are estimated. Default for both items and persons is NULL, which is equivalent to a vector of NA so that all items and persons are estimated.

Value

A list whose elements are:

item_measures a vector of item measures for each item

person_measures
: a vector of person measures for each person

item_std_errors
: a vector of standard errors for the items

person_std_errors
: a vector of standard errors for the persons

item_reliability
: reliability value for the items

person_reliability
: reliability value for the persons

infit_items if misfit = TRUE, a vector of infit statistics for the items

outfit_items if misfit = TRUE, a vector of outfit statistics for the items

infit_persons if misfit = TRUE, a vector of infit statistics for the persons

outfit_persons if misfit = TRUE, a vector of outfit statistics for the persons

Note

The axis origin is set by convention at the mean item measure. All item measures and person measures that cannot be estimated will return as NA (e.g., if a person responds with a single rating category to all items, that person's person measure cannot be estimated).

rasch is the basis for the "successive dichotomizations" in msd and is repeatedly called by msd when there are three or more rating categories.

The accuracy of rasch can be tested using the simdata function (see Examples).

Author(s)

Chris Bradley (cbradley05@gmail.com)

See Also

msd simdata

Examples

```
# Simple example using a randomly generated ratings matrix
d <- as.numeric(sample(0:1, 200, replace = TRUE))
dm <- matrix(d, nrow = 20, ncol = 10)
m1 <- rasch(dm, misfit = TRUE)

# Anchor first 5 item measures and first 10 person measures
im <- m1$item_measures
im[6:length(im)] <- NA
pm <- m1$person_measures
pm[11:length(pm)] <- NA
m2 <- rasch(dm, items = im, persons = pm)

# To test the accuracy of rasch using simdata, set the true mean item measure to
# zero (axis origin in rasch is the mean item measure).  Note that the threshold for
# dichotomous data is at 0.
im <- runif(100, -2, 2)
im <- im - mean(im)
pm <- runif(100, -2, 2)
th <- 0
d <- simdata(im, pm, th, missingProb = 0.15, minRating = 0)
m <- rasch(d)

# Compare rasch parameters to true values.  Linear regression should
# yield a slope very close to 1 and an intercept very close to 0.
lm(m$item_measures ~ im)
lm(m$person_measures ~ pm)
```

---

simdata *Simulated Rating Scale Data*

Description

Generates simulated rating scale data given item measures, person measures and rating category thresholds.

Usage

```
simdata(items, persons, thresholds, missingProb = 0, minRating = 0)
```

Arguments

items	a numeric vector of item measures with no NA.
persons	a numeric vector of person measures with no NA.
thresholds	a numeric vector of ordered rating category thresholds with no NA.
missingProb	a number between 0 and 1 specifying the probability of missing data.
minRating	integer representing the smallest ordinal rating category. Default is 0 (see Details).

Details

It is assumed that the set of ordinal rating categories consists of all integers from the lowest rating category specified by `minRating` to the highest rating category, which is `minRating + length(thresholds)`.

Value

A numeric matrix of simulated rating scale data.

Note

`simdata` can be used to test the accuracy of `msd` (see Examples).

Author(s)

Chris Bradley (cbradley05@gmail.com)

See Also

`msd`

Examples

```
# Use simdata to test the accuracy of msd. First, randomly generate item
# measures, person measures and thresholds with 15 percent missing data and
# ordinal rating categories from 0 to 5. Then, set mean item measure to zero
# (axis origin in msd is the mean item measure) and mean threshold to zero
# (any non-zero mean threshold is reflected in the person measures).
im <- runif(100, -2, 2)
pm <- runif(100, -2, 2)
th <- sort(runif(5, -2, 2))
im <- im - mean(im)
th <- th - mean(th)
d <- simdata(im, pm, th, missingProb = 0.15, minRating = 0)
m <- msd(d)

# Compare msd parameters to true values.  Linear regression should
# yield a slope very close to 1 and an intercept very close to 0.
lm(m$item_measures ~ im)
lm(m$person_measures ~ pm)
lm(m$thresholds ~ th)
```

`thresh` *Rating Category Thresholds*

Description

Estimates rating category thresholds for `msd` given rating scale data, item measures and person measures.

Usage

```
thresh(data, items, persons)
```

Arguments

data	a numeric matrix of ordinal rating scale data whose entries are integers with missing data set to NA. Rows are persons and columns are items. The ordinal rating scale is assumed to go from the smallest integer to the largest integer in data in integer steps.
items	a numeric vector of item measures with missing values set to NA (see Details).
persons	a numeric vector of person measures with missing values set to NA (see Details).

Details

The length of items must equal the number of columns in data and the length of persons must equal the number of rows in data. Neither items nor persons can consist of only NA.

Value

A list whose elements are:

thresholds	a vector of average rating category thresholds used by the persons when rating the items
threshold_std_errors	a vector of standard errors for the thresholds

Note

thresh is a special case of msd when item measures and person measures are known.

Author(s)

Chris Bradley (cbradley05@gmail.com)

See Also

msd

Examples

```
# Using randomly generated values
d <- as.numeric(sample(0:5, 1000, replace = TRUE))
m <- matrix(d, nrow = 50, ncol = 20)
im <- runif(20, -2, 2)
pm <- runif(50, -2, 2)
th1 <- thresh(m, items = im, persons = pm)

# Anchor first 10 item measures and first 10 person measures
im[11:length(im)] <- NA
pm[11:length(pm)] <- NA
th2 <- thresh(m, items = im, persons = pm)
```

Index