

Package ‘compareGroups’

May 13, 2012

Type Package

Title descriptives by groups

Version 1.1

Date 2012-05-13

Depends R (>= 2.13.1), xtable, Hmisc, gdata, survival, oce, epitools

Suggests tcltk2, car

Author Isaac Subirana, Héctor Sanz, Joan Vila

Maintainer Isaac Subirana <isubirana@imim.es>

Description This package is meant to facilitate the construction of bivariate table (descriptives of several variables between groups) and generate reports in LaTeX, HTML or (CSV) plain text. Also, it includes a graphical interface (GUI) to build these tables more easily.

License GPL (>= 2)

LazyLoad yes

Encoding latin1

BuildVignettes False

Repository CRAN

Date/Publication 2012-05-13 14:48:09

R topics documented:

compareGroups-package	2
cGroupsGUI	2
compareGroups	3
createTable	9
export2csv	12

export2html	13
export2latex	14
regicor	16
varinfo	17

Index	19
--------------	-----------

compareGroups-package *Descriptives analysis by groups*

Description

A package which quickly and efficiently generates descriptives by groups (bivariate tables) in plain text, HTML or as LATEX code. Depending on the nature of the variable, different descriptive statistics can be calculated (mean, median, frequencies, incidences) and different tests are computed as appropriate (t-test, ANOVA, Kruskal-Wallis, Fisher, log-rank, ...). In order to facilitate construction of the bivariate table, we have implemented a graphical user interface (GUI).

Details

Package:	compareGroups
Type:	Package
Version:	1.0
Date:	2011-10-04
License:	GPL version 2 or newer
LazyLoad:	yes

Main functions: `compareGroups`, `createTable`, `export2latex`, `export2html`, `export2csv`

Author(s)

Héctor Sanz <hsanz@imim.es>, Isaac Subirana <isubirana@imim.es> and Joan Vila <jvila@imim.es>

Maintainer: Isaac Subirana <isubirana@imim.es>

cGroupsGUI	<i>Graphical user interface to create tables of descriptives by groups (bivariate tables)</i>
------------	---

Description

This function allows the user to build tables in an easy and intuitive way and to modify several options, using a graphical interface.

Usage

```
cGroupsGUI(X)
```

Arguments

X a matrix or a data.frame. 'X' must exist in .GlobalEnv. If missing, [regicor](#) data is loaded

Details

See the vignette for more detailed examples illustrating the use of this function.

Note

If a data.frame or a matrix is passed through 'X' argument or is loaded by the 'Load data' GUI menu, this object is placed in the .GlobalEnv. Manipulating this data.frame or matrix while GUI is opened may produce an error in executing the GUI operations.

See Also

[compareGroups](#), [createTable](#)

Examples

```
## Not run:  
data(regicor)  
cGroupsGUI(regicor)  
  
## End(Not run)
```

compareGroups	<i>Descriptives by groups</i>
---------------	-------------------------------

Description

This function performs descriptives by groups for several variables. Depending on the nature of these variables, different descriptive statistics are calculated (mean, median, frequencies or K-M probabilities) and different tests are computed as appropriate (t-test, ANOVA, Kruskal-Wallis, Fisher, log-rank, ...).

Usage

```
compareGroups(X, ...)  
## Default S3 method:  
compareGroups(X, y = NULL, Xext = NULL, selec = NA, method = 1, timemax = NA,  
alpha = 0.05, min.dis = 5, max.ylev = 5, max.xlev = 10, include.label = TRUE,  
Q1 = 0.25, Q3 = 0.75, simplify = TRUE, ref = 1, ref.no = NA, fact.ratio = 1,  
ref.y = 1, p.corrected = TRUE, compute.ratio = TRUE, ...)
```

```
## S3 method for class 'formula'
compareGroups(X, data, subset, na.action=NULL, include.label=TRUE, ...)
## S3 method for class 'compareGroups'
plot(x, file, bivar = FALSE, z=1.5, n.breaks = "Sturges", ...)
```

Arguments

X	either a data.frame or a matrix (then method 'compareGroups.default' is called), or a formula (then method 'compareGroups.formula' is called). When 'X' is a formula, it must be an object of class "formula" (or one that can be coerced to that class). Right side of ~ must have the terms in an additive way, and left side of ~ must contain the name of the grouping variable or can be left in blank (in this latter case descriptives for whole sample are calculated and no test is performed).
y	a vector variable that distinguishes the groups. It must be either a numeric, character, factor or NULL. Default value is NULL which means that descriptives for whole sample are calculated and no test is performed.
Xext	a data.frame or a matrix with the same rows / individuals contained in X, and maybe with different variables / columns than X. This argument is used by compareGroups.default in the sense that the variables specified in the argument selec are searched in Xext and/or in the <code>.GlobalEnv</code> . If Xext is NULL, then Xext is created from variables of X plus y. Default value is NULL.
selec	a list with as many components as row-variables. If list length is 1 it is recycled for all row-variables. Every component of 'selec' is an expression that will be evaluated to select the individuals to be analyzed for every row-variable. Otherwise, a named list specifying 'selec' row-variables is applied. '.else' is a reserved name that defines the selection for the rest of the variables; if no '.else' variable is defined, default value is applied for the rest of the variables. Default value is NA; all individuals are analyzed (no subsetting).
method	integer vector with as many components as row-variables. If its length is 1 it is recycled for all row-variables. It only applies for continuous row-variables (for factor row-variables it is ignored). Possible values are: 1 - forces analysis as "normal-distributed"; 2 - forces analysis as "continuous non-normal"; 3 - forces analysis as "categorical"; and 4 - NA, which performs a Shapiro-Wilks test to decide between normal or non-normal. Otherwise, a named vector specifying 'method' row-variables is applied. '.else' is a reserved name that defines the method for the rest of the variables; if no '.else' variable is defined, default value is applied. Default value is 1.
timemax	double vector with as many components as row-variables. If its length is 1 it is recycled for all row-variables. It only applies for 'Surv' class row-variables (for all other row-variables it is ignored). This value indicates at which time the K-M probability is to be computed. Otherwise, a named vector specifying 'timemax' row-variables is applied. '.else' is a reserved name that defines the 'timemax' for the rest of the variables; if no '.else' variable is defined, default value is applied. Default value is NA; K-M probability is then computed at the median of observed times.
alpha	double between 0 and 1. Significance threshold for the <code>shapiro.test</code> normality test for continuous row-variables. Default value is 0.05.

min.dis	an integer. If a non-factor row-variable contains less than 'min.dis' different values and 'method' argument is set to NA, then it will be converted to a factor. Default value is 5.
max.ylev	an integer indicating the maximum number of levels of grouping variable ('y'). If 'y' contains more than 'max.ylev' levels, then the function 'compareGroups' produces an error. Default value is 5.
max.xlev	an integer indicating the maximum number of levels when the row-variable is a factor. If the row-variable is a factor (or converted to a factor if it is a character, for example) and contains more than 'max.xlev' levels, then it is removed from the analysis and a warning is printed. Default value is 10.
data	an optional data frame, list or environment (or object coercible by 'as.data.frame' to a data frame) containing the variables in the model. If they are not found in 'data', the variables are taken from 'environment(formula)'.
subset	an optional vector specifying a subset of individuals to be used in the computation process. It is applied to all row-variables. 'subset' and 'selec' are added in the sense of '&' to be applied in every row-variable.
na.action	a function which indicates what should happen when the data contain NAs. The default is NULL, and that is equivalent to na.pass , which means no action. Value na.exclude can be useful if it is desired to removed all individuals with some NA in any variable.
include.label	logical, indicating whether or not variable labels have to be shown in the results. Default value is TRUE
Q1	double between 0 and 1, indicating the quantile to be displayed as the first number inside the square brackets in the bivariate table. To compute the minimum just type 0. Default value is 0.25 which means the first quartile.
Q3	double between 0 and 1, indicating the quantile to be displayed as the second number inside the square brackets in the bivariate table. To compute the maximum just type 1. Default value is 0.75 which means the third quartile.
simplify	logical, indicating whether levels with no values must be removed for grouping variable and for row-variables. Default value is TRUE.
ref	an integer vector with as many components as row-variables. If its length is 1 it is recycled for all row-variables. It only applies for categorical row-variables. Or a named vector specifying which row-variables 'ref' is applied (a reserved name is '.else' which defines the reference category for the rest of the variables); if no '.else' variable is defined, default value is applied for the rest of the variables. Default value is 1.
ref.no	character specifying the name of the level to be the reference for Odds Ratio or Hazard Ratio. This is especially useful for yes/no variables. Default value is NA which means that category specified in 'ref' is the one selected to be the reference.
fact.ratio	a double vector with as many components as row-variables indicating the units for the HR / OR (note that it does not affect the descriptives). If its length is 1 it is recycled for all row-variables. Otherwise, a named vector specifying 'fact.ratio' row-variables is applied. '.else' is a reserved name that defines the reference category for the rest of the variables; if no '.else' variable is defined, default value is applied. Default value is 1.

<code>ref.y</code>	an integer indicating the reference category of y variable for computing the OR, when y is a binary factor. Default value is 1.
<code>p.corrected</code>	logical, indicating whether p-values for pairwise comparisons must be corrected. It only applies when there is a grouping variable with more than 2 categories. Default value is TRUE.
<code>compute.ratio</code>	logical, indicating whether Odds Ratio (for a binary response) or Hazard Ratio (for a time-to-event response) must be computed. Default value is TRUE.
<code>x</code>	an object of class 'compareGroups'.
<code>file</code>	a character string giving the name of the file. A pdf file is saved with an appendix added to 'file' corresponding to the row-variable name. If missing, multiple devices are opened, one for each row-variable of 'x' object.
<code>bivar</code>	logical. If <code>bivar=TRUE</code> , it plots a boxplot or a barplot (for a continuous or categorical row-variable, respectively) stratified by groups. If <code>bivar=FALSE</code> , it plots a normality plot (for continuous row-variables) or a barplot (for categorical row-variables). Default value is FALSE.
<code>z</code>	double. Indicates threshold limits to be placed in the deviation from normality plot. It is considered that too many points beyond this threshold indicates that current variable is far to be normal-distributed. Default value is 1.5.
<code>n.breaks</code>	same as argument 'breaks' of hist .
<code>...</code>	further arguments passed to 'compareGroups.default' or other methods.

Details

Depending whether the row-variable is considered as continuous normal-distributed (1), continuous non-normal distributed (2) or categorical (3), the following descriptives and tests are performed:

1- mean, standard deviation and t-test or ANOVA

2- median, 1st and 3rd quartiles (by default), and Kruskal-Wallis test

3- or absolute and relative frequencies and chi-squared or exact Fisher test when the expected frequencies is less than 5 in some cell

Also, a row-variable can be of class 'Surv'. Then the probability of 'event' at a fixed time (set up with 'timemax' argument) is computed and a logrank test is performed.

When there are more than 2 groups, it also performs pairwise comparisons adjusting for multiple testing (Tukey when row-variable is normal-distributed and Benjamini & Hochberg method otherwise), and computes p-value for trend. The p-value for trend is computed from the Pearson test when row-variable is normal and from the Spearman test when it is continuous non normal. If row-variable is of class 'Surv', the score test is computed from a Cox model where the grouping variable is introduced as an integer variable predictor. If the row-variable is categorical, the p-value for trend is computed as

$$1 - \text{pchisq}(\text{cor}(\text{as.integer}(x), \text{as.integer}(y))^{2 * (\text{length}(x) - 1)}, 1)$$

where 'x' is the row-variable and 'y' is the grouping variable.

If there are two groups, the Odds Ratio is computed for each row-variable. While, if the response is of class 'Surv' (i.e. time to event) Hazard Ratios are computed.

The p-values for Hazard Ratios are computed using the logrank or Wald test under a Cox proportional hazard regression when row-variable is categorical or continuous, respectively.

See the vignette for more detailed examples illustrating the use of this function and the methods used.

Value

An object of class 'compareGroups'.

'print' returns a table sample size, overall p-values, type of variable ('categorical', 'normal', 'non-normal' or 'Surv') and the subset of individuals selected.

'summary' returns a much more detailed list. Every component of the list is the result for each row-variable, showing frequencies, mean, standard deviations, quartiles or K-M probabilities as appropriate. Also, it shows overall p-values as well as p-trends and pairwise p-values among the groups.

'plot' displays, for all the analyzed variables, normality plots (with the Shapiro-Wilks test), barplots or Kaplan-Meier plots depending on whether the row-variable is continuous, categorical or time-to-response, respectively. Also, bivariate plots can be displayed with stratified by groups boxplots or barplots, setting 'bivar' argument to TRUE.

An update method for 'compareGroups' objects has been implemented and works as usual to change all the arguments of previous analysis.

A subset, '[', method has been implemented for 'compareGroups' objects. The subsetting indexes can be either integers (as usual), row-variables names or row-variable labels.

Combine by rows, 'rbind', method has been implemented for 'compareGroups' objects. It is useful to distinguish row-variable groups.

See examples for further illustration about all previous issues.

Note

Arguments 'X', 'y' and 'Xext' from the compareGroups.default method are not recommended to be used. Use 'X', 'data' and 'subset' arguments from the compareGroups.formula method instead.

By default, the labels of the variables (row-variables and grouping variable) are displayed in the resulting tables. These labels are taken from the "label" attribute of each variable. And if this attribute is NULL, then the name of the variable is displayed, instead. To label non-labeled variables, or to change their labels, use the function [label](#).

There may be no equivalence between the intervals of the OR / HR and p-values. For example, when the response variable is binary and the row-variable is continuous, p-value is based on Mann-Whitney U test or t-test depending on whether row-variable is normal distributed or not, respectively, while the confidence interval is build using the Wald method ($\log(\text{OR}) \pm 1.95 \cdot \text{se}$). Or when the answer is of class 'Surv', p-value is computed with the logrank test, while confidence intervals are based on the Wald method ($\log(\text{HR}) \pm 1.95 \cdot \text{se}$). Finally, when the response is binary and the row variable is categorical, the p-value is based on the chi-squared or Fisher test when appropriate, while confidence intervals are constructed from the median-unbiased estimation method (see [oddsratio](#)).

Subjects selection criteria specified in 'selec' and 'subset' arguments are combined using '&' to be applied to every row-variable.

Currently, 'plot' method only saves in pdf format.

See Also

[createTable](#)

Examples

```
require(compareGroups)

# load REGICOR data
data(regicor)

# compute a time-to-cardiovascular event variable
regicor$tcv <- with(regicor, Surv(tocv, as.integer(cv=='Yes')))
label(regicor$tcv)<-"Cardiovascular"

# compute a time-to-overall death variable
regicor$death <- with(regicor, Surv(todeath, as.integer(death=='Yes')))
label(regicor$death) <- "Mortality"

# descriptives by sex
res <- compareGroups(sex ~ .-id-tocv-cv-todeath-death, data = regicor)
res

# summary of the first 4 row-variables
summary(res[1:4])

# univariate plots of all row-variables
## Not run:
plot(res)

## End(Not run)

# plot of all row-variables by sex
## Not run:
```

```

plot(res, bivar = TRUE)

## End(Not run)

# update changing the response: time-to-cardiovascular event.
# note that time-to-death must be removed since it is not possible
# not compute descriptives of a 'Surv' class object by another 'Surv' class object.
update(res, tcv ~ . + sex - tdeath - tcv)

```

createTable	<i>Table of descriptives by groups: bivariate table</i>
-------------	---

Description

This functions builds a "compact" and "nice" table with the descriptives by groups.

Usage

```

createTable(x, hide = NA, digits = NA, type = NA, show.p.overall = TRUE,
            show.all, show.p.trend, show.p.mul = FALSE, show.n, show.ratio =
            FALSE, show.descr = TRUE, hide.no = NA, digits.ratio = NA)
## S3 method for class 'createTable'
print(x, which.table = "descr", nmax = TRUE, ...)

```

Arguments

x	an object of class 'compareGroups'
hide	a vector (or a list) with integers or characters with as many components as row-variables. If its length is 1 it is recycled for all row-variables. Each component specifies which category (the literal name of the category if it is a character, or the position if it is an integer) must be hidden and not shown. This argument only applies to categorical row-variables, and for continuous row-variables it is ignored. If NA, all categories are displayed. Or a named vector (or a named list) specifying which row-variables 'hide' is applied, and for the rest of row-variables default value is applied. Default value is NA.
digits	an integer vector with as many components as row-variables. If its length is 1 it is recycled for all row-variables. Each component specifies the number of significant decimals to be displayed. Or a named vector specifying which row-variables 'digits' is applied (a reserved name is '.else' which defines 'digits' for the rest of the variables); if no '.else' variable is defined, default value is applied for the rest of the variables. Default value is NA which puts the 'appropriate' number of decimals (see vignette for further details).
type	an integer that indicates whether absolute and/or relative frequencies are displayed: 1 - only relative frequencies; 2 or NA - absolute and relative frequencies in brackets; 3 - only absolute frequencies.

<code>show.p.overall</code>	logical indicating whether p-value of overall groups significance ('p.overall' column) is displayed or not. Default value is TRUE.
<code>show.all</code>	logical indicating whether the '[ALL]' column (all data without stratifying by groups) is displayed or not. Default value is FALSE if grouping variable is defined, and FALSE if there are no groups.
<code>show.p.trend</code>	logical indicating whether p-trend is displayed or not. It is always FALSE when there are less than 3 groups. If this argument is missing, there are more than 2 groups and the grouping variable is an ordered factor, then p-trend is displayed. By default, p-trend is not displayed, and it is displayed when there are more than 2 groups and the grouping variable is of class ordered-factor.
<code>show.p.mul</code>	logical indicating whether the pairwise (between groups) comparisons p-values are displayed or not. It is always FALSE when there are less than 3 groups. Default value is FALSE.
<code>show.n</code>	logical indicating whether number of individuals analyzed for each row-variable is displayed or not in the 'descr' table. Default value is FALSE and it is TRUE when there are no groups.
<code>show.ratio</code>	logical indicating whether OR / HR is displayed or not. Default value is FALSE.
<code>show.descr</code>	logical indicating whether descriptives (i.e. mean, proportions, ...) are displayed. Default value is TRUE.
<code>hide.no</code>	character specifying the name of the level to be hidden for all categorical variables with 2 categories. The result is one row for the variable with only the name displayed and not the category. This is especially useful for yes/no variables. It is ignored for the categorical row-variables with 'hide' argument different from NA. Default value is NA which means that no category is hidden.
<code>digits.ratio</code>	The same as 'digits' argument but applied for the Hazard Ratio or Odds Ratio.
<code>which.table</code>	character indicating which table is printed. Possible values are 'descr', 'avail' or 'both' (partial matching allowed), printing descriptives by groups table, availability data table or both tables, respectively. Default value is 'descr'.
<code>nmax</code>	logical, indicating whether to show the maximum number of responses ("n" of a complete, or least-missing, variable) available by group. Default value is TRUE.
<code>...</code>	other arguments passed to <code>print.default</code> .

Value

An object of class 'createTable', which contains a list of 2 matrix:

<code>descr</code>	a character matrix of descriptives for all row-variables by groups and p-values in a 'compact' format
<code>avail</code>	a character matrix indicating the number of available data for each group, the type of variable (categorical, continuous-normal or continuous-non-normal) and the individuals selection made (if non selection 'ALL' is displayed).

'print' prints these two tables in a 'nice' format.

'summary' prints the 'available' info table (it is a short form of `print(x, which.table = 'avail')`).

'update' modifies previous results from 'createTable'.

subsetting, '[' , can also be applied to 'createTable' objects in the same way as 'compareGroups' objects.

combine by rows, 'rbind', method can be applied to 'createTable' objects, but only if all 'createTable' objects have the same columns. It is useful to distinguish row-variable groups. The resulting object is of class 'rbind.createTable' and 'createTable'.

combine by columns, 'cbind', method can be applied to 'createTable' objects, but only if all 'createTable' objects have the same rows. It may be used when combining different tables referring to different subsets of people (for example, men and women). The resulting object is of class 'cbind.createTable' and 'createTable' and has its own 'print' method.

See the vignette for more details.

See Also

[compareGroups](#), [export2latex](#), [export2csv](#), [export2html](#)

Examples

```
require(compareGroups)

# load REGICOR data
data(regicor)

# compute a time-to-cardiovascular event variable
regicor$tcv <- with(regicor, Surv(tocv, as.integer(cv=='Yes')))
label(regicor$tcv) <- "Cardiovascular incidence"

# compute a time-to-overall death variable
regicor$death <- with(regicor, Surv(todeath, as.integer(death=='Yes')))
label(regicor$death) <- "Mortality incidence"

# remove variables used to create time-to variables
regicor <- remove.vars(regicor, c("todeath", "death", "tocv", "cv"))

# descriptives by time-to-cardiovascular event, taking 'no' category as
# the reference in computing HRs.
res <- compareGroups(tcv ~ .-id-death, regicor, ref.no='no')

# build table showing HR and hiding the 'no' category
restab <- createTable(res, show.ratio = TRUE, hide.no = 'no')
restab

# prints available info table
summary(restab)

# Adds the 'available data' column
update(restab, show.n=TRUE)

# Descriptive of the entire cohort
update(restab, x = update(res, ~ . ))
```

```

# .. changing the response variable to sex
# Odds Ratios (OR) are displayed instead of Hazard Ratios (HR).
# note that now it is possible to compute descriptives by time-to-death
# or time-to-cv but not the ORs .
# We set timemax to 5 years, to report the probability of death and CV at 5 years:
update(restab, x = update(res, sex ~ . - sex + tdeath + tcv, timemax = 5*365.25))

## Combining tables:

# a) By rows: takes the first four variables as a group and the rest as another group:
rbind("First group of variables"=restab[1:4],"Second group of variables"=
  restab[5:length(res)])

# b) By columns: puts stratified tables by sex one beside the other:
res1<-compareGroups(year ~ . - id - sex, regicor)
restab1<-createTable(res1, hide.no = 'no')
restab2<-update(restab1, x = update(res1, subset = sex == 'Male'))
restab3<-update(restab1, x = update(res1, subset = sex == 'Female'))
cbind("ALL" = restab1, "MALES" = restab2, "FEMALES" = restab3)

```

export2csv

Exporting descriptives table to plain text (CSV) format

Description

This function takes the result of `createTable` and exports the tables to plain text (CSV) format.

Usage

```
export2csv(x, file, which.table="descr", sep=",", nmax = TRUE, ...)
```

Arguments

<code>x</code>	an object of class <code>'createTable'</code> .
<code>file</code>	file where table in CSV format will be written. Note that the name of the folder and the basename of the file must be written but not the extension (.csv). Also, another file with the extension <code>'_appendix'</code> is written with the available data table.
<code>which.table</code>	character indicating which table is printed. Possible values are <code>'descr'</code> , <code>'avail'</code> or <code>'both'</code> (partial matching allowed), exporting descriptives by groups table, available data table or both tables, respectively. Default value is <code>'descr'</code> .
<code>sep</code>	character. The variable separator. If <code>'sep'</code> is set to <code>'</code> , it uses <code>write.csv</code> , and when is set to <code>;</code> it uses <code>write.csv2</code> . Default value is <code>'</code> .
<code>nmax</code>	logical, indicating whether to show the maximum number of responses ("n" of a complete, or least-missing, variable) available by group. Default value is <code>TRUE</code> .
<code>...</code>	other arguments passed to <code>write.table</code> .

See Also

[createTable](#), [export2latex](#), [export2html](#)

Examples

```
## Not run:
require(compareGroups)
data(regicor)
res <- compareGroups(sex ~. -id-todeath-death-tocv-cv, regicor)
export2csv(createTable(res, hide.no = 'n'), file="table1")

## End(Not run)
```

export2html

Exporting descriptives table to HTML format

Description

This function takes the result of `createTable` and exports the tables to HTML format.

Usage

```
export2html(x, file, which.table="descr", nmax = TRUE, ...)
```

Arguments

<code>x</code>	an object of class 'createTable'.
<code>file</code>	file where table in HTML format will be written. Note that the the folder and the basename of the file must be written but not the extension (.html). Also, another file with the extension '_appendix' is written with the available data table.
<code>which.table</code>	character indicating which table is printed. Possible values are 'descr', 'avail' or 'both' (partial matching allowed), exporting descriptives by groups table, availability data table or both tables, respectively. Default value is 'descr'.
<code>nmax</code>	logical, indicating whether to show the maximum number of responses ("n" of a complete, or least-missing, variable) available by group. Default value is TRUE.
<code>...</code>	currently ignored.

See Also

[createTable](#), [export2latex](#), [export2csv](#)

Examples

```
## Not run:
require(compareGroups)
data(regicor)
res <- compareGroups(sex ~. -id-todeath-death-tocv-cv, regicor)
export2html(createTable(res, hide.no = 'n'), file="table1")

## End(Not run)
```

export2latex

Exporting descriptives table to LaTeX format

Description

This function takes the result of `createTable` and exports the tables to LaTeX format.

Usage

```
export2latex(x, ...)
## S3 method for class 'createTable'
export2latex(x, file, which.table = 'descr', size = 'same',
             nmax = TRUE, caption = NULL, loc.caption = 'top', label = NULL, ...)
## S3 method for class 'cbind.createTable'
export2latex(x, file, which.table = 'descr', size = 'same',
             nmax = TRUE, caption = NULL, loc.caption = 'top', label = NULL, ...)
```

Arguments

<code>x</code>	an object of class 'createTable'.
<code>file</code>	Name of file where the resulting code should be saved. If file is missing, output is displayed on screen. Note that the folder and the basename of the file must be written but not the extension (.tex). Also, another file with the extension '_appendix' is written with the available data table.
<code>which.table</code>	character indicating which table is exported. Possible values are 'descr', 'avail' or 'both' (partial matching allowed), exporting descriptives by groups table, availability data table or both tables, respectively. Default value is 'descr'.
<code>size</code>	character indicating the size of the table elements. Possible values are: 'tiny', 'scriptsize', 'footnotesize', 'small', 'normalsize', 'large', 'Large', 'LARGE', 'huge', 'Huge' or 'same' (partial matching allowed). Default value is 'same' which means that font size of the table is the same as specified in the main LaTeX document.
<code>nmax</code>	logical, indicating whether to show the maximum number of responses ("n" of a complete, or least-missing, variable) available by group. Default value is TRUE.

caption	character specifying the table caption for descriptives and available data table. If <code>which.table='both'</code> the first element of 'caption' will be assigned to descriptives table and the second to available data table. Default value is NULL, which writes 'Summary descriptives table by groups of 'y'' for descriptives table and 'Available data by groups of 'y'' for the available data table.
label	character specifying the table label for descriptives and available data table. This may be useful to cite the tables elsewhere in the LaTeX document. If <code>which.table='both'</code> the first element of 'label' will be assigned to descriptives table and the second to available data table. Default value is NULL, which assigns no label to the table/s.
loc.caption	character specifying the table caption location. Possible values are 'top' or 'bottom' (partial matching allowed). Default value is 'top'.
...	currently ignored.

Value

List of two possible components corresponding to the code of 'descr' table and 'avail' table. Each component of the list is a character corresponding to the LaTeX code of these tables which can be helpful for post-processing.

Note

The table is created in Latex language using the longtable environment. Therefore, it is necessary to type `\includepackage{longtable}` in the preamble of the LaTeX main document where the table code is inserted.

See Also

[createTable](#), [export2csv](#), [export2html](#)

Examples

```
## Not run:
require(compareGroups)
data(regicor)
res <- compareGroups(sex ~. -id-todeath-death-tocv-cv, regicor)
export2latex(createTable(res, hide.no = 'n'), file="table1")

## End(Not run)
```

regicor *REGICOR cross-sectional data*

Description

These data come from 3 different cross-sectional surveys of individuals representative of the population from a north-west Spanish province (Girona), REGICOR study.

Usage

```
data(regicor)
```

Format

A data frame with 2294 observations on the following 21 variables:

id Individual id
year a factor with levels 1995 2000 2005. Recruitment year
age Patient age at recruitment date
sex a factor with levels male female. Sex
smoker a factor with levels Never smoker Current or former < 1y Never or former >= 1y.
Smoking status
sbp Systolic blood pressure
dbp Diastolic blood pressure
histhtn a factor with levels Yes No. History of hypertension
txhtn a factor with levels No Yes. Hypertension (HTN) treatment
chol Total cholesterol (mg/dl)
hdl HDL cholesterol (mg/dl)
triglyc Triglycerides (mg/dl)
ldl LDL cholesterol (mg/dl)
histchol a factor with levels Yes No. History of hypercholesterolemia
txchol a factor with levels No Yes. Cholesterol treatment
height Height (cm)
weight Weight (Kg)
bmi Body mass index
phyact Physical activity (Kcal/week)
pcs Physical component summary
mcs Mental component summary
death a factor with levels No Yes. Overall death
todeath Days to overall death or end of follow-up
cv a factor with levels No Yes. Cardiovascular event
tocv Days to cardiovascular event or end of follow-up

Details

The variables collected in the REGICOR study were mainly cardiovascular risk factors (hundreds of variables were collected in the different questionnaires and blood measurements), but the variables present in this data set are just a few of them. Also, for reasons of confidentiality, the individuals in this data set are a 30% approx. random subsample of the original one.

Each variable of this data.frame contains label describing them in the attribute "label".

For more information, see the vignette.

Note

Variables death, todeath, cv, tocv are not real but they have been simulated at random to complete the data example with some time-to-event variables.

Source

For reasons of confidentiality, the whole data set is not publicly available. For more information about the study these data come from, visit www.regicor.org.

Examples

```
require(compareGroups)
data(regicor)
summary(regicor)
```

varinfo

Variable names and labels extraction

Description

This functions builds and prints a table with the variable names and their labels.

Usage

```
varinfo(x, ...)
## S3 method for class 'compareGroups'
varinfo(x, ...)
```

Arguments

x an object of class 'compareGroups'
... other arguments currently ignored

Details

By default, a compareGroup descriptives table lists variables by label (if one exists) rather than by name. If researchers have assigned detailed labels to their variables, this function is very useful to quickly locate the original variable name if some modification is required. This function simply lists all "Analyzed variable names" by "Orig varname" (i.e. variable name in the data.frame) and "Shown varname" (i.e., label).

Value

A 'matrix' with two columns

Orig varname actual variable name in the 'data.frame' or in the 'parent environment'.

Shown varname names of the variable shown in the resulting tables.

Note

If a variable has no "label" attribute, then the 'original varname' is the same as the 'shown varname'. The first variable in the table corresponds to the grouping variable. To label non-labeled variables or to change the label, use the function [label](#).

See Also

[compareGroups](#), [createTable](#)

Examples

```
require(compareGroups)
data(regicor)
res<-compareGroups(sex ~ . ,regicor)
createTable(res, hide.no = 'no')
varinfo(res)
```

Index

- *Topic **datasets**
 - regicor, [16](#)
- *Topic **misc**
 - compareGroups, [3](#)
 - createTable, [9](#)
- *Topic **package**
 - compareGroups-package, [2](#)
- *Topic **utilities**
 - cGroupsGUI, [2](#)
 - export2csv, [12](#)
 - export2html, [13](#)
 - export2latex, [14](#)
 - varinfo, [17](#)
- .GlobalEnv, [4](#)

- cGroupsGUI, [2](#)
- compareGroups, [2](#), [3](#), [3](#), [11](#), [18](#)
- compareGroups-package, [2](#)
- createTable, [2](#), [3](#), [8](#), [9](#), [13](#), [15](#), [18](#)

- export2csv, [2](#), [11](#), [12](#), [13](#), [15](#)
- export2html, [2](#), [11](#), [13](#), [13](#), [15](#)
- export2latex, [2](#), [11](#), [13](#), [14](#)

- hist, [6](#)

- label, [7](#), [18](#)

- na.exclude, [5](#)
- na.pass, [5](#)

- oddsratio, [8](#)

- plot.compareGroups (compareGroups), [3](#)
- print.compareGroups (compareGroups), [3](#)
- print.createTable (createTable), [9](#)
- print.default, [10](#)
- print.summary.compareGroups (compareGroups), [3](#)
- print.summary.createTable (createTable), [9](#)

- regicor, [3](#), [16](#)

- shapiro.test, [4](#)
- summary.compareGroups (compareGroups), [3](#)
- summary.createTable (createTable), [9](#)

- update.compareGroups (compareGroups), [3](#)

- varinfo, [17](#)

- write.csv, [12](#)
- write.csv2, [12](#)
- write.table, [12](#)