

# Package ‘CRUF’

March 5, 2020

**Title** Clinical Research Utilities Functions

**Version** 0.5.1

**URL** <https://github.com/Ygall/CRUF>

**BugReports** <https://github.com/Ygall/CRUF/issues>

**Description** Miscellaneous functions for clinical research data analysis. Format table of descriptive statistics, regression models, pvalues according to medical journals standards.

**Depends** R (>= 3.4)

**License** GPL-3

**Encoding** UTF-8

**LazyLoad** true

**LazyData** true

**Imports** utils, stats, survminer, survival, miceadds, aod

**Suggests** knitr, rmarkdown, testthat

**VignetteBuilder** knitr

**RoxygenNote** 7.0.2

**NeedsCompilation** no

**Author** Yves Gallien [aut, cre] (<<https://orcid.org/0000-0002-1245-9260>>)

**Maintainer** Yves Gallien <yves.gallien@gmail.com>

**Repository** CRAN

**Date/Publication** 2020-03-05 20:50:02 UTC

## R topics documented:

as_numeric_factor . . . . .	2
B7 . . . . .	2
boys . . . . .	3
logistic_cluster_multivariate . . . . .	4
logistic_cluster_univariate . . . . .	4
logistic_multivariate . . . . .	5

logistic_univariate . . . . .	6
pval_format . . . . .	7
select_noms . . . . .	7
step_lrcl_pval . . . . .	8
survival_univariate . . . . .	8
tabkris_2 . . . . .	9

<b>Index</b>	<b>12</b>
--------------	-----------

---

as_numeric_factor	<i>Numeric factor</i>
-------------------	-----------------------

---

### Description

Coerce a factor to a numeric vector

### Usage

```
as_numeric_factor(x)
```

### Arguments

x	Factor to coerce
---	------------------

### Value

A vector as a numeric

### Examples

```
vec <- as.factor(c(8:10))

as.numeric(vec) # Return a false value
as_numeric_factor # Return the value of numeric vector
```

---

B7	<i>Growth of Dutch boys</i>
----	-----------------------------

---

### Description

Height, weight, head circumference and puberty of 748 Dutch boys.

### Format

A data frame with 53 rows of 46 variables of a Behcet disease

**Details**

Random sample of 10% from the cross-sectional data used to construct the Dutch growth references 1997. Variables gen and phb are ordered factors. reg is a factor.

---

boys

*Growth of Dutch boys*

---

**Description**

Height, weight, head circumference and puberty of 748 Dutch boys.

**Format**

A data frame with 748 rows on the following 9 variables:

**age** Decimal age (0-21 years)

**hgt** Height (cm)

**wgt** Weight (kg)

**bmi** Body mass index

**hc** Head circumference (cm)

**gen** Genital Tanner stage (G1-G5)

**phb** Pubic hair (Tanner P1-P6)

**tv** Testicular volume (ml)

**reg** Region (north, east, west, south, city)

**Details**

Random sample of 10% from the cross-sectional data used to construct the Dutch growth references 1997. Variables gen and phb are ordered factors. reg is a factor.

**Source**

Fredriks, A.M., van Buuren, S., Burgmeijer, R.J., Meulmeester JF, Beuker, R.J., Brugman, E., Roede, M.J., Verloove-Vanhorick, S.P., Wit, J.M. (2000) Continuing positive secular growth change in The Netherlands 1955-1997. *Pediatric Research*, **47**, 316-323.

Fredriks, A.M., van Buuren, S., Wit, J.M., Verloove-Vanhorick, S.P. (2000). Body index measurements in 1996-7 compared with 1980. *Archives of Disease in Childhood*, **82**, 107-112.

logistic\_cluster\_multivariate

*Multivariate Logistic Regression with cluster*

---

### Description

A function used to generate result table for multivariate logistic regression model using a cluster variable. Compute robust variance using sandwich

### Usage

```
logistic_cluster_multivariate(fit)
```

### Arguments

fit                    Class glm.cluster. Multivariate model to format

### Value

A dataframe of the multivariate parameters formatted

---

logistic\_cluster\_univariate

*Univariate Logistic Regression with cluster*

---

### Description

A function used to generate multiple result table for univariate logistic regression model with  $y \sim x$  using a cluster variable. For each specified `y_names`, a result table is computed, including all `x_names` variables. Compute robust variance using sandwich

### Usage

```
logistic_cluster_univariate(  
  data,  
  y_names,  
  x_names,  
  cluster,  
  twobytwo = TRUE,  
  formula = "(y ~ x)",  
  collapse = FALSE,  
  ref_label = "1",  
  digits = 2  
)
```

**Arguments**

data	A dataframe including all the variables needed in all the models
y_names	Vector. Name(s) of response variable(s)
x_names	Vector. Name(s) of predictor variable(s)
cluster	Character. Name of the clustering variable.
twobytwo	Logical. Either to include the two by two table for each variable. Default is TRUE.
formula	Formula for logistic regression to customize. Default is (y ~ x).
collapse	"NULL", "OR", "CI". Collapse columns in one column. "OR" collapses OR, Upper and Lower CI. "CI" collapses Upper and Lower CI.
ref_label	Character. Set the label for reference estimate.
digits	Numeric. Number of digits to display.

**Value**

The returned value is a list of length y\_names, which consists of a dataframe having the univariate logistic regressions of the x\_names.

---

logistic\_multivariate *Multivariate Logistic Regression*

---

**Description**

A function used to generate result table for multivariate logistic regression model.

**Usage**

```
logistic_multivariate(fit)
```

**Arguments**

fit	Class glm. Multivariate model to format
-----	---

**Value**

A dataframe of the multivariate parameters formatted

---

logistic\_univariate     *Univariate Logistic Regression*

---

### Description

A function used to generate multiple result table for univariate logistic regression model with  $y \sim x$ . For each specified `y_names`, a result table is computed, including all `x_names` variables.

### Usage

```
logistic_univariate(  
  data,  
  y_names,  
  x_names,  
  twobytwo = TRUE,  
  formula = "(y ~ x)",  
  collapse = FALSE,  
  ref_label = "1",  
  digits = 2  
)
```

### Arguments

<code>data</code>	A dataframe including all the variables needed in all the models
<code>y_names</code>	Vector. Name(s) of response variable(s)
<code>x_names</code>	Vector. Name(s) of predictor variable(s)
<code>twobytwo</code>	Logical. Either to include the two by two table for each variable. Default is TRUE.
<code>formula</code>	Formula for logistic regression to customize. Default is $(y \sim x)$ .
<code>collapse</code>	"NULL", "OR", "CI". Collapse columns in one column. "OR" collapses OR, Upper and Lower CI. "CI" collapses Upper and Lower CI.
<code>ref_label</code>	Character. Set the label for reference estimate.
<code>digits</code>	Numeric. Number of digits to display.

### Value

The returned value is a list of length `y_names`, which consists of a dataframe having the univariate logistic regressions of the `x_names`.

---

pval_format	<i>p-value format</i>
-------------	-----------------------

---

**Description**

Format a p-value into R display system with stars

**Usage**

```
pval_format(pval)
```

**Arguments**

pval            Numeric.

**Value**

"\*\*\*\*" if < 0.001, "\\\*\\\*" if < 0.01, "\\\*" if < 0.05, "." if < 0.1

**Examples**

```
pval_format(0.00025)
```

```
pval_format(0.20)
```

---

select_noms	<i>Name selection</i>
-------------	-----------------------

---

**Description**

Name selection

**Usage**

```
select_noms(fichier)
```

**Arguments**

fichier            Le fichier a passer

**Value**

Permet de sélectionner les noms dans un vecteur

---

step_lrcl_pval	<i>Backward stepwise selection with pvalue for logistic regression with clustering</i>
----------------	--

---

**Description**

Backward stepwise selection with pvalue for logistic regression with clustering

**Usage**

```
step_lrcl_pval(fitcl, cluster, threshold = 0.05, verbose = TRUE)
```

**Arguments**

fitcl	Initial multivariate model
cluster	Character. Name of the clustering variable of the model
threshold	Numeric [0,1].
verbose	Whether to display messages or not. Default TRUE

**Value**

A final multivariate model

---

survival_univariate	<i>Univariate Survival Regression</i>
---------------------	---------------------------------------

---

**Description**

Univariate Survival Regression

**Usage**

```
survival_univariate(
  data,
  time,
  time2 = NULL,
  event,
  names = NULL,
  strata = NULL,
  cluster = NULL,
  test = "LRT"
)
```



**Arguments**

data	A dataframe including all the variable needed, one variable for time to event and one variable for event indicator.
time	Name of the variable used for time to event or for start time if Start-Stop format
time2	Stop time if the data are in Start-Stop format
event	Name of the column used for event indicator.
names	Names of the variables to display. Lenght must be minus 2 the number of column of data, excluding time and event
strata	Name of the variable used for analysis with strata
cluster	Name of the variable used for analysis with cluster
test	Which test to use for p-value, possible values are "LRT" for Likelihood Ratio Test, "Wald" for Wald Test and "LogRank" for Log-Rank Test"

**Value**

Return a table with model parameters for every variable included in data.

---

tabkris_2	<i>Data description function</i>
-----------	----------------------------------

---

**Description**

tabkris\_2 computes descriptive statistics for data

**Usage**

```
tabkris_2(
  data,
  names = NULL,
  varint = NULL,
  lang = "en",
  method = NULL,
  test = FALSE,
  pres_quant = c("med"),
  pres_quali = c("n", "per"),
  default_method = c("cont", "bino", "cate", "ordo"),
  default_test = c("stud", "chisq", "chisq", "chisq"),
  explicit_na = FALSE,
  digits = 2,
  return_table = TRUE,
  auto_detect = TRUE,
  lev_co = 10,
  verbose = FALSE
)
```

**Arguments**

<code>data</code>	Dataframe to describe or a "descatable" object
<code>names</code>	Vectors of variables to display in the final table, length of <code>ncol(data)</code>
<code>varint</code>	Variable to stratify on, factor only
<code>lang</code>	Language to display, default "en", "fr"
<code>method</code>	Vectors of variables to customize the methods used for description, length of data columns
<code>test</code>	Either a logical indicating statistical tests execution or a vectors of variables to customize the tests, length of data columns. Default FALSE
<code>pres_quant</code>	Descriptive statistics for quantitative variables. Possible values are "mean" for mean, SD, "med" for median, IQR, "range" for range
<code>pres_quali</code>	Descriptive statistics for qualitative variables. Possible values are "n" for number, "total" to add "/ total" and "per" for percentages
<code>default_method</code>	Default method to compute the table for each variable. Default <code>default_method = c("cont", "bino", "cate", "ordo")</code>
<code>default_test</code>	Default test to apply for each variable type. Default <code>c("stud", "chisq", "chisq", "chisq")</code> . Available "stud", "wilcox", "kruskal", "chisq", "fish"
<code>explicit_na</code>	Whether to display NA in description, Default FALSE
<code>digits</code>	Number of significant number to display, default 2
<code>return_table</code>	Whether to return a dataframe or an object to customize option easily, default TRUE
<code>auto_detect</code>	Whether to automatically detect variable type, transforming to factors numeric variable with moderate levels (< 10), default TRUE. Possible to set the cut-off number with <code>lev_co</code>
<code>lev_co</code>	Numeric. When <code>auto_detect</code> is TRUE, set the number of level to cutoff for categorical variables
<code>verbose</code>	Logical. Display information about transformation of variables. default FALSE

**Details**

The `tabkris_2` function is a function to describe a set of data. Main purpose is to create a typical table one in biomedical literature, either a patient characteristic table or population characteristic table.

`names` is a vector to name the variable of data. Default will use the `colnames` of data.

`varint` is a variable to stratify the analysis. It must be included in the initial dataset. It will not be displayed in the final table if chosen as the stratifying variable

`lang` is useful to choose the language for the final display. The default is english. French is also supported.

`default_method` and `method` are used to set the methods used for display. `default_method` must be length 4, to set the default method for continuous, binomial, categorical and ordered variable. `method` must be length of data columns, used to fine-tune every method for each variable.

default\_test and test are used to set the tests performed. default\_test must be length 4, to set the default method for continuous, binomial, categorical and ordered variable. test must be length of data columns, used to fine-tune every test for each variable.

pres\_quant is used to set the display of quantitative variable. mean (SD), median [IQR] and range are available, default is median.

pres\_quali is used to set the display of qualitative variable. "n" for number, "total" to add "/" total" and "per" for percentages, default is "n / per".

explicit\_na is used to display.

digits is the number of digits to display for numbers. Usually if  $n < 100$ ,  $digits = 0$  if  $100 < n < 200$ ,  $digits = 1$  else  $digits = 2$ .

return\_table choose if the user wants to directly display a table or if the user wants to get an object with parametrable objects.

auto\_detect will test if each column can be coerced to a factor (i.e. having between 2 and 10 levels) and change the type of variable if so.

lev\_co will set the number of maximum levels to coerce a column in a factor

### Value

Depending on argument return\_table, an object of class data.frame, which is the descriptive table or an object of class "desctable", which is a customizable object.

### Author(s)

Yves Gallien <yves.gallien@gmail.com>, 2019

### See Also

<https://github.com/Ygall/CRUF> for manual and examples.

### Examples

```
tabkris_2(boys)
```

# Index

## \*Topic **datasets**

B7, [2](#)

boys, [3](#)

as\_numeric\_factor, [2](#)

B7, [2](#)

boys, [3](#)

logistic\_cluster\_multivariate, [4](#)

logistic\_cluster\_univariate, [4](#)

logistic\_multivariate, [5](#)

logistic\_univariate, [6](#)

pval\_format, [7](#)

select\_noms, [7](#)

step\_lrcl\_pval, [8](#)

survival\_univariate, [8](#)

tabkris\_2, [9](#)