

Package ‘eatTools’

February 9, 2021

Type Package

Title Miscellaneous Functions for the Analysis of Educational Assessments

Version 0.5.0

Depends R (>= 3.5.0)

Imports stats, data.table

Date 2021-01-25

Description Miscellaneous functions for data cleaning and data analysis of educational assessments. Includes functions for descriptive analyses, character vector manipulations and weighted statistics. Mainly a lightweight dependency for the packages 'eatRep', 'eatGADS', 'eatPrep' and 'eatModel' (which will be subsequently submitted to 'CRAN'). The function for defining (weighted) contrasts in weighted effect coding refers to te Grotenhuis et al. (2017) <doi:10.1007/s00038-016-0901-1>. Functions for weighted statistics refer to Wolter (2007) <doi:10.1007/978-0-387-35099-8>.

License GPL (>= 2)

URL <https://github.com/weirichs/eatTools>

LazyData true

LazyLoad yes

Suggests testthat, covr

NeedsCompilation no

Author Sebastian Weirich [aut, cre],
Martin Hecht [aut],
Karoline Sachse [aut],
Benjamin Becker [aut],
Nicole Mahler [aut]

Maintainer Sebastian Weirich <sebastian.weirich@iqb.hu-berlin.de>

Repository CRAN

Date/Publication 2021-02-09 10:50:02 UTC

R topics documented:

asNumericIfPossible	2
catch_asNumericIfPossible	3
contr.wec.weighted	4
crop	5
descr	6
do_call_rbind_withName	7
eatTools	8
existsBackgroundVariables	8
facToChar	9
gsubAll	10
halveString	10
insert.col	11
multiseq	12
num.to.cat	12
recodeLookup	13
removeNonNumeric	14
removeNumeric	14
removePattern	15
set.col.type	16
tablePattern	17
tableUnlist	18
whereAre	18
wideToLong	19
wtdTable	20
wtdVar	21
Index	22

asNumericIfPossible	<i>Convert a Vector, Matrix or Data Frame Into Numeric Values If Possible</i>
---------------------	---

Description

This function converts vectors and matrices of all kinds to numeric. The function can also be used to convert all columns of a data.frame to class numeric for which this conversion is possible i.e. without creating NA when it fails. Non-convertible columns are maintained.

Usage

```
asNumericIfPossible(x, maintain.factor.scores = TRUE, force.string = TRUE,
transform.factors = TRUE, varName = NULL)
```

Arguments

<code>x</code>	A vector or data frame which should be converted.
<code>maintain.factor.scores</code>	Logical: If TRUE, conversion of the factor levels is attempted (like in <code>as.numeric(as.character(f))</code>). If FALSE, the internal codes of the factor are returned (like in <code>as.numeric(f)</code>). See 'Details'. This argument is only evaluated if <code>transform.factors = TRUE</code> .
<code>force.string</code>	Logical indicating whether columns should be force to numeric, even if NAs are induced. If FALSE, affected columns are maintained. If TRUE, conversion is forced.
<code>transform.factors</code>	Logical indicating whether columns of class factor should be converted. If FALSE, columns of class factor are maintained. If TRUE, conversion of factors is attempted.
<code>varName</code>	Optional: Name of the corresponding variable. Doesn't have to be changed by user.

Details

In R, factors may represent ordered categories or categorical variables. Depending on the meaning of the variable, a conversion of the nominal values (of a factor variable) to numeric values may be desirable or not. The arguments `transform.factors` and `maintain.factor.scores` specify if and how factor variables should be treated. See examples.

Author(s)

Sebastian Weirich, Karoline Sachse, Benjamin Becker

Examples

```
dat <- data.frame(X1 = c("1",NA,"0"), X2 = c("a",NA,"b"),
                 X3 = c(TRUE,FALSE,FALSE), X4 = as.factor(c("a",NA,"b")),
                 X5 = as.factor(c("5","6","7")), stringsAsFactors = FALSE)

str(dat)
asNumericIfPossible(dat)
asNumericIfPossible(dat, transform.factors=TRUE,
                    maintain.factor.scores=FALSE)
asNumericIfPossible(dat, transform.factors=TRUE,
                    maintain.factor.scores=TRUE)
```

catch_asNumericIfPossible

Use asNumericIfPossible with modified warning.

Description

This function uses `asNumericIfPossible` but lets the user change the warning issued by `asNumericIfPossible`. Suited for use in other R packages.

Usage

```
catch_asNumericIfPossible(x, warn, maintain.factor.scores = TRUE,
  force.string = TRUE, transform.factors = TRUE)
```

Arguments

x	A vector or data frame which should be converted.
warn	A character vector of length 1 with the desired warning.
maintain.factor.scores	Logical: If TRUE, conversion of the factor levels is attempted (like in <code>as.numeric(as.character(f))</code>). If FALSE, the internal codes of the factor are returned (like in <code>as.numeric(f)</code>). See 'Details'. This argument is only evaluated if <code>transform.factors = TRUE</code> .
force.string	Logical indicating whether columns should be force to numeric, even if NAs are induced. If FALSE, affected columns are maintained. If TRUE, conversion is forced.
transform.factors	Logical indicating whether columns of class factor should be converted. If FALSE, columns of class factor are maintained. If TRUE, conversion of factors is attempted.

Details

For details see [asNumericIfPossible](#)

Author(s)

Benjamin Becker

Examples

```
char <- c("a", "b", 1)
catch_asNumericIfPossible(x = char, warn = "Vector could not be converted")
```

contr.wec.weighted	<i>Calculates contrasts for a weighted factor variable based on weighted effect coding</i>
--------------------	--

Description

Function works equivalent to `contr.wec` from the `wec` package, but allows for weighted contrasts.

Usage

```
contr.wec.weighted(x, omitted, weights)
```

Arguments

x	grouping variable of class factor
omitted	Label of the factor label that should be taken as the omitted category
weights	Numeric vector of non-negative weights

Value

Returns a contrast matrix based on weighted effect coding.

Author(s)

Sebastian Weirich, based upon the `contr.wec` function of the `wec` package

Examples

```
### exemplary data according to wec paper
dat <- data.frame ( group = as.factor(c(rep(1,3), rep(2,2))), wgt = c(2/3, 4/3, 2, 3/8, 5/8))
### default contrasts
contrasts(dat[, "group"])
### weighted effect coding for weighted data
contr.wec.weighted(x= dat[, "group"], omitted=1, weights=dat[, "wgt"])
### equal to weighted effect coding: wec::contr.wec(x= dat[, "group"], omitted=1)
contr.wec.weighted(x= dat[, "group"], omitted=1, weights=rep(1, nrow(dat)))
```

crop

Remove Trailing and Leading Characters From Character Strings

Description

Similarly to the function `trim` from the `gdata` package, this function can be used to remove trailing and leading spaces from character strings. However, in contrast to `trim`, any character can be removed by `crop`.

Usage

```
crop(x, char = " ")
```

Arguments

x	character string
char	character to be removed from beginning and end of x

Author(s)

Martin Hecht, Sebastian Weirich

Examples

```
str <- c(" 12 kk ", "op j   q ", "110")
crop(str)
crop(str, "op")
```

 descr

Descriptive statistics for one or several variables

Description

Function computes descriptive statistics for one variable or several variables within a data frame.

Usage

```
descr (variable, na = NA, p.weights = NULL, na.rm = FALSE, verbose=TRUE)
```

Arguments

variable	one variable or a data.frame with several variables
na	optional values with should be considered a missing values
p.weights	optional: vector with individual weights if weighted statistics should be computed
na.rm	logical: should missings be removed prior to estimation?
verbose	logical: Print messages to console?

Value

a data frame with the following columns

N	number of observations
N.valid	number of non-missing observations
Missing	number of missings
Minimum	minimum of numeric variables
Maximum	maximum of numeric variables
Sum	sum of numeric variables
Mean	arithmetic mean of numeric variables
std.err	standard error of the arithmetic mean. Note: for weighted means, standard error is estimated according to Cochran (1977): $\sigma_x^2 = n/((n-1) * w_s^2) * \text{Sigma}(w_i^2 * (x_i - x))$.
sig	p value
Median	median of numeric variables
SD	standard deviation of numeric variables
Var	variance of numeric variables

Author(s)

Sebastian Weirich

References

Cochran W. G. (1977). *Sampling Techniques* (3rd Edn). Wiley, New York

Examples

```
data(mtcars)
descr(mtcars)
```

```
do_call_rbind_withName
```

Row bind a list while assigning names to rows

Description

Use `do.call(rbind, ...)` on a list of `data.frames` while creating a new variable (`colName`) which contains, for example, the original list naming (`name`).

Usage

```
do_call_rbind_withName(df_list, name = names(df_list), colName)
```

Arguments

<code>df_list</code>	A list of <code>data.frames</code> .
<code>name</code>	Vector of names to fill <code>colName</code> . Default uses the names of <code>df_list</code> .
<code>colName</code>	A single character; name for the new column.

Value

Returns a `data.frame`.

Author(s)

Benjamin Becker

Examples

```
### create example list
df_list <- lapply(mtcars, function(x) {
  data.frame(m = mean(x), sd = sd(x))
})

### transform to a single data.frame
do_call_rbind_withName(df_list, colName = "variable")
```

eatTools	<i>eatTools: Miscellaneous Functions for the Analysis of Educational Assessments</i>
----------	--

Description

The eatTools package provides various groups of functions. The main groups of functions include: transformation of vector types, modification of character variables, descriptive analyses and weighted statistics. The package's purpose is mainly to function as a lightweight dependency for other packages.

Transformation of vector types

The functions `asNumericIfPossible` and `catch_asNumericIfPossible` transform character and factor variables to numeric. `facToChar` transforms factor variables to character. `set.col.type` allows manually setting the type of multiple variables within a `data.frame`.

Modification of character variables

Multiple convenience functions exist for modification of character variables: removing certain pattern (`removePattern`), removing numerics (`removeNumeric`) and removing non numerics (`removeNonNumeric`), substituting multiple patterns within a string (`gsubAll`) and splitting strings into multiple or a fixed number of parts but at specific position (`halveString`)

Descriptive Statistics

The function `descr` provides simple descriptive statistics for a `data.frame`, but in a format especially useful for further automated processing (long format `data.frame`).

Weighted Statistics

`wtdVar` provides calculation of weighted variances (this can be done also by the package `Hmisc`, which has, however, a very high number of dependencies). `wtdTable` provides a weighted frequency table.

existsBackgroundVariables	<i>Internally needed function for consistency checks and data preparation.</i>
---------------------------	--

Description

Function is necessary for `eatRep` and `eatModel` as well and therefore exported to namespace.

Usage

```
existsBackgroundVariables (dat, variable)
```


Arguments

dat	A data frame
variable	column number or variable name

Value

a structured list of variable names

Author(s)

Sebastian Weirich

Examples

```
data(mtcars)
existsBackgroundVariables(mtcars, 2:4)
```

facToChar	<i>Transform columns in a data frame</i>
-----------	--

Description

Function transforms all data frame columns of a specific class into another class.

Usage

```
facToChar ( dataFrame, from = "factor", to = "character")
```

Arguments

dataFrame	a data frame
from	which column class should be transformed?
to	target column class

Value

a data frame

Author(s)

Sebastian Weirich

Examples

```
data(mtcars)
### original classes
sapply(mtcars, class)
mtcars1 <- facToChar(mtcars, from = "numeric", to = "character")
sapply(mtcars1, class)
```

gsubAll *Pattern matching and replacement*

Description

Function is a wrapper for gsub() which allows to replace more than one pattern.

Usage

```
gsubAll ( string, old, new)
```

Arguments

string	a character vector where matches are sought
old	character vector containing strings to be matched in the given character vector named string.
new	a replacement for matched pattern

Value

character vector with replaced patterns

Author(s)

Benjamin Becker

Examples

```
### replace all numbers by words
txt <- "1 example for 2 reasons in 4 seasons"
gsubAll ( txt, old = as.character(1:4), new = c("one", "two", "three", "four"))
```

halveString *Split string exactly in two parts*

Description

strsplit splits a string according to a specific sign. The number of occurrences of the splitting sign defines the number of splits. halveString allows to split the string in only two parts, no matter how often the splitting sign occurs.

Usage

```
halveString (string, pattern, first = TRUE )
```

Arguments

string	A character vector.
pattern	character vector (or object which can be coerced to such) to use for splitting.
first	Logical: Relevant if the pattern occurs more than one time in the string. Defines whether the first (default) or last occurrence is used for splitting.

Value

A data.frame with two columns

Author(s)

Sebastian Weirich

Examples

```
str <- c("John_Bolton", "Richard_Milhouse_Nixon", "Madonna")
strsplit(str, split = "_")
halveString(str, pattern = "_")
halveString(str, pattern = "_", first=FALSE)
```

insert.col

Insert Columns into a Data Frame in a Specific Position

Description

Insert columns into a data frame in specific position

Usage

```
insert.col(dat, toinsert, after)
```

Arguments

dat	A data frame
toinsert	Column name(s) or column number(s) of the columns to be reinserted
after	Column name or column number after which the columns specified in insert should be reinserted.

Value

A data frame with columns in specified positions.

multiseq	<i>multiple sequences</i>
----------	---------------------------

Description

creates a sequence for every unique value in a vector

Usage

```
multiseq(v)
```

Arguments

v a vector

Value

a vector with multiple sequences

Author(s)

Martin Hecht

Examples

```
v <- c("a", "a", "a", "c", "b", "b", "a")
multiseq(v)
```

num.to.cat	<i>Transform continuous variables into ordered factors</i>
------------	--

Description

Function is useful if parameters on the ‘PISA’ metric should be transformed into competence levels.

Usage

```
num.to.cat(x, cut.points, cat.values = NULL)
```

Arguments

x Numeric vector.

cut.points Numeric vector with cut scores.

cat.values Optional: vector with labels for the cut scores. Note: if specified, length of cat.values should be length(cut.points)+1.

Value

Vector with factor values.

Author(s)

Sebastian Weirich

Examples

```
values <- rnorm(10,0,1.5) * 100 + 500
num.to.cat(x = values, cut.points = 390+0:3*75)
num.to.cat(x = values, cut.points = 390+0:3*75, cat.values = c("1a", "1b", 2:4))
```

recodeLookup

Recode a variable according to a lookup table

Description

Recodes the values of a variable. Function resembles the [recode](#) function from the `car` package, but uses a lookup table to specify old and new values.

Usage

```
recodeLookup(var, lookup)
```

Arguments

<code>var</code>	a vector (e.g. numeric, character, or factor)
<code>lookup</code>	a data.frame with exact two columns. First column contains old values, second column new values. Values which do not occur in the old column remain unchanged.

Value

a vector of the same length as `var` with recoded values

Examples

```
num_var <- sample(1:10, size = 10, replace = TRUE)
lookup <- data.frame(old = c(2, 4, 6), new = c(200,400,600))
num_var2<- recodeLookup(num_var, lookup)
```

removeNonNumeric *Removes all non-numeric characters from a string.*

Description

Function removes all non-numeric characters from a string.

Usage

```
removeNonNumeric ( string)
```

Arguments

string a character vector

Value

a character string

Author(s)

Sebastian Weirich

Examples

```
str <- c(".d1.nh.120", "empty", "110", ".nh.dgd", "only.nh")
removeNonNumeric(str)
```

removeNumeric *Removes alphanumeric characters from a string.*

Description

Function removes alphanumeric characters from a string.

Usage

```
removeNumeric ( string)
```

Arguments

string a character vector

Value

a character string

Author(s)

Sebastian Weirich

Examples

```
str <- c(".d1.nh.120", "empty", "110", ".nh.dgd", "only.nh")
removeNumeric(str)
```

removePattern	<i>Removes a specified pattern from a string.</i>
---------------	---

Description

Function remove a specified string from a character vector.

Usage

```
removePattern ( string, pattern)
```

Arguments

string	a character vector
pattern	a character pattern

Value

a character string

Author(s)

Sebastian Weirich

Examples

```
str <- c(".d1.nh.120", "empty", "110", ".nh.dgd", "only.nh")
removePattern(str, ".nh.")
```

set.col.type *Set the Class of Columns in a Data Frame*

Description

This function converts the classes of columns to character, numeric, logical, integer or factor.

Usage

```
set.col.type(dat, col.type = list("character" = NULL), verbose = FALSE, ...)
```

Arguments

dat	A data frame
col.type	A named list of column names that are to be converted. The names of the list indicate the class to which the respective column should be converted (character, numeric, numeric.if.possible, logical, integer or factor)
verbose	if TRUE details about converted columns are printed on the console
...	Additional arguments to be passed to asNumericIfPossible

Details

Use col.type="numeric.if.possible" if conversion to numeric should be tested upfront, see asNumericIfPossible for details.

Value

A data frame with column classes changed according to the specifications in col.type

Author(s)

Martin Hecht, Karoline Sachse

See Also

asNumericIfPossible

Examples

```
str(d <- data.frame("var1" = 1, "var2" = TRUE, "var3" = FALSE,
  "var4" = as.factor(1), "var5" = as.factor("a"), "var6" = "b",
  stringsAsFactors = FALSE))

str(set.col.type(d))
str(set.col.type(d, list("numeric" = NULL)))
str(set.col.type(d, list("character" = c("var1" , "var2"),
  "numeric" = "var3", "logical" = "var4")))
str(set.col.type(d, list("numeric.if.possible" = NULL)))
```



```
str(set.col.type(d, list("numeric.if.possible" = NULL),
  transform.factors = TRUE))
str(set.col.type(d, list("numeric.if.possible" = NULL), transform.factors = TRUE,
  maintain.factor.scores = FALSE))
```

tablePattern	<i>Creates skeleton for frequency tables with desired values</i>
--------------	--

Description

Function takes values and creates a frequency table including these values. Models behavior of factor variables.

Usage

```
tablePattern(x, pattern = NULL, weights, na.rm = TRUE,
  useNA = c("no", "ifany", "always"))
```

Arguments

x	a vector
pattern	desired values for table output
weights	optional: weights
na.rm	should missing values be removed
useNA	whether to include [NA] values in the table

Value

a frequency table

Author(s)

Sebastian Weirich

Examples

```
grades <- c(1,1,3,4,2,3,4,5,5,3,2,1)
table(grades)
tablePattern(grades, pattern = 1:6)
```

tableUnlist	<i>Frequency table for data frames, e.g. across multiple columns</i>
-------------	--

Description

Replaces the somehow buggy function combination `table(unlist(data))`.

Usage

```
tableUnlist(dataFrame, useNA = c("no", "ifany",  
  "always"))
```

Arguments

dataFrame	Data frame with more than one column.
useNA	whether to include NA values in the table. See help file of <code>table</code> for more details.

Value

A frequency table

Author(s)

Sebastian Weirich

Examples

```
dat <- data.frame ( matrix ( data = sample(0:1,200,replace=TRUE), nrow=20, ncol=10))  
tableUnlist(dat)
```

whereAre	<i>Matches a scalar with elements of a vector.</i>
----------	--

Description

The function closely resembles the `match` function, but allows for multiple matches.

Usage

```
whereAre(a,b,verbose=TRUE)
```

Arguments

a	a scalar
b	a numeric or character vector
verbose	logical: print messages on console?

Value

A numeric vector

Author(s)

Sebastian Weirich

Examples

```
a <- 12
b <- c(10, 11, 12, 10, 11, 12)
match(a, b)
whereAre(a=a, b=b)
```

wideToLong	<i>Transform wide format data sets into the long format necessary for eatRep analyses</i>
------------	---

Description

Data from large-scale assessments often are provided in the wide format. This function easily transform data into the long format required by eatRep.

Usage

```
wideToLong (datWide, noImp, imp, multipleColumns = TRUE, variable.name = "variable",
            value.name = "value")
```

Arguments

datWide	Data set in the wide format, i.e. one row per person
noImp	character vector of non-imputed variables which are desired for following analyses
imp	Named list of character vectors which include the imputed variables which are desired for following analyses
multipleColumns	Logical: use one column for each imputed variable (if more than one imputed variable is used)? Alternatively, only one column for all imputed variables is used (this is the default behavior of the <code>melt</code> function from the <code>reshape2</code> package).
variable.name	Applies only if <code>multipleColumns = "TRUE"</code> : name of variable used to store measured variable names
value.name	Applies only if <code>multipleColumns = "TRUE"</code> : name of variable used to store values

Value

A data.frame in the long format.

Author(s)

Sebastian Weirich

Examples

```
### create arbitrary wide format large-scale assessment data for two
### subjects, each with three imputations
datWide <- data.frame ( id = paste0("P",1:5), weight = abs(rnorm(5,10,1)),
  country = c("USA", "BRA", "TUR", "GER", "AUS"),
  sex = factor(c("female", "male", "female", "female", "male")),
  matrix(data = rnorm(n=15, mean = 500, sd = 75),
    nrow=5, dimnames = list(NULL, paste0("mat.pv", 1:3))),
  matrix(data = rnorm(n=15, mean = 480, sd = 80),
    nrow=5, dimnames = list(NULL, paste0("sci.pv", 1:3))),
  stringsAsFactors=FALSE)
datLong <- wideToLong(datWide = datWide, noImp = c("id", "weight", "country", "sex"),
  imp = list ( math = paste0("mat.pv", 1:3),
  science = paste0("sci.pv", 1:3)))
datLong2<- wideToLong(datWide = datWide, noImp = c("id", "weight", "country", "sex"),
  imp = list ( math = paste0("mat.pv", 1:3),
  science = paste0("sci.pv", 1:3)),
  multipleColumns = FALSE, variable.name = "varName",
  value.name = "val")
```

wtdTable

Computed weighted frequency tables

Description

This functions works quite equally as the `wtd.table` function from the Hmisc package.

Usage

```
wtdTable(x , weights , na.rm = FALSE)
```

Arguments

<code>x</code>	a character or category or factor vector
<code>weights</code>	a numeric vector of non-negative weights
<code>na.rm</code>	set to FALSE to suppress checking for NAs. If TRUE, NAs are removed from <code>x</code> as well as from <code>weights</code> prior to variance estimation.

Value

a frequency table

Examples

```
x <- c(50, 1, 50)
w <- c(1, 4, 1)
wtdTable(x, w)
```

wtdVar

Computed weighted variance

Description

This functions works quite equally as the [wtd.var](#) function from the `Hmisc` package.

Usage

```
wtdVar(x , weights , na.rm = FALSE)
```

Arguments

x	numeric vector
weights	a numeric vector of non-negative weights
na.rm	set to FALSE to suppress checking for NAs. If TRUE, NAs are removed from x as well as from weights prior to variance estimation.

Value

a scalar

Author(s)

Benjamin Becker

Examples

```
x <- c(50, 1, 25)
w <- c(1, 4, 1)
wtdVar(x, w)
```

Index

asNumericIfPossible, [2](#), [4](#), [8](#)

catch_asNumericIfPossible, [3](#), [8](#)
contr.wec.weighted, [4](#)
crop, [5](#)

descr, [6](#), [8](#)
do_call_rbind_withName, [7](#)

eatTools, [8](#)
existsBackgroundVariables, [8](#)

facToChar, [8](#), [9](#)

gsubAll, [8](#), [10](#)

halveString, [8](#), [10](#)

insert.col, [11](#)

melt, [19](#)
multiseq, [12](#)

num.to.cat, [12](#)

recode, [13](#)
recodeLookup, [13](#)
removeNonNumeric, [8](#), [14](#)
removeNumeric, [8](#), [14](#)
removePattern, [8](#), [15](#)

set.col.type, [8](#), [16](#)

tablePattern, [17](#)
tableUnlist, [18](#)

whereAre, [18](#)
wideToLong, [19](#)
wtd.table, [20](#)
wtd.var, [21](#)
wtdTable, [8](#), [20](#)
wtdVar, [8](#), [21](#)