

Package ‘ipmisc’

January 9, 2021

Type Package

Title Miscellaneous Functions for Data Cleaning and Analysis

Version 5.0.2

Maintainer Indrajeet Patil <patilindrajeet.science@gmail.com>

Description Provides functions needed for data cleaning and formatting and forms data cleaning and wrangling backend for the following packages: 'broomExtra', 'ggstatsplot', 'groupedstats', 'pairwiseComparisons', 'statsExpressions', and 'tidyBF'.

License GPL-3 | file LICENSE

URL <https://indrajeetpatil.github.io/ipmisc/>,
<https://github.com/IndrajeetPatil/ipmisc>

BugReports <https://github.com/IndrajeetPatil/ipmisc/issues>

Depends R (>= 3.6.0)

Imports dplyr, magrittr, rlang, tibble, tidyr, zeallot

Suggests ggplot2, rmarkdown, parameters, spelling, testthat

Encoding UTF-8

Language en-US

LazyData true

RoxygenNote 7.1.1

Config/testthat/edition 3

Config/testthat/parallel true

Config/testthat/start-first watcher, parallel*

NeedsCompilation no

Author Indrajeet Patil [aut, cre, cph]
(<<https://orcid.org/0000-0003-1995-6531>>)

Repository CRAN

Date/Publication 2021-01-09 02:20:06 UTC

R topics documented:

bugs_long	2
iris_long	3
long_to_wide_converter	4
specify_decimal_p	5
stats_type_switch	6
Index	7

bugs_long	<i>Tidy version of the "Bugs" dataset.</i>
-----------	--------------------------------------------

Description

Tidy version of the "Bugs" dataset.

Usage

```
bugs_long
```

Format

A data frame with 372 rows and 6 variables

- subject. Dummy identity number for each participant.
- gender. Participant's gender (Female, Male).
- region. Region of the world the participant was from.
- education. Level of education.
- condition. Condition of the experiment the participant gave rating for (**LDLF**: low frighteningness and low disgustingness; **LFHD**: low frighteningness and high disgustingness; **HFHD**: high frighteningness and low disgustingness; **HFHD**: high frighteningness and high disgustingness).
- desire. The desire to kill an arthropod was indicated on a scale from 0 to 10.

Details

This data set, "Bugs", provides the extent to which men and women want to kill arthropods that vary in frighteningness (low, high) and disgustingness (low, high). Each participant rates their attitudes towards all anthropods. Subset of the data reported by Ryan et al. (2013).

Source

<https://www.sciencedirect.com/science/article/pii/S0747563213000277>

Examples

```
dim(iris_long)
head(iris_long)
dplyr::glimpse(iris_long)
```

iris_long

Edgar Anderson's Iris Data in long format.

Description

Edgar Anderson's Iris Data in long format.

Usage

```
iris_long
```

Format

A data frame with 600 rows and 5 variables

- id. Dummy identity number for each flower (150 flowers in total).
- Species. The species are *Iris setosa*, *versicolor*, and *virginica*.
- condition. Factor giving a detailed description of the attribute (Four levels: "Petal.Length", "Petal.Width", "Sepal.Length", "Sepal.Width").
- attribute. What attribute is being measured ("Sepal" or "Petal").
- measure. What aspect of the attribute is being measured ("Length" or "Width").
- value. Value of the measurement.

Details

This famous (Fisher's or Anderson's) iris data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are *Iris setosa*, *versicolor*, and *virginica*.

This is a modified dataset from datasets package.

Source

<https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/iris.html>

Examples

```
dim(iris_long)
head(iris_long)
dplyr::glimpse(iris_long)
```

`long_to_wide_converter`*Converts dataframe from long/tidy to wide format with NAs removed*

Description

This conversion is helpful mostly for repeated measures design.

Usage

```
long_to_wide_converter(  
  data,  
  x,  
  y,  
  subject.id = NULL,  
  paired = TRUE,  
  spread = TRUE,  
  ...  
)
```

Arguments

<code>data</code>	A dataframe (or a tibble) from which variables specified are to be taken. A matrix or tables will not be accepted.
<code>x</code>	The grouping variable from the dataframe <code>data</code> .
<code>y</code>	The response (a.k.a. outcome or dependent) variable from the dataframe <code>data</code> .
<code>subject.id</code>	In case of repeated measures design (<code>paired = TRUE</code> , i.e.), this argument specifies the subject or repeated measures id. Note that if this argument is <code>NULL</code> (which is the default), the function assumes that the data has already been sorted by such an id by the user and creates an internal identifier. So if your data is not sorted and you leave this argument unspecified, the results can be inaccurate.
<code>paired</code>	Logical that decides whether the experimental design is repeated measures/within-subjects or between-subjects. The default is <code>FALSE</code> .
<code>spread</code>	Logical that decides whether the dataframe needs to be converted from long/tidy to wide (default: <code>TRUE</code>), or the data needs to be returned as it is but with the NAs removed.
<code>...</code>	Currently ignored.

Value

A dataframe in the wide (or Cartesian) format.

Examples

```
# for reproducibility
library(ipmisc)
set.seed(123)

# repeated measures design
long_to_wide_converter(
  data = bugs_long,
  x = condition,
  y = desire,
  subject.id = subject,
  paired = TRUE
)

# independent measures design
long_to_wide_converter(
  data = ggplot2::msleep,
  x = vore,
  y = brainwt,
  paired = FALSE,
  spread = FALSE
)
```

specify_decimal_p *Formatting numeric values*

Description

Function to format an R object for pretty printing with a specified (k) number of decimal places. The function also allows really small p -values to be denoted as " $p < 0.001$ " rather than " $p = 0.000$ ". Note that if `p.value` is set to `TRUE`, the minimum value of k allowed is 3. If k is set to less than 3, the function will ignore entered k value and use $k = 3$ instead. **Important:** This function is not vectorized.

Usage

```
specify_decimal_p(x, k = 3L, p.value = FALSE, ...)
```

```
format_num(x, k = 3L, p.value = FALSE, ...)
```

Arguments

<code>x</code>	A numeric value.
<code>k</code>	Number of digits after decimal point (should be an integer) (Default: $k = 3L$).
<code>p.value</code>	Decides whether the number is a p -value (Default: <code>FALSE</code>).
<code>...</code>	Currently ignored.

Value

Formatted numeric value.

Examples

```
specify_decimal_p(x = 0.0000123, k = 2, p.value = TRUE)
specify_decimal_p(x = 0.008675, k = 2, p.value = TRUE)
specify_decimal_p(x = 0.003458, k = 3, p.value = FALSE)
```

stats_type_switch	<i>Switch type of statistics.</i>
-------------------	-----------------------------------

Description

Relevant mostly for ggstatsplot and statsExpressions packages, where there are four types of statistics are supported: parametric, non-parametric, robust, and Bayesian. This switch function converts strings entered by users to a common pattern.

Usage

```
stats_type_switch(type)
```

Arguments

type	Character string describing the type of statistics.
------	-----------------------------------------------------

Examples

```
stats_type_switch("p")
stats_type_switch("bf")
```

Index

* datasets

bugs_long, [2](#)

iris_long, [3](#)

bugs_long, [2](#)

format_num(specify_decimal_p), [5](#)

iris_long, [3](#)

long_to_wide_converter, [4](#)

specify_decimal_p, [5](#)

stats_type_switch, [6](#)