Package 'Counternull'

October 12, 2022

Type Package

Title Creating Null and Counternull Distributions to Find Counternull Values

Version 0.1.0

Description Calculates the difference in average change over time for variables in given dataset. Generates a randomization matrix to resample data for permutation testing. Creates and plots null distributions and calculates P-Values. Identifies potential counternull values by generating and plotting counternull distributions.
Rosenthal and Rubin (1994) <doi:10.1111/j.1467-9280.1994.tb00281.x>.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Depends R (>= 2.10)

RoxygenNote 7.1.1

Imports stats, effsize, graphics

URL https://github.com/ymabene/Counternull

BugReports https://github.com/ymabene/Counternull/issues

NeedsCompilation no

Author Mabene Yasmine [aut, cre],
Bind Marie [aut],
Harvard University [cph]

Maintainer Mabene Yasmine <ymabene@stanford.edu>

Repository CRAN

Date/Publication 2021-12-17 19:42:13 UTC

R topics documented:

`create_counternull_distribution` *Creates a counternull distribution for a given value*

Description

Resamples data to create counternull distribution. Calculate and prints P-value. Returns vector with counternull distribution data points (test statistics created from resampling). Observed test statistic is indicated in distribution using dashed black line. No effect is indicated with gray dashed line. Counternull value is indicated with red dashed line.

Usage

```
create_counternull_distribution(
  sample_data,
  extreme,
  rand_matrix,
  permutation_counter_function,
  counternull_value,
  test_stat,
  variable,
  iterations,
  pairs
)
```

Arguments

sample_data	Sample data set. Data should have column indicating exposure (1) or non exposure (0) for each group (row) that is measured. Each measured outcome (variable) should be represented by an additional column.
extreme	Direction which is defined as more extreme for test statistic in distribution (0 for less or 1 for greater)
rand_matrix	Matrix with unique randomizations for exposure assignment
permutation_counter_function	Function used to create permutation vector for counternull distribution
counternull_value	Number to test out as counternull value
test_stat	Observed test statistic.
variable	Variable measured Format: sample_data$column
iterations	Number of unique arrangements of exposure assignments used to generate distribution (At most the number of rows in rand_matrix)
pairs	Number of pairs of units there are to measure in dataset (One pair = control unit + experimental unit)

Value

Counternull test statistics (Vector)

Examples

```
create_counternull_distribution(sample_district_1DS,0,rand_matrix_1DS,
permutation_counter_diff_means,
-3323,find_test_stat_diff_means(sample_district_1DS,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre),
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre,128,7)

create_counternull_distribution(sample_district_1DS,0,rand_matrix_1DS,
permutation_counter_t,
-3127,find_test_stat_t(sample_district_1DS,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre),
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre,128,7)

create_counternull_distribution(sample_district_1DS,0,rand_matrix_1DS,
permutation_counter_paired_t,
-3127,find_test_stat_paired_t(sample_district_1DS,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre),
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre,128,7)
```

`create_null_distribution` *Creates a null distribution*

Description

Resamples data to create null distribution. Calculates and prints observed test statistic and P-value. Returns vector with null distribution data points (test statistics created from resampling). Observed test statistic is indicated in null distribution using dashed black line.

Usage

```
create_null_distribution(
  sample_data,
  extreme,
  rand_matrix,
  permutation_null_function,
  test_stat,
  variable,
  iterations
)
```

Arguments

`sample_data`	Sample data set. Data should have column indicating exposure (1) or non exposure (0) for each group (row) that is measured. Each measured outcome (variable) should be represented by an additional column.
`extreme`	Direction which is defined as more extreme for test statistic in distribution (0 for less or 1 for greater)
`rand_matrix`	Matrix with unique randomizations for exposure assignment
`permutation_null_function`	Function used to create permutation vector for null distribution
`test_stat`	Observed test statistic.
`variable`	Variable measured Format: sample_data$column
`iterations`	Number of unique arrangements of exposure assignments used to generate distribution (At most the number of rows in rand_matrix)

Value

Vector with all generated test statistics in null distribution

Examples

```
create_null_distribution(sample_district_1DS,0,rand_matrix_1DS,
permutation_null_diff_means,find_test_stat_diff_means(sample_district_1DS,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre),
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre,128)

create_null_distribution(sample_district_1DS,0,rand_matrix_1DS,
permutation_null_cohens_d,find_test_stat_cohens_d(sample_district_1DS,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre),
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre,128)

create_null_distribution(sample_district_1DS,0,rand_matrix_1DS,
permutation_null_t,find_test_stat_t(sample_district_1DS,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre),
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre,128)

create_null_distribution(sample_district_1D,0,rand_matrix_1D,
permutation_null_diff_means,find_test_stat_diff_means(sample_district_1D,
sample_district_1D$charge_prosecuted_1000_rate_post -
sample_district_1D$charge_prosecuted_1000_rate_pre),
sample_district_1D$charge_prosecuted_1000_rate_post -
sample_district_1D$charge_prosecuted_1000_rate_pre,10000)
```

`find_counternull_values` *Finds counternull values*

Description

Finds and prints full range of counternull values, the test statistic and p-value along with null and counternull distribution if counternull values are found. Otherwise only null distribution is displayed. Observed test statistic is indicated in distribution using dashed black line. No effect is indicated with gray dashed line. Counternull value is indicated with red dashed line. Counternull values are returned if found. 0 is returned otherwise.

Usage

```
find_counternull_values(
  obs_pval,
  sample_data,
  extreme,
  rand_matrix,
  permutation_null_function,
```

```
  permutation_counter_function,
  low,
  high,
  test_stat,
  variable,
  iterations,
  pairs
)
```

Arguments

obs_pval	P-value from null distribution
sample_data	Sample data set. Data should have column indicating exposure (1) or non exposure (0) for each group (row) that is measured. Each measured outcome (variable) should be represented by an additional column.
extreme	Direction which is defined as more extreme for test statistic in distribution (0 for less or 1 for greater)
rand_matrix	Matrix with all possible randomizations of exposure assignment
permutation_null_function	Function used to create permutation vector for null distribution
permutation_counter_function	Function used to create permutation vector for counternull distribution
low	Lower bound of counternull value search
high	Upper bound of counternull value search
test_stat	Observed test statistic. (You can use built in functions to find various test statistics in given dataset)
variable	Variable measured. Format: sample_data$column
iterations	Number of unique arrangements of exposure assignments used to generate distribution (At most the number of rows in rand_matrix)
pairs	Number of pairs of units there are to measure in dataset (One pair = control unit + experimental unit)

Value

Vector of Counternull Values (Numeric 0 if none are found)

Examples

```
find_counternull_values(.375,sample_district_1DS,0,rand_matrix_1DS,
permutation_null_diff_means,permutation_counter_diff_means,
-8000,0, find_test_stat_diff_means(sample_district_1DS,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre),
sample_district_1DS$charge_prosecuted_1000_rate_post-
sample_district_1DS$charge_prosecuted_1000_rate_pre,128,7)
```

```
find_counternull_values(.375,sample_district_1DS,0,rand_matrix_1DS,
permutation_null_t,permutation_counter_t,
-8000,0, find_test_stat_t(sample_district_1DS,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre),
sample_district_1DS$charge_prosecuted_1000_rate_post-
sample_district_1DS$charge_prosecuted_1000_rate_pre,128,7)

find_counternull_values(.375,sample_district_1DS,0,rand_matrix_1DS,
permutation_null_paired_t,permutation_counter_paired_t,
-8000,0, find_test_stat_paired_t(sample_district_1DS,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre),
sample_district_1DS$charge_prosecuted_1000_rate_post-
sample_district_1DS$charge_prosecuted_1000_rate_pre,128,7)
```

`find_test_stat_cohens_d` *Finds Cohen's D (observed test statistic) for data set*

Description

Finds Cohen's D test statistic for experimental (exposed) and control (non exposed) group for measured outcome in dataset

Usage

```
find_test_stat_cohens_d(sample_data, variable)
```

Arguments

`sample_data`	Sample data set. Data should have column indicating exposure (1) or non exposure (0) for each group (row) that is measured. Each measured outcome (variable) should be represented by an additional column.
`variable`	Variable measured Format: sample_data$column

Value

Observed Test Statistic (Numeric)

Examples

```
find_test_stat_cohens_d(sample_district_1DS,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre)

find_test_stat_cohens_d(sample_district_1D,
sample_district_1D$charge_prosecuted_1000_rate_post -
sample_district_1D$charge_prosecuted_1000_rate_pre)
```

find_test_stat_diff_means

Finds difference of means (observed test statistic) for data set

Description

Finds difference in the average change over time between experimental (exposed) and control (non exposed) group for measured outcome in dataset

Usage

```
find_test_stat_diff_means(sample_data, variable)
```

Arguments

sample_data	Sample data set. Data should have column indicating exposure (1) or non exposure (0) for each group (row) that is measured. Each measured outcome (variable) should be represented by an additional column.
variable	Variable measured Format: sample_data$column

Value

Observed Test Statistic (Numeric)

Examples

```
find_test_stat_diff_means(sample_district_1DS,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre)

find_test_stat_diff_means(sample_district_1D,
sample_district_1D$charge_prosecuted_1000_rate_post -
sample_district_1D$charge_prosecuted_1000_rate_pre)
```

find_test_stat_paired_t

Finds paired t statistic (observed test statistic) for data set

Description

Finds paired t statistic between experimental (exposed) and control (non exposed) group for measured outcome in dataset

Usage

```
find_test_stat_paired_t(sample_data, variable)
```

Arguments

sample_data	Sample data set. Data should have column indicating exposure (1) or non exposure (0) for each group (row) that is measured. Each measured outcome (variable) should be represented by an additional column.
variable	Variable measured Format: sample_data$column

Value

Observed Test Statistic (Numeric)

Examples

```
find_test_stat_paired_t(sample_district_1DS,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre)

find_test_stat_paired_t(sample_district_1D,
sample_district_1D$charge_prosecuted_1000_rate_post -
sample_district_1D$charge_prosecuted_1000_rate_pre)
```

---

find_test_stat_t *Finds t statistic (observed test statistic) for data set*

Description

Finds t statistic between experimental (exposed) and control (non exposed) group for measured outcome in dataset

Usage

```
find_test_stat_t(sample_data, variable)
```

Arguments

sample_data	Sample data set. Data should have column indicating exposure (1) or non exposure (0) for each group (row) that is measured. Each measured outcome (variable) should be represented by an additional column.
variable	Variable measured Format: sample_data$column

Value

Observed Test Statistic (Numeric)

Examples

```
find_test_stat_t(sample_district_1DS,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre)

find_test_stat_t(sample_district_1D,
sample_district_1D$charge_prosecuted_1000_rate_post -
sample_district_1D$charge_prosecuted_1000_rate_pre)
```

---

permutation_counter_cohens_d

Creates Cohen's D permutation vector for counternull distribution

---

Description

Resamples data to create counternull distribution. Returns vector with test statistics in counternull distribution.

Usage

```
permutation_counter_cohens_d(
  sample_data,
  rand_matrix,
  counternull_value,
  variable,
  iterations,
  pairs
)
```

Arguments

sample_data	Sample data set. Data should have column indicating exposure (1) or non exposure (0) for each group (row) that is measured. Each measured outcome (variable) should be represented by an additional column.
rand_matrix	Matrix with unique randomizations for exposure assignment
counternull_value	Number to test out as counternull value
variable	Variable measured Format: sample_data$column
iterations	Numbers of unique arrangements of exposure assignments used to generate distribution (At most the number of rows in rand_matrix)
pairs	Number of pairs of units there are to measure in dataset (One pair = control unit + experimental unit)

Value

Vector with all generated test statistics in counternull distribution

Examples

```
permutation_counter_cohens_d(sample_district_1DS, rand_matrix_1DS, -3323,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre,128,7)
```

`permutation_counter_diff_means` *Creates difference of means permutation vector for counternull distribution*

Description

Resamples data to create counternull distribution. Returns vector with test statistics in counternull distribution.

Usage

```
permutation_counter_diff_means(
  sample_data,
  rand_matrix,
  counternull_value,
  variable,
  iterations,
  pairs
)
```

Arguments

`sample_data`	Sample data set. Data should have column indicating exposure (1) or non exposure (0) for each group (row) that is measured. Each measured outcome (variable) should be represented by an additional column.
`rand_matrix`	Matrix with unique randomizations for exposure assignment
`counternull_value`	Number to test out as counternull value
`variable`	Variable measured Format: sample_data$column
`iterations`	Numbers of unique arrangements of exposure assignments used to generate distribution (At most the number of rows in rand_matrix)
`pairs`	Number of pairs of units there are to measure in dataset (One pair = control unit + experimental unit)

Value

Vector with all generated test statistics in null distribution

Examples

```
permutation_counter_diff_means(sample_district_1DS, rand_matrix_1DS, -3323,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre,128,7)
```

`permutation_counter_paired_t` *Creates paired T statistic permutation vector for counternull distribution*

Description

Resamples data to create counternull distribution. Returns vector with test statistics in counternull distribution.

Usage

```
permutation_counter_paired_t(
  sample_data,
  rand_matrix,
  counternull_value,
  variable,
  iterations,
  pairs
)
```

Arguments

`sample_data`	Sample data set. Data should have column indicating exposure (1) or non exposure (0) for each group (row) that is measured. Each measured outcome (variable) should be represented by an additional column.
`rand_matrix`	Matrix with unique randomizations for exposure assignment
`counternull_value`	Number to test out as counternull value
`variable`	Variable measured Format: sample_data$column
`iterations`	Numbers of unique arrangements of exposure assignments used to generate distribution (At most the number of rows in rand_matrix)
`pairs`	Number of pairs of units there are to measure in dataset (One pair = control unit + experimental unit)

Value

Vector with all generated test statistics in counternull distribution

Examples

```
permutation_counter_paired_t(sample_district_1DS, rand_matrix_1DS, -3323,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre,128,7)
```

`permutation_counter_t` *Creates T statistic permutation vector for counternull distribution*

Description

Resamples data to create counternull distribution. Returns vector with test statistics in counternull distribution.

Usage

```
permutation_counter_t(
  sample_data,
  rand_matrix,
  counternull_value,
  variable,
  iterations,
  pairs
)
```

Arguments

`sample_data`	Sample data set. Data should have column indicating exposure (1) or non exposure (0) for each group (row) that is measured. Each measured outcome (variable) should be represented by an additional column.
`rand_matrix`	Matrix with unique randomizations for exposure assignment
`counternull_value`	Number to test out as counternull value
`variable`	Variable measured Format: sample_data$column
`iterations`	Numbers of unique arrangements of exposure assignments used to generate distribution (At most the number of rows in rand_matrix)
`pairs`	Number of pairs of units there are to measure in dataset (One pair = control unit + experimental unit)

Value

Vector with all generated test statistics in counternull distribution

Examples

```
permutation_counter_t(sample_district_1DS, rand_matrix_1DS, -3323,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre,128,7)
```

permutation_null_cohens_d

Creates Cohen's D vector for null distribution

Description

Resamples data to create null distribution. Returns vector with test statistics in null distribution.

Usage

```
permutation_null_cohens_d(rand_matrix, variable, iterations)
```

Arguments

rand_matrix	Matrix with unique randomizations for exposure assignment
variable	Variable measured Format: sample_data$column
iterations	Numbers of unique arrangements of exposure assignments used to generate distribution (At most the number of rows in rand_matrix)

Value

Vector with all generated test statistics in null distribution

Examples

```
permutation_null_cohens_d(rand_matrix_1DS,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre,128)
```

permutation_null_diff_means

Creates difference of means permutation vector for null distribution

Description

Resamples data to create null distribution. Returns vector with test statistics in null distribution.

Usage

```
permutation_null_diff_means(rand_matrix, variable, iterations)
```

Arguments

rand_matrix	Matrix with unique randomizations for exposure assignment
variable	Variable measured Format: sample_data$column
iterations	Numbers of unique arrangements of exposure assignments used to generate distribution (At most the number of rows in rand_matrix)

Value

Vector with all generated test statistics in null distribution

Examples

```
permutation_null_diff_means(rand_matrix_1DS,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre,128)
```

---

permutation_null_paired_t

Creates Paired T statistic vector for null distribution

Description

Resamples data to create null distribution. Returns vector with test statistics in null distribution.

Usage

```
permutation_null_paired_t(rand_matrix, variable, iterations)
```

Arguments

rand_matrix	Matrix with unique randomizations for exposure assignment
variable	Variable measured Format: sample_data$column
iterations	Numbers of unique arrangements of exposure assignments used to generate distribution (At most the number of rows in rand_matrix)

Value

Vector with all generated test statistics in null distribution

Examples

```
permutation_null_paired_t(rand_matrix_1DS,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre,128)
```

permutation_null_t *Creates T statistic vector for null distribution*

Description

Resamples data to create null distribution. Returns vector with test statistics in null distribution.

Usage

```
permutation_null_t(rand_matrix, variable, iterations)
```

Arguments

rand_matrix	Matrix with unique randomizations for exposure assignment
variable	Variable measured Format: sample_data$column
iterations	Numbers of unique arrangements of exposure assignments used to generate distribution (At most the number of rows in rand_matrix)

Value

Vector with all generated test statistics in null distribution

Examples

```
permutation_null_t(rand_matrix_1DS,
sample_district_1DS$charge_prosecuted_1000_rate_post -
sample_district_1DS$charge_prosecuted_1000_rate_pre,128)
```

rand_matrix_1D *Randomization matrix of body camera assignments for District1D*

Description

This matrix contains 10,000 possible body camera assignments for District1D. 0 means no camera. 1 means camera.

Usage

```
rand_matrix_1D
```

Format

A matrix with 10000 columns (body camera assignment options):

0 Body Camera On

1 Body Camera Off

rand_matrix_1DS *Randomization matrix of body camera assignments for District1DS*

Description

This matrix contains all possible body camera assignments for District1DS. 0 means no camera. 1 means camera.

Usage

```
rand_matrix_1DS
```

Format

A matrix with 128 columns (body camera assignment options):

0 Body Camera On

1 Body Camera Off

sample_district_1D *Sample data for Police District1D revealing body camera assignment and behavioral outcomes*

Description

This CSV dataset is taken from a study measuring impact of body cameras on police behavioral outcomes in Washington D.C. police districts. It includes the body camera assignments for police officers (142 pairs) in District1D as well as their ID numbers and rates of different behavioral outcomes pre and post body camera assignment.

Usage

```
sample_district_1D
```

Format

A table with 225 behavioral outcomes:

z Body Camera Assignment

block_id ID Number

district District

district_block_id District ID

columns 5-229 Behavioral Outcomes

References

doi: 10.1073/pnas.1814773116

`sample_district_1DS` *Sample data for Police District1DS revealing body camera assignment and behavioral Outcomes*

Description

This CSV dataset is taken from a study measuring impact of body cameras on police behavioral outcomes in Washington D.C. police districts. It includes the body camera assignments for police officers (7 pairs) in District1DS as well as their ID numbers and rates of different behavioral outcomes pre and post body camera assignment.

Usage

```
sample_district_1DS
```

Format

A table with 225 behavioral outcomes:

z Body Camera Assignment

block_id ID Number

district District

district_block_id District ID

columns 5-229 Behavioral Outcomes

References

doi: 10.1073/pnas.1814773116

Index