

Package ‘descr’

October 11, 2009

Version 0.3.2

Date 2009-10-10

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

Depends xtable

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

License GPL (>= 2)

URL <http://jalvesaq.googlepages.com/descr>

Repository CRAN

Date/Publication 2009-10-11 14:41:11

R topics documented:

compmeans	2
crosstab	3
CrossTable	4
descr	6
file.head	7
freq	8
fromUTF8	9
fwf2csv	10
histkdnc	11
LogRegR2	12
toUTF8	12
Index	14

compmeans

Means of a numerical vector according to a factor with boxplot

Description

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

Usage

```
compmeans(x, f, w, sort = FALSE, maxlevels = 60,  
          plot = getOption("descr.plot"), xlab = deparse(substitute(f)),  
          ylab = deparse(substitute(x)), ...)
```

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.
plot	Logical: if TRUE (default), a boxplot is produced. You may put options (descr.plot = FALSE) in your '.Rprofile' to change the default function behavior.
xlab	A title for the x axis (see title).
ylab	A title for the y axis (see title).
...	Further arguments to be passed to boxplot .

Value

A object of class table.

Author(s)

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

crosstab	<i>Cross tabulation with mosaic plot</i>
----------	--

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 = 5, 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, plot = getOption("descr.plot"),
         main = "", xlab = deparse(substitute(x)), ylab = deparse(substitute(y)),
         col = gray.colors(length(levels(y)), 0.9, 0.3), ...)
```

Arguments

<code>x, 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 .
<code>expected</code>	See CrossTable .
<code>prop.r</code>	See CrossTable .
<code>prop.c</code>	See CrossTable .
<code>prop.t</code>	See CrossTable .
<code>prop.chisq</code>	See CrossTable .
<code>chisq</code>	See CrossTable .
<code>fisher</code>	See CrossTable .
<code>mcnemar</code>	See CrossTable .
<code>resid</code>	See CrossTable .
<code>sresid</code>	See CrossTable .
<code>asresid</code>	See CrossTable .
<code>missing.include</code>	See CrossTable .
<code>format</code>	See CrossTable .
<code>dnn</code>	See CrossTable .
<code>plot</code>	Logical: if TRUE (default), a mosaic plot is produced. You may put options (<code>descr.plot = FALSE</code>) in your <code>‘.Rprofile’</code> to change the default function behavior.

<code>main</code>	An overall title for the plot (see title).
<code>xlab</code>	A title for the x axis (see title).
<code>ylab</code>	A title for the y axis (see title).
<code>col</code>	A specification for the default plotting color. (See section ‘Color Specification’ of par).
<code>...</code>	Further arguments to be passed to mosaicplot .

Details

`crosstab` invokes the [CrossTable](#) function in the **gmodels** package with all boolean options set to `FALSE` and `"SPSS"` as the default `format` option.

Author(s)

Jakson A. Aquino (jalvesaq@gmail.com)

See Also

[CrossTable](#)

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 = 5, 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, ...)
```

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>expected</code>	If TRUE, <code>chisq</code> will be set to TRUE and expected cell counts from the χ^2 will be included
<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>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
<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
<code>sresid</code>	If TRUE, standardized residual will be included
<code>asresid</code>	If TRUE, adjusted standardized residual will be included
<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>...</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: A limited `xtable` method is provided by the package; expected values, proportions and other information are not included in the generated table.

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 `'hstest'`. See `?chisq.test` for details

`chisq.corr`: Results from the corrected Chi-Square test. A list with class `'hstest'`. See `?chisq.test` for details. ONLY included in the case of a 2 x 2 table.

`fisher.ts`: Results from the two-sided Fisher Exact test. A list with class `'hstest'`. 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 `'hstest'`. See `?fisher.test` for details. ONLY included if `'fisher' = TRUE` and in the case of a 2 x 2 table.

`fisher.gt`: Results from the Fisher Exact test with `HA = "greater"`. A list with class `'hstest'`. See `?fisher.test` for details. ONLY included if `'fisher' = TRUE` and in the case of a 2 x 2 table.

`mcnemar`: Results from the McNemar test. A list with class `'hstest'`. See `?mcnemar.test` for details. ONLY included if `'mcnemar' = TRUE`.

`mcnemar.corr`: Results from the corrected McNemar test. A list with class `'hstest'`. See `?mcnemar.test` for details. ONLY included if `'mcnemar' = TRUE` and in the case of a 2 x 2 table.

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

Author(s)

Jakson Aquino has splited 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

`table`, `xtabs`, `prop.table`

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(warfbreaks$wool, warfbreaks$tension, dnn = c("Wool", "Tension"))
```

descr

Summary of an object

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`

Prints first lines of a file.

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)

`freq`*Frequency table with barplot*

Description

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

Usage

```
freq(x, w, plot = getOption("descr.plot"), y.axis = c("count", "percent"), ...)
```

Arguments

<code>x</code>	The factor from which the frequency of values is desired.
<code>w</code>	An optional vector for a weighted frequency table.
<code>plot</code>	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.
<code>y.axis</code>	Character string, indicating what variable to use in the y axis: "count" or "percent".
<code>...</code>	Further arguments to be passed to <code>barplot</code> .

Details

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

Value

A list with class `freqtable` containing the following components:

<code>label</code>	A character string corresponding to either the attribute "label" of <code>x</code> or, if <code>x</code> does not have this attribute, the name of <code>x</code> .
<code>freqtable</code>	A matrix with frequencies, sums and standard deviations.

Author(s)

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

`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

<code>x</code>	A R object, usually a variable of a data frame or a data frame.
<code>to</code>	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⟩](mailto: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)
```

Arguments

<code>fwffile</code>	The fixed width format file.
<code>csvfile</code>	The csv file to be created. The fields will be separated by tab characters and there will be no quotes around strings.
<code>names</code>	A character vector with column names.
<code>begin</code>	A numeric vector with the begin offset of values in the fixed width format file.
<code>end</code>	A numeric vector with the end offset of values in the fixed width format file.

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

<code>v</code>	The object from which the histogram is desired.
<code>breaks</code>	See hist .
<code>include.lowest</code>	See hist .
<code>right</code>	See hist .
<code>main</code>	See hist .
<code>col</code>	See hist .
<code>xlab</code>	See hist .
<code>...</code>	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`*Pseudo R² of logistic regression*

Description

The function calculates multiple R² analogues (pseudo R²) of logistic regression.

Usage

```
LogRegR2(model)
```

Arguments

`model` A logistic regression model.

Details

The function calculates McFaddens R², 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)
```

`toUTF8`*Conversion to UTF-8 encoding*

Description

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

Usage

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

Arguments

<code>x</code>	A R object, usually a variable of a data frame or a data frame.
<code>from</code>	A string indicating the original 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 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⟩](mailto:jalvesaq@gmail.com).

See Also

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

Index

*Topic **category**
 CrossTable, 4

*Topic **file**
 compmeans, 2
 crosstab, 3
 descr, 6
 file.head, 7
 freq, 8
 fromUTF8, 9
 fwf2csv, 10
 histkdnc, 11
 LogRegR2, 12
 toUTF8, 12

*Topic **univar**
 CrossTable, 4

barplot, 8
boxplot, 2

chisq.test, 5
compmeans, 2
crosstab, 3
CrossTable, 3, 4, 4

descr, 6

file.head, 7
freq, 8
fromUTF8, 9
fwf2csv, 10

hist, 11
histkdnc, 11

iconv, 9, 13

label, 9, 13
LogRegR2, 12

mosaicplot, 4

par, 4

prop.table, 6

read.fwf, 10

summary, 7

table, 6
title, 2, 4
toUTF8, 12

xtabs, 6