Package 'repsd'

April 18, 2023

Title Root Expected Proportion Squared Difference for Detecting DIF

Version 1.0.1

Description Root Expected Proportion Squared Difference (REPSD) is a nonparametric differential item functioning (DIF) method that (a) allows practitioners to explore for DIF related to small, fine-grained focal groups of examinees, and (b) compares the focal group directly to the composite group that will be used to develop the reported test score scale. Using your provided response matrix with a column that identifies focal group membership, this package provides the REPSD values, a simulated null distribution of possible REPSD values, and the simulated p-values identifying items possibly displaying DIF without requiring enormous sample sizes.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.2.3

Depends R (>= 3.5.0), progress

Suggests colorDF, testthat (>= 3.0.0)

LazyData true

Config/testthat/edition 3

NeedsCompilation no

Author Anne Corrine Huggins-Manley [aut], Anthony William Raborn [aut, cre] (<https://orcid.org/0000-0002-8083-4739>)

Maintainer Anthony William Raborn <anthony.w.raborn@gmail.com>

Repository CRAN

Date/Publication 2023-04-18 07:10:04 UTC

R topics documented:

`estimate_impact` *Estimate the effect size difference between focal and composite group abilities*

Description

Estimate the effect size difference between focal and composite group abilities

Usage

```
estimate_impact(responses = timmsData, focal_column = 21, focal_id = 1)
```

Arguments

`responses`	The `data.frame` of responses, including the `focal_column`.
`focal_column`	The `numeric` location of the focal column.
`focal_id`	The `numeric`, `character`, or `logical` value that identifies the focal group.

Value

A `numeric` estimate of the impact as the effect size D, e.g., the standardized mean theta difference between the focal group and the composite (total) group abilities. This estimate is rounded to 3 decimal places.

`null_repsd` *null_repsd*

Description

null_repsd

Usage

```
null_repsd(
  item_count = 20,
  focal_sample = 88,
  focal_prop = 0.09,
  numStrata = 4,
  impact = estimate_impact(),
  item_params_a = timmsDiscrim,
  item_params_b = timmsDiffic,
  anchorItems = NULL,
  iterations = 10000,
  verbose = TRUE
)
```

Arguments

`item_count`	numeric. How many items?
`focal_sample`	numeric. How large is the focal sample?
`focal_prop`	numeric, between 0 and 1 (exclusive). What is the proportion of the focal sample compared to the rest of the data?
`numStrata`	numeric. How many strata for matching should be used?
`impact`	numeric. What is the expected, standardized mean difference between the focal group's mean theta and the composite group's mean theta (i.e., standardized focal mean - composite mean). See details for further explanation.
`item_params_a`	numeric vector. What are the discrimination parameters of the items in the data set?
`item_params_b`	numeric vector. What are the difficulty parameters of the items in the data set?
`anchorItems`	either `NULL` or a vector of the anchorItems names or numeric column locations. If `NULL`, all items are used for calculating the total test score for stratifying individuals. If a vector, the specified items are used to calculate the total test score for stratifying individuals.
`iterations`	numeric. How many iterations for the function to run? Defaults to 10000.
`verbose`	logical. If `TRUE` (default), prints a `progress::progress_bar()` in the console to allow tracking of the state of the distribution generation.

Value

An `item_count` x `iterations` data.frame with simulated repsd values for each item.

plot_repsd *REPSD Null vs Observed Histogram*

Description

REPSD Null vs Observed Histogram

Usage

```
plot_repsd(repsd_values, null_values, pvalues, which_item, bins = 30)
```

Arguments

repsd_values	A numerical vector of repsd values, the output of repsd()$repsd_each_item.
null_values	A matrix of the repsd null distribution, the output of null_repsd().
pvalues	A numerical vector of the repds p-values, the output of repsd_pval()$p.value
which_item	A numerical indicator of the specific item to plot.
bins	A numerical indicator on the number of bins to output in the histogram.

Value

A plot of the REPSD null distribution for the indicated item with the observed REPSD value as a red line and the observed p-value

Examples

```
example_repsd <-
    repsd()
example_null <-
    null_repsd(iterations = 100)
example_pvals <-
    repsd_pval(
                alpha = .05,
                null_dist = example_null,
                items_repsd = example_repsd$repsd_each_item
                )
# Only one plot
plot_repsd(repsd_values = example_repsd$repsd_each_item,
           null_values = example_null,
           pvalues = example_pvals$p.value,
           which_item = 18,
           bins = 10)
# Multiple plots on the same plot
oldpar <- par()
par(mfrow = c(2,2))
for (i in c(1,8,16,18)) {
  plot_repsd(
             repsd_values = example_repsd$repsd_each_item,
```

```
              null_values = example_null,
              pvalues = example_pvals$p.value,
              which_item = 18,
              bins = 10
              )
}
par(mfrow = oldpar$mfrow)
```

---

repsd *repsd*

Description

repsd

Usage

```
repsd(
  responses = timmsData,
  focalColumn = 21,
  focalGroupID = 1,
  anchorItems = NULL,
  numStrata = 4
)
```

Arguments

responses	data.frame, matrix, or similar object which includes the item responses and the focal group ID column.
focalColumn	numeric or character. The location or name of the column that holds the focal group data.
focalGroupID	numeric or character. The value that identifies the focal group.
anchorItems	either NULL or a vector of the anchorItems names or numeric column locations. If NULL, all items are used for calculating the total test score for stratifying individuals. If a vector, the specified items are used to calculate the total test score for stratifying individuals.
numStrata	numeric. How many strata for matching should be used?

Value

Matrix of repsd values for each item.

repsd_pval *Calculating p-values for repsd*

Description

Calculating p-values for repsd

Usage

```
repsd_pval(
  alpha = 0.05,
  null_dist = null_repsd(),
  items_repsd = repsd()$repsd_each_item,
  responses = timmsData,
  focalColumn = 21,
  verbose = TRUE
)
```

Arguments

alpha	numeric. The alpha level to calculate significance.
null_dist	A data.frame-type object with the null distribution simulation for each item as columns.
items_repsd	A numeric vector of the repsd values for each item.
responses	The data.frame of item responses and the focal column.
focalColumn	The column number for the focal column. Removed from the final data.
verbose	Logical. Do you want to print the results to console (TRUE, default) or return the results invisibly (FALSE)?

Details

Calculates the p-values for repsd for the data set. It can be used as a wrapper function by providing the null_repsd() function and the repsd_each_item output of the repsd() function (each with proper arguments) as the arguments to null_dist and items_repsd, respectively.

Value

If the colorDF package is installed and accessible, a colorDF with the significant items highlighted. Otherwise, a data.frame. Both have columns with the items names, the repsd value, the p.value, and the sig (0 = false, 1 = true) for each item.

`timmsData` *Sample data from TIMMS*

Description

Dataset including 977 observations on 20 items and 1 group identifying variable.

Usage

```
timmsData
```

Format

A data frame with 977 rows and 21 columns:

MA13011 0 (incorrect) or 1 (correct) response on this math item

MA13012 0 (incorrect) or 1 (correct) response on this math item

MA13013 0 (incorrect) or 1 (correct) response on this math item

MA13015 0 (incorrect) or 1 (correct) response on this math item

MA13016 0 (incorrect) or 1 (correct) response on this math item

MA13017 0 (incorrect) or 1 (correct) response on this math item

MA13018 0 (incorrect) or 1 (correct) response on this math item

MA33086 0 (incorrect) or 1 (correct) response on this math item

MA33225C 0 (incorrect) or 1 (correct) response on this math item

MA33225E 0 (incorrect) or 1 (correct) response on this math item

MA33142 0 (incorrect) or 1 (correct) response on this math item

MA33044 0 (incorrect) or 1 (correct) response on this math item

MA33179 0 (incorrect) or 1 (correct) response on this math item

MA33076 0 (incorrect) or 1 (correct) response on this math item

MA33140 0 (incorrect) or 1 (correct) response on this math item

MA33007 0 (incorrect) or 1 (correct) response on this math item

MA33214 0 (incorrect) or 1 (correct) response on this math item

MA33171 0 (incorrect) or 1 (correct) response on this math item

MA33039 0 (incorrect) or 1 (correct) response on this math item

MA33180 0 (incorrect) or 1 (correct) response on this math item

middle_school_or_lower_for_parents_highest_ed 0 (higher than middle school) or 1 (middle school or lower) indicator for parents' highest education level

timmsDiffic *Sample TIMMS item difficulties*

Description

A vector of the 20 item difficulty parameters b for the timmsData items.

Usage

```
timmsDiffic
```

Format

An object of class numeric of length 20.

timmsDiscrim *Sample TIMMS item discriminations*

Description

A vector of the 20 item discrimination parameters a for the timmsData items.

Usage

```
timmsDiscrim
```

Format

An object of class numeric of length 20.

Index