

Package ‘descr’

January 2, 2012

Version 0.9.7

Date 2011-12-12

Title Descriptive statistics

Author Jakson Aquino. Includes R source code and/or documentation
written by Dirk Enzmann, Marc Schwartz, and Nitin Jain

Maintainer Jakson Aquino <jalvesaq@gmail.com>

Encoding UTF-8

Imports xtable, ENmisc

Description This package contains functions to describe weighted
categorical variables and functions to facilitate the character encoding conversion of objects.

License GPL (>= 2)

Repository CRAN

Date/Publication 2011-12-13 08:36:19

R topics documented:

compmeans	2
crosstab	3
CrossTable	5
descr	8
file.head	8
forODFTable	9
freq	10
fromUTF8	11
fwf2csv	12
histkdnc	13
LogRegR2	14
plot.CrossTable	15
plot.freqtable	16
toUTF8	16
xtable.CrossTable	17

compmeans	<i>Means of a numerical vector according to a factor</i>
-----------	--

Description

Calculates the means of a numerical vector according to a factor.

Usage

```
compmeans(x, f, w, sort = FALSE, maxlevels = 60,
          user.missing, plot = getOption("descr.plot"),
          relative.widths = TRUE, col = "lightgray", ...)
```

Arguments

x	A numeric vector.
f	A factor.
w	Optional vector with weights.
sort	If TRUE, sorts the lines by the means values.
maxlevels	Maximum number of levels that x converted into factor should have.
user.missing	Character vector, indicating what levels of f must be treated as missing values.
plot	Logical: if TRUE (default), a boxplot is produced. You may put <code>options(descr.plot = FALSE)</code> in your <code>‘.Rprofile’</code> to change the default function behavior.
relative.widths	If TRUE, the boxes widths will be proportional to the number of elements in each level of f.
col	Vector with the boxes colors.
...	Further arguments to be passed to either boxplot or wtd.boxplot .

Value

A matrix with class `c("matrix", "meanscomp")` with labels attributes for x and f. The returned object can be plotted, generating a [boxplot](#) of x grouped by f.

Author(s)

Jakson A. Aquino, <jalvesaq@gmail.com>.

See Also

[wtd.boxplot](#), [boxplot](#).

Examples

```

educ <- sample(c(1, 2), 200, replace = TRUE, prob = c(0.3, 0.7))
educ <- factor(educ, levels = c(1, 2), labels = c("Low", "High"))
income <- rnorm(200, 1000, 100) + 100 * as.numeric(educ)
attr(educ, "label") <- "Education level"
attr(income, "label") <- "Income"
w <- sample(c(10, 15, 19), 200, replace = TRUE)

compmeans(income, educ, col = c("#FFD0E0", "#D8E8FF"))
comp <- compmeans(income, educ, w, plot = FALSE)
comp
plot(comp)
## Not run:
print(xtable(comp))
# If the decimal separator in your country is a comma:
# options(OutDec = ",")
print(xtable(comp, align = "lrrr", display = c("s", "f", "d", "f")))

## End(Not run)

```

crosstab

Cross tabulation with mosaic plot

Description

This function is a wrapper for [CrossTable](#), adding a mosaic plot and making it easier to do a weighted cross-tabulation.

Usage

```

crosstab(x, y, weight = NULL, digits = 3, max.width = NA,
         expected = FALSE, prop.r = FALSE, prop.c = FALSE,
         prop.t = FALSE, prop.chisq = FALSE, chisq = FALSE,
         fisher = FALSE, mcnemar = FALSE, resid = FALSE,
         sresid = FALSE, asresid = FALSE,
         missing.include = FALSE, format = "SPSS",
         dnn = NULL, user.missing.x, user.missing.y,
         plot = getOption("descr.plot"),
         xlab = deparse(substitute(x)),
         ylab = deparse(substitute(y)), main = "", ...)

```

Arguments

<code>x</code> , <code>y</code>	Vectors in a matrix or a dataframe.
<code>weight</code>	An optional vector for a weighted cross tabulation.
<code>digits</code>	See CrossTable .
<code>max.width</code>	See CrossTable .

expected	See CrossTable .
prop.r	See CrossTable .
prop.c	See CrossTable .
prop.t	See CrossTable .
prop.chisq	See CrossTable .
chisq	See CrossTable .
fisher	See CrossTable .
mcnemar	See CrossTable .
resid	See CrossTable .
sresid	See CrossTable .
asresid	See CrossTable .
missing.include	See CrossTable .
format	See CrossTable .
dnn	See CrossTable .
user.missing.x	An optional character vector with the levels of x that should be treated as missing values.
user.missing.y	An optional character vector with the levels of y that should be treated as missing values.
plot	Logical: if TRUE (default), a mosaic plot is produced. You may put <code>options(descr.plot = FALSE)</code> in your <code>.Rprofile</code> to change the default function behavior.
main	An overall title for the plot (see title).
xlab	A title for the x axis (see title).
ylab	A title for the y axis (see title).
...	Further arguments to be passed to mosaicplot .

Details

`crosstab` invokes the [CrossTable](#) with all boolean options set to FALSE and "SPSS" as the default format option. The returned `CrossTable` object can be plotted as a [mosaicplot](#).

Author(s)

Jakson A. Aquino <jalvesaq@gmail.com>

See Also

[CrossTable](#), [plot.CrossTable](#), [xtable.CrossTable](#).

Examples

```

educ <- sample(c(1, 2), 200, replace = TRUE, prob = c(0.3, 0.7))
educ <- factor(educ, levels = c(1, 2), labels = c("Low", "High"))
income <- sample(c(1, 2, 3), 200, replace = TRUE,
                prob = c(0.3, 0.4, 0.3))
income <- factor(income, levels = c(1, 2, 3),
                labels = c("Low", "Middle", "High"))
attr(educ, "label") <- "Education level"
attr(income, "label") <- "Income level"
w <- sample(c(10, 15, 19), 200, replace = TRUE)

crosstab(income, educ, ylab = "Education", xlab = "Income")
ct <- crosstab(income, educ, w, expected = TRUE, plot = FALSE)
ct
plot(ct, inv.y = TRUE)
## Not run:
print(xtable(ct))
print(xtable(ct, decimal.mark = ",", digits = 1,
             multirow = TRUE, hline = TRUE, align = "llrrr"),
      sanitize.text.function = function(x) x,
      include.rownames = FALSE)

## End(Not run)

```

CrossTable

Cross tabulation with tests for factor independence

Description

An implementation of a cross-tabulation function with output similar to S-Plus `crosstabs()` and SAS Proc Freq (or SPSS format) with Chi-square, Fisher and McNemar tests of the independence of all table factors.

Usage

```

CrossTable(x, y, digits = 3, max.width = NA, expected = FALSE,
           prop.r = TRUE, prop.c = TRUE, prop.t = TRUE,
           prop.chisq = TRUE, chisq = FALSE, fisher = FALSE,
           mcnemar = FALSE, resid = FALSE, sresid = FALSE,
           asresid = FALSE, missing.include = FALSE,
           format = c("SAS", "SPSS"), dnn = NULL, xlab, ylab,
           ...)

```

Arguments

<code>x</code>	A vector or a matrix. If <code>y</code> is specified, <code>x</code> must be a vector.
<code>y</code>	A vector in a matrix or a dataframe.
<code>digits</code>	Number of digits after the decimal point for cell proportions.

<code>max.width</code>	In the case of a 1 x n table, the default will be to print the output horizontally. If the number of columns exceeds <code>max.width</code> , the table will be wrapped for each successive increment of <code>max.width</code> columns. If you want a single column vertical table, set <code>max.width</code> to 1.
<code>prop.r</code>	If TRUE, row proportions will be included.
<code>prop.c</code>	If TRUE, column proportions will be included.
<code>prop.t</code>	If TRUE, table proportions will be included.
<code>expected</code>	If TRUE, <code>chisq</code> will be set to TRUE and expected cell counts from the χ^2 will be included. The value of <code>expected</code> will be set to TRUE if <code>chisq</code> , <code>resid</code> , <code>sresid</code> , <code>asresid</code> or <code>prop.chisq</code> is TRUE.
<code>prop.chisq</code>	If TRUE, chi-square contribution of each cell will be included.
<code>chisq</code>	If TRUE, the results of a chi-square test will be included. The value of <code>chisq</code> will be set to TRUE if <code>expected</code> , <code>resid</code> , <code>sresid</code> or <code>prop.chisq</code> is TRUE.
<code>fisher</code>	If TRUE, the results of a Fisher Exact test will be included.
<code>mcnemar</code>	If TRUE, the results of a McNemar test will be included.
<code>resid</code>	If TRUE, residual (Pearson) will be included. The value of <code>resid</code> will be set to FALSE if <code>format = "SAS"</code> .
<code>sresid</code>	If TRUE, standardized residual will be included. The value of <code>sresid</code> will be set to FALSE if <code>format = "SAS"</code> .
<code>asresid</code>	If TRUE, adjusted standardized residual will be included. The value of <code>asresid</code> will be set to FALSE if <code>format = "SAS"</code> .
<code>missing.include</code>	If TRUE, then remove any unused factor levels.
<code>format</code>	Either SAS (default) or SPSS, depending on the type of output desired.
<code>dnn</code>	the names to be given to the dimensions in the result (the <code>dimnames</code> names).
<code>xlab</code>	A title for the x axis when plotting the CrossTable object (see title).
<code>ylab</code>	A title for the y axis when plotting the CrossTable object (see title).
<code>...</code>	Optional arguments passed to <code>chisq.test</code> .

Details

A summary table will be generated with cell row, column and table proportions and marginal totals and proportions. Expected cell counts can be printed if desired (if `'chisq = TRUE'`). In the case of a 2 x 2 table, both corrected and uncorrected values will be included for appropriate tests. In the case of tabulating a single vector, cell counts and table proportions will be printed.

Note 1: If `'x'` is a vector and `'y'` is not specified, no statistical tests will be performed, even if any are set to TRUE.

Note 2: `'x'` and `'y'` labels will be truncated if the table is not going to fit to the screen, according to the value of `getOption("width")`.

Value

A list of class `CrossTable` containing parameters used by the `print.CrossTable` method and the following components:

`t`: An n by m matrix containing table cell counts.

`prop.row`: An n by m matrix containing cell row proportions.

`prop.col`: An n by m matrix containing cell column proportions.

`prop.tbl`: An n by m matrix containing cell table proportions.

`chisq`: Results from the Chi-Square test. A list with class `'htest'`. See [chisq.test](#) for details.

`chisq.corr`: Results from the corrected Chi-Square test. A list with class `'htest'`. See [chisq.test](#) for details. ONLY included in the case of a 2×2 table.

`fisher.ts`: Results from the two-sided Fisher Exact test. A list with class `'htest'`. See [fisher.test](#) for details. ONLY included if `'fisher' = TRUE`.

`fisher.lt`: Results from the Fisher Exact test with `HA = "less"`. A list with class `'htest'`. See [fisher.test](#) for details. ONLY included if `'fisher' = TRUE` and in the case of a 2×2 table.

`fisher.gt`: Results from the Fisher Exact test with `HA = "greater"`. A list with class `'htest'`. See [fisher.test](#) for details. ONLY included if `'fisher' = TRUE` and in the case of a 2×2 table.

`mcnemar`: Results from the McNemar test. A list with class `'htest'`. See [mcnemar.test](#) for details. ONLY included if `'mcnemar' = TRUE`.

`mcnemar.corr`: Results from the corrected McNemar test. A list with class `'htest'`. See [mcnemar.test](#) for details. ONLY included if `'mcnemar' = TRUE` and in the case of a 2×2 table.

`resid/sresid/asresid`: Pearson Residuals (from chi-square tests).

Author(s)

Jakson Aquino has split the function `CrossTable` (from the package `gmodels`) in two: `CrossTable` and `print.CrossTable`. The `gmodels`'s function was developed by Marc Schwartz (original version posted to `r-devel` on Jul 27, 2002. SPSS format modifications added by Nitin Jain based upon code provided by Dirk Enzmann).

See Also

[forODFTable](#), [table](#), [prop.table](#), [xtabs](#), [crosstab](#).

Examples

```
# Simple cross tabulation of education versus prior induced
# abortions using infertility data
data(Infert, package = "datasets")
CrossTable(Infert$education, Infert$induced, expected = TRUE)
CrossTable(Infert$education, Infert$induced, expected = TRUE,
           format = "SAS")
CrossTable(Infert$education, Infert$induced, expected = TRUE,
           format = "SPSS")
CrossTable(warbreaks$wool, warbreaks$tension,
           dnn = c("Wool", "Tension"))
```

descr	<i>Summary of an object</i>
-------	-----------------------------

Description

Wrapper for the function `summary` of **base** package, including information about variable label. The function prints the `label` attribute of the object and, then, invokes `summary(object)`. If the object is a data frame, the function prints the `label` and invokes `summary` for each variable in the data frame.

Usage

```
descr(x)
```

Arguments

`x` The object to be described.

Value

Null.

Author(s)

Jakson Aquino <jalvesaq@gmail.com>

See Also

[summary](#)

file.head	<i>Prints first lines of a file.</i>
-----------	--------------------------------------

Description

The function prints the first lines of a file, optionally truncating the lines according to the screen width. The lines are truncated at `getOption("width") - 2`.

Usage

```
file.head(file, n, truncate.cols = TRUE)
```

Arguments

`file` Character: The name of the file whose first lines should be printed.
`n` The number of lines to show.
`truncate.cols` Logical: if TRUE truncate the lines.

Value

NULL.

Author(s)

Jakson A. Aquino <jalvesaq@gmail.com>

forODFTable

Convert an object of class CrossTable into a matrix for odfTable

Description

The function converts an object of class `CrossTable` into a matrix to be printed by `'odfTable()'` of `odfWeave` package.

Usage

```
forODFTable(x, digits = 1, ...)
```

Arguments

`x` A object of class `'CrossTable'`.
`digits` See [round](#).
`...` Optional arguments passed to [format](#).

Value

A matrix.

Author(s)

Jakson A. Aquino, <jalvesaq@gmail.com>.

See Also

[CrossTable](#)

Examples

```
## Not run:  
library(odfWeave)  
data(infert, package = "datasets")  
x <- crosstab(infert$education, infert$induced, expected = TRUE)  
  
# Use the function directly:  
odfTable(forODFTable(x))  
  
# Create a method for odfTable:
```

```
odfTable.CrossTable <- function(x) odfTable(forODFTable(x))
odfTable(x)
methods(odfTable)

## End(Not run)
```

freq

Frequency table

Description

Prints a frequency table of the selected object. Optionally, the frequency might be weighted.

Usage

```
freq(x, w, user.missing, plot = getOption("descr.plot"), ...)
```

Arguments

x	The factor from which the frequency of values is desired.
w	An optional vector for a weighted frequency table.
user.missing	Character vector, indicating what levels must be treated as missing values while calculating valid percents. Levels representing user missing values are not shown in the barplot .
plot	Logical: if TRUE (default), a barplot is produced. You may put <code>options(descr.plot = FALSE)</code> in your <code>‘.Rprofile’</code> to change the default function behavior.
...	Further arguments to be passed to plot.freqtable .

Details

A column with cumulative percents are added to the frequency table if x is an ordered factor.

Value

A matrix with class `c("matrix", "freqtable")` with the attribute `"xlab"` which is a character string corresponding to either the attribute `"label"` of x or, if x does not have this attribute, the name of x. The returned object can be plotted, generating a [barplot](#).

Author(s)

Jakson A. Aquino <jalvesaq@gmail.com>, based on function written by Dirk Enzmann

Examples

```
x <- c(rep(1, 100), rep(2, 120), rep(3, 10), rep(NA, 12))
w <- c(rep(1.1, 122), rep(0.9, 120))
x <- factor(x, levels = c(1, 2, 3),
           labels = c("No", "Yes", "No answer"))
attr(x, "label") <- "Do you agree?"

freq(x, y.axis = "percent")
f <- freq(x, w, user.missing = "No answer", plot = FALSE)
f
plot(f)
## Not run:
print(xtable(f))
# If the decimal separator in your country is a comma:
# options(OutDec = ",")
print(xtable(f, align = "lrrr", display = c("s", "d", "f", "f")))

## End(Not run)
```

fromUTF8

Conversion from UTF-8 encoding

Description

Converts the encoding of some attributes of an object from UTF-8 into other encoding.

Usage

```
fromUTF8(x, to = "WINDOWS-1252")
```

Arguments

x	A R object, usually a variable of a data frame or a data frame.
to	A string indicating the desired encoding. Common values are "LATIN1" and "WINDOWS-1252". Type <code>iconvlist()</code> for the complete list of available encodings.

Details

The function converts the attribute `label` of `x` from UTF-8 into the specified encoding. If `x` is a factor, the levels are converted as well. If `x` is a `data.frame`, the function makes the conversions in all of its variables.

Value

The object with its label and levels converted.

Author(s)

Jakson A. Aquino, <jalvesaq@gmail.com>.

See Also

[iconv](#), [label](#)

fwf2csv

Fast conversion of a fwf file into a csv one

Description

Convert fixed width formatted file into a tab separated one.

Usage

```
fwf2csv(fwffile, csvfile, names, begin, end,  
        verbose = getOption("verbose"))
```

Arguments

fwffile	The fixed width format file.
csvfile	The csv file to be created. The fields will be separated by tab characters and there will be no quotes around strings.
names	A character vector with column names.
begin	A numeric vector with the begin offset of values in the fixed width format file.
end	A numeric vector with the end offset of values in the fixed width format file.
verbose	Logical: if TRUE a message about the number of saved lines is printed.

Details

The return value is NULL, but csvfile is created if the function is successful. The file is a text table with fields separated by tabular characters without quotes around the strings. This function is useful if you have a very big fixed width formatted file to read and [read.fwf](#) would be too slow.

Value

NULL.

Author(s)

Jakson A. Aquino <jalvesaq@gmail.com>

Examples

```
## Not run: begin <- c(1, 3, 6, 9, 10, 11, 13)
end <- c(2, 5, 8, 9, 10, 12, 16)
names <- c("state", "municp", "house", "cond", "sex",
           "age", "income")
fwf2csv("example.txt", "example.csv", names, begin, end)
df <- read.table("example.csv", header = TRUE,
                sep = "\t", quote = "")

## End(Not run)
```

histkdnc

Histogram with kernel density and normal curve

Description

Plots a histogram with kernel density and normal curve.

Usage

```
histkdnc(v, breaks = 0, include.lowest = T, right = T,
         main = "Histogram with kernel density and normal curve",
         col = grey(0.90), xlab = deparse(substitute(v)), ...)
```

Arguments

v	The object from which the histogram is desired.
breaks	See hist .
include.lowest	See hist .
right	See hist .
main	See hist .
col	See hist .
xlab	See hist .
...	Further arguments to be passed to hist .

Details

The function plots a histogram of the object x with its kernel density and a normal curve with the same mean and standard deviation of x .

Value

NULL.

Author(s)

Dirk Enzmann (modified by Jakson Aquino<jalvesaq@gmail.com>).

LogRegR2	NA
----------	----

Description

The function calculates multiple R^2 analogues (pseudo R^2) of logistic regression.

Usage

```
LogRegR2(model)
```

Arguments

model A logistic regression model.

Details

The function calculates McFaddens R^2 , Cox & Snell Index, and Nagelkerke Index of a logistic regression model.

Value

A object of class `list` with the calculated indexes.

Author(s)

Dirk Enzmann

Examples

```
# Suppose that "happy" is a factor and "income" is a numeric
# vector in a data frame:
## Not run:
m <- glm(happy ~ income, family=binomial(link="logit"))
LogRegR2(m)

## End(Not run)
```

plot.CrossTable	<i>Mosaic plot from object of class CrossTable</i>
-----------------	--

Description

This function receives a [CrossTable](#) object as its main argument and produces a mosaicplot.

Usage

```
## S3 method for class 'CrossTable'  
plot(x, xlab, ylab, main = "", col,  
      inv.x = FALSE, inv.y = FALSE, ...)
```

Arguments

x	A object of class CrossTable.
xlab	See plot.default .
ylab	See plot.default .
main	See plot.default .
col	A specification for the default plotting color. (See section ‘Color Specification’ of par).
inv.x	A logical value indicating whether the order of the levels of the x variable should be inverted.
inv.y	A logical value indicating whether the order of the levels of the y variable should be inverted.
...	Further arguments to be passed to mosaicplot .

Author(s)

Jakson A. Aquino <jalvesaq@gmail.com>

See Also

[CrossTable](#), [crosstab](#).

plot.freqtable	<i>Bar plot from object of class freqtable</i>
----------------	--

Description

This function receives a freqtable object as its main argument and produces a barplot.

Usage

```
## S3 method for class 'freqtable'
plot(x, y.axis = "count", ...)
```

Arguments

x	A object of class freqtable.
y.axis	Character string, indicating what variable to use in the y axis, "count" or "percent", when plotting the frequency table.
...	Further arguments to be passed to barplot .

Author(s)

Jakson A. Aquino <jalvesaq@gmail.com>

toUTF8	<i>Conversion to UTF-8 encoding</i>
--------	-------------------------------------

Description

Converts the encoding of some attributes of an object to UTF-8

Usage

```
toUTF8(x, from = "WINDOWS-1252")
```

Arguments

x	A R object, usually a variable of a data frame or a data frame.
from	A string indicating the original encoding. Common values are "LATIN1" and "WINDOWS-1252". Type iconvlist() for the complete list of available encodings.

Details

The function converts the attribute `label` of `x` from the specified encoding into UTF-8. If `x` is a factor, the levels are converted as well. If `x` is a `data.frame`, the function makes the conversions in all of its variables.

Value

The object with its label and levels converted.

Author(s)

Jakson A. Aquino, <jalvesaq@gmail.com>.

See Also

[iconv](#), [label](#)

xtable.CrosTable	<i>CrosTable method for xtable</i>
------------------	------------------------------------

Description

The method creates an object of class xtable.

Usage

```
## S3 method for class 'CrosTable'
xtable(x, multirow = FALSE, hline = FALSE,
       caption = NULL, label = NULL, align = NULL, digits = 1,
       display = NULL, ...)
```

Arguments

x	A object of class CrossTable.
multirow	A logical value indicating whether the command \multirow should be added to the table. See the Details section below.
hline	A logical value indicating whether the command \hline should be added to the table. See the Details section below.
caption	See xtable .
label	See xtable .
align	See xtable .
display	See xtable .
digits	See round .
...	Further arguments to be passed to format .

Details

If either multirow or hline is TRUE, the `sanitize.text.function` argument of [print.xtable](#) must be defined. You will also have to add `\usepackage{multirow}` to your Rnoweb document. See the Example section of [crosstab](#).

Author(s)

Jakson A. Aquino <jalvesaq@gmail.com>

See Also

[CrossTable](#), [crosstab](#), [print.xtable](#).

Index

barplot, [10](#), [16](#)
boxplot, [2](#)

chisq.test, [6](#), [7](#)
compmeans, [2](#)
crosstab, [3](#), [7](#), [15](#), [17](#), [18](#)
CrossTable, [3](#), [4](#), [5](#), [9](#), [15](#), [18](#)

descr, [8](#)

file.head, [8](#)
fisher.test, [7](#)
format, [9](#), [17](#)
forODFTable, [7](#), [9](#)
freq, [10](#)
fromUTF8, [11](#)
fwf2csv, [12](#)

hist, [13](#)
histkdnc, [13](#)

iconv, [12](#), [17](#)

label, [12](#), [17](#)
LogRegR2, [14](#)

mcnemar.test, [7](#)
mosaicplot, [4](#), [15](#)

par, [15](#)
plot.CrossTable, [4](#), [15](#)
plot.default, [15](#)
plot.freqtable, [10](#), [16](#)
print.xtable, [17](#), [18](#)
prop.table, [7](#)

read.fwf, [12](#)
round, [9](#), [17](#)

summary, [8](#)

table, [7](#)

title, [4](#), [6](#)
toUTF8, [16](#)

wtd.boxplot, [2](#)

xtable, [17](#)
xtable.CrossTable, [4](#), [17](#)
xtabs, [7](#)