

# Package ‘Rsampling’

August 29, 2016

**Title** Ports the Workflow of “Resampling Stats” Add-in to R

**Version** 0.1.1

**Description** Resampling Stats (<http://www.resample.com>) is an add-in for running randomization tests in Excel worksheets. The workflow is (1) to define a statistic of interest that can be calculated from a data table, (2) to randomize rows ad/or columns of a data table to simulate a null hypothesis and (3) and to score the value of the statistic from many randomizations. The relative frequency distribution of the statistic in the simulations is then used to infer the probability of the observed value be generated by the null process (probability of Type I error). This package intends to translate this logic for R for teaching purposes. Keeping the original workflow is favored over performance.

**Depends** R (>= 3.0.0)

**Imports** graphics, stats, utils

**License** GPL-2

**LazyData** true

**Suggests** knitr, rmarkdown

**VignetteBuilder** knitr

**RoxygenNote** 5.0.1

**Encoding** UTF-8

**NeedsCompilation** no

**Author** Paulo Prado [aut, cre],  
Andr'e Chalom [aut],  
Alexandre Oliveira [aut]

**Maintainer** Paulo Prado <[prado@ib.usp.br](mailto:prado@ib.usp.br)>

**Repository** CRAN

**Date/Publication** 2016-06-07 14:46:12

## R topics documented:

azteca . . . . .	2
basefunctions . . . . .	3
dplot . . . . .	4
embauba . . . . .	5
peucetia . . . . .	6
pielou . . . . .	6
rhyzophora . . . . .	7
Rsampling . . . . .	8
splot . . . . .	9
zfsample . . . . .	9

<b>Index</b>	<b>11</b>
--------------	-----------

---

azteca	<i>An experiment on ant recruitment</i>
--------	---

---

### Description

Number of *Azteca* ants recruited by leaf extracts of their host plant , *Cecropia* trees.

### Usage

azteca

### Format

A data frame with 21 rows and 3 variables:

**plant** plant id, integer

**extract.new** number of recruited ants in the leaf that received drops of smashed new leaves extract

**extract.old** number of recruited ants in the leaf that received drops of smashed old leaves extract

### Details

The ant colonies live in the hollow trunk of *Cecropia* and can detect and expel leaf-chewing insects. To test if this response is more intense in young leaves, drops of extract of smashed young and old leaves were poured in two neighbor leaves of the same plant. After 7 minutes the number of recruited ants in each leaf was recorded.

### Source

Kondrat, H. 2012. Estímulos químicos de folhas novas promovem recrutamento eficiente de formigas associadas à embaúba *Cecropia glaziovii* (Urticaceae). Curso de campo "Ecologia da Mata Atlântica" (G. Machado; P.I. Prado & A.M.Z. Martini, eds.). Universidade de São Paulo, São Paulo. <http://ecologia.ib.usp.br/curso/2012/PDF/PI-Hebert.pdf>

## Description

Functions to run (un)restricted sampling with or without replacement in a dataframe.

## Usage

```
within_rows(dataframe, cols = 1:ncol(dataframe), replace = FALSE,  
            FUN = base::sample)
```

```
within_columns(dataframe, cols = 1:ncol(dataframe), stratum = rep(1,  
                        nrow(dataframe)), replace = FALSE, FUN = base::sample)
```

```
normal_rand(dataframe, cols = 1:ncol(dataframe), stratum = rep(1,  
                        nrow(dataframe)), replace = FALSE, FUN = base::sample)
```

```
rows_as_units(dataframe, stratum = rep(1, nrow(dataframe)), replace = FALSE,  
              length.out = NULL)
```

```
columns_as_units(dataframe, cols = 1:ncol(dataframe), replace = FALSE,  
                 length.out = NULL)
```

## Arguments

dataframe	a dataframe with the data to be shuffled or resampled.
cols	columns of dataframe that should be selected to be resampled/shuffled. Defaults for all columns.
replace	(logical) should the data be permuted (FALSE) or resampled with replacement (TRUE) ?
FUN	function used for the sampling procedure. The default is <code>sample</code> , and a new function <code>zfsample</code> is provided for sampling with fixed zeroes.
stratum	factor or integer vector that separates data in groups or strata. Randomizations will be performed within each level of the stratum. Needs at least two observations in each level. Default is a single-level stratum.
length.out	(integer) specifies the size of the resulting data set. For <code>columns_as_units</code> , a data.frame with <code>length.out</code> columns will be returned, and for <code>rows_as_units</code> , a data.frame with <code>length.out</code> rows will be returned. Note that if <code>length.out</code> is larger than the relevant dimension, <code>replace</code> must also be specified.

## Value

a dataframe with the same structure of those input in dataframe with values randomized accordingly.

## Details

Each function performs as close as possible the corresponding options in Resampling Stats add-in for Excel (www.resample.com) for permutation (shuffling) and sampling with replacement (resampling) values in tabular data:

- `normal_rand` corresponds to the 'normal shuffle' and 'normal resample' option. For shuffling (`replace=FALSE`) the data is permuted over all cells of dataframe. For resampling (`replace=TRUE`) data from any cell can be sampled and attributed to any other cell.
- `within_rows` and `within_columns` correspond to the options with the same names. The randomization is done within each row or column of dataframe. So for shuffling the values of each row/column are permuted independently and for resampling the values are sampled independently from each row/column and attributed only to cells of the row/column they were sampled.
- `rows_as_units` and `columns_as_units` also correspond to the options with the same names. Each row or column dataframe is shuffled or resampled as whole. Only the placement of rows and columns in the dataframe change. The values and their position within each row/column remains the same.

All functions assemble the randomized values in a dataframe of the same configuration of the original. Columns that were not selected to be randomized with argument `cols` are then bound to the resulting dataframe. The order and names of the rows and columns are preserved, except if `length.out` is specified. In this case, the randomized rows/columns may be shifted to the end of the table.

When both `stratum` and `length.out` are used, the function will try to keep the proportion of each strata close to the original.

## References

Statistics.com LCC. 2009. Resampling Stats Add-in for Excel User's Guide. <http://www.resample.com/content/software/excel/userguide/RXLHelp.pdf>

---

dplot

*Statistic distribution plot*

---

## Description

Plots the distribution of the statistic of interest. Has switches to plot the extreme values and null hypothesis rejection region (also known as critical region).

## Usage

```
dplot(dist, svalue, pside = c("Two sided", "Greater", "Lesser"),
      extreme = TRUE, vline = TRUE, rejection = TRUE, ...)
```

**Arguments**

<code>dist</code>	the statistic distribution, as generated by <code>Rsampling</code> (numeric vector)
<code>svalue</code>	the result of applying the statistic over the original data
<code>pside</code>	the alternative hypothesis for the hypothesis testing
<code>extreme</code>	logical. should extreme points be highlighted in the plot?
<code>vline</code>	logical. should the <code>svalue</code> be displayed as a vertical line?
<code>rejection</code>	logical. should the critical region be highlighted?
<code>...</code>	further arguments to be passed to <code>hist</code> function.

**See Also**

See the package vignettes for more information about how to interpret this graph

---

embauba	<i>Vine infestation on Cecropia trees</i>
---------	---

---

**Description**

Presence/absence data of vines on Cecropia trees of two morphotypes.

**Usage**

```
embauba
```

**Format**

A data frame with 152 rows (plants) and 2 variables:

**morphotype** the tree morphotype, factor with two levels

**with.vines** does the tree harbor vines? Logical.

**Details**

Two morphotypes of Cecropia trees differ in the occupancy by ant colonies. Ants attack and drive out other insects that get to the trees. To test if this protection also affects infestation by vines, trees of similar size of both morphs were sampled and inspected for the presence of vines.

**Source**

Mello, T.J. 2012. Infestação por lianas e comportamento de poda por formigas em *Cecropia* (Urticaceae). Curso de campo "Ecologia da Mata Atlântica" (G. Machado; P.I. Prado & A.M.Z. Martini, eds.). Universidade de São Paulo, São Paulo. <http://ecologia.ib.usp.br/curso/2012/PDF/PI-Thayna.pdf>

---

 peucetia

*Preference of hunting spiders by hairy leaves*


---

**Description**

Occupancy of *Peucetia* spiders on parts of an experimental arena covered by leaves with or without trichomes.

**Usage**

peucetia

**Format**

A data frame with 27 rows (trials) and 6 variables:

**t1,t2,t3,t4,t5,t6** Is the spider on the part covered by hairy leaves? Logical, for each of 6 successive inspections (time 1, 2, ...)

**Details**

Spiders of the genus *Peucetia* do not make webs and hunt actively on the vegetation. The data is from an experiment to test if spiders prefer to stay in hairy leaves, that can stick their prey. The spiders were kept in Petri dishes that had half of lower plate covered with hairy leaves. The other half was covered by leaves without trichomes. The placement of each spider was recorded 6 times at each 30 min.

**Source**

Werneck, R.T. 2010. Lar, viscoso lar. Experimento de seleção de habitat e forrageio de aranhas em plantas com tricomas glandulares. Curso de campo "Ecologia da Mata Atlântica" (G. Machado; P.I. Prado & A.A. Oliveira, eds.). Universidade de São Paulo, São Paulo. <http://ecologia.ib.usp.br/curso/2010/pages/pdf/PI/relatorios/rachel.pdf>

---

 pielou

*Aphids recorded on goldenrods*


---

**Description**

Occurrences of aphids of the genus *Dactynotus* on plants of the genus *Solidago* in Canada

**Usage**

pielou

**Format**

A dataframe with 10 rows (aphid species of the genus *Dactynotus*) and 12 columns (plant species of the genus *Solidago*). Each entry is the number of records of a given aphid species on a plant species.

**Details**

Data from a field survey by E.C. Pielou in Ontario to exemplify a method to calculate niche overlap and niche width. The niche overlap gauges the overall similarity of the plant ranges used by the aphids. The niche width expresses how diverse is the average diversity of plants used by the aphids.

**Source**

Pielou, E.C. 1972. Niche width and niche overlap: a method for measuring them. *Ecology*, 53: 687–692.

---

 rhizophora

*Allometry in mangrove trees*


---

**Description**

Canopy to height ratio and variables of root area in mangrove trees sampled in two soil types.

**Usage**

rhizophora

**Format**

A data frame with 24 rows (trees) and 4 variables:

**soil.instability** soil type according to instability; factor with two levels (high / medium)

**canopy.trunk** ratio between canopy and trunk area, both in m<sup>2</sup>, numeric

**root** area covered by aerial roots, numeric (m<sup>2</sup>)

**n.roots** number of aerial roots, integer

**Details**

Data from a field practical exercise to test if mangrove trees in more unstable soil allocates more biomass in supporting roots.

**Source**

Prado, A. *et al.* 2013. Variações na morfologia de sustentação em *Rhizophora mangle* (Rizophoraceae) em diferentes condições de inundação do solo. Curso de campo "Ecologia da Mata Atlântica" (G. Machado, P.I. Prado & A.M.Z. Martini eds.). Universidade de São Paulo, São Paulo. <http://ecologia.ib.usp.br/curso/2013/pdf/P04-2.pdf>

Rsampling

*Repeats randomizations and scores summary statistics***Description**

Repeats resampling/shuffling of dataframes and scores the values returned by user-define function which is applied to each randomized dataframe.

**Usage**

```
Rsampling(type = c("normal_rand", "rows_as_units", "columns_as_units",
  "within_rows", "within_columns"), dataframe, statistics, ntrials = 10000,
  simplify = TRUE, progress = "text", fix.zeroes = FALSE, ...)
```

**Arguments**

type	character; the name of the randomization function to be applied to dataframe. See <a href="#">randomization functions</a> .
dataframe	a dataframe with the data to be shuffled or resampled.
statistics	a function that calculates the statistics of interest from the dataframe. The first argument should be the dataframe with the data and preferably should return a (named) vector, data frame, matrix or array.
ntrials	integer; number of randomizations to perform.
simplify	logical; should the result be simplified to a vector, matrix or higher dimensional array if possible?
progress	which kind of progress bar should be used (currently unimplemented!)
fix.zeroes	logical; for normal_rand, within_rows or within_columns, should zeroes in the dataframe be kept in place? See the help on <a href="#">zfsample</a> for more details.
...	further arguments to be passed to the randomization functions (e.g., cols, replace, stratum).

**Value**

a list of objects returned by the function defined by `statistics` or a vector, matrix or array when `simplify=TRUE` and simplification can be done (see [simplify2array](#)).

**Details**

This function corresponds to *Repeat and score* in Resampling Stats add-in for Excel ([www.resample.com](http://www.resample.com)). The randomization function defined by `type` is applied `ntrials` times on the data provided by `dataframe`. At each trial the function defined by argument `statistics` is applied to the resulting dataframe and the resulting objects are returned.

**References**

Statistics.com LCC. 2009. Resampling Stats Add-in for Excel User's Guide. <http://www.resample.com/content/software/excel/userguide/RXMLHelp.pdf>



---

splot	<i>Spaghetti plot</i>
-------	-----------------------

---

**Description**

Quick plot of paired differences, for exploratory purposes.

**Usage**

```
splot(p1, p2, highlight = TRUE, col.dif = c("black", "grey"),
      groups.names = c(1, 2), ...)
```

**Arguments**

p1, p2	vectors of paired values (numerical vectors)
highlight	should positive and negative differences within pairs highlighted with different colors? Logical
col.dif	color vector if highlight = TRUE
groups.names	labels for the groups names; numeric or character vector of length two.
...	further arguments to be passed to plot function.

---

zfsample	<i>Zero-fixed (re)sampling</i>
----------	--------------------------------

---

**Description**

This function builds on [sample](#) to provide sampling from a vector, but with all zero entries fixed. This way, `zfsample(c(0, 1, 0, 2))` may result in (0,1,0,2) or (0,2,0,1), but the positions that were initially zero will remain zeroed.

**Usage**

```
zfsample(x, replace = FALSE)
```

**Arguments**

x	Either a vector of one or more elements from which to choose, or a positive integer.
replace	Should sampling be with replacement?

**Value**

a vector of the same length of 'x' with elements drawn from 'x'.

**Details**

The actual sampling is done by [sample](#), so its help page should be checked for details on the parameter handling. The parameter 'size' is always passed as `length(x)`, and 'prob' is not supported.

**Examples**

```
# Sampling without replacement
zfsample(c(0,1,2,0,3,4,0))
# Sampling with replacement
zfsample(c(0,1,2,0,3,4,0), replace=TRUE)
# With no zeroes, zfsample just calls sample
set.seed(42); s1<-sample(c(1,2,3,4,5,6))
set.seed(42); s2<-zfsample(c(1,2,3,4,5,6))
all.equal(s1, s2)
```

# Index

## \*Topic **datasets**

- azteca, [2](#)
- embauba, [5](#)
- peucetia, [6](#)
- pielou, [6](#)
- rhyzophora, [7](#)

azteca, [2](#)

basefunctions, [3](#)

columns\_as\_units (basefunctions), [3](#)

dplot, [4](#)

embauba, [5](#)

normal\_rand (basefunctions), [3](#)

peucetia, [6](#)

pielou, [6](#)

randomization functions, [8](#)

rhyzophora, [7](#)

rows\_as\_units (basefunctions), [3](#)

Rsampling, [5](#), [8](#)

sample, [3](#), [9](#), [10](#)

simplify2array, [8](#)

splot, [9](#)

within\_columns (basefunctions), [3](#)

within\_rows (basefunctions), [3](#)

zfsample, [3](#), [8](#), [9](#)