# Package: r2spss (via r-universe)

September 4, 2024

**Type** Package

**Title** Format R Output to Look Like SPSS

**Version** 0.3.3

**Date** 2022-06-03

**Description** Create plots and LaTeX tables that look like SPSS output
for use in teaching materials. Rather than copying-and-pasting
SPSS output into documents, R code that mocks up SPSS output
can be integrated directly into dynamic LaTeX documents with
tools such as knitr. Functionality includes statistical
techniques that are typically covered in introductory
statistics classes: descriptive statistics, common hypothesis
tests, ANOVA, and linear regression, as well as box plots,
histograms, scatter plots, and line plots (including profile
plots).

**License** GPL (>= 3)

**URL** <https://github.com/aalfons/r2spss>

**BugReports** <https://github.com/aalfons/r2spss/issues>

**Depends** R (>= 3.5.0), ggplot2 (>= 3.3.0)

**Imports** graphics, scales, stats, car

**Suggests** knitr

**LazyLoad** yes

**VignetteBuilder** knitr

**Author** Andreas Alfons [aut, cre]
(<<https://orcid.org/0000-0002-2513-3788>>)

**Maintainer** Andreas Alfons <alfons@ese.eur.nl>

**Encoding** UTF-8

**RoxygenNote** 7.2.0

**Repository** https://aalfons.r-universe.dev

**RemoteUrl** https://github.com/aalfons/r2spss

**RemoteRef** HEAD

**RemoteSha** 84c1856d406f6f49b8cff8fb72bdeb4ddf4413b4

1

# Contents

---

r2spss-package                 *Format R Output to Look Like SPSS*

---

### Description

Create plots and LaTeX tables that look like SPSS output for use in teaching materials. Rather than copying-and-pasting SPSS output into documents, R code that mocks up SPSS output can be integrated directly into dynamic LaTeX documents with tools such as knitr. Functionality includes statistical techniques that are typically covered in introductory statistics classes: descriptive statistics, common hypothesis tests, ANOVA, and linear regression, as well as box plots, histograms, scatter plots, and line plots (including profile plots).

### Details

The DESCRIPTION file: This package was not yet installed at build time.

Index: This package was not yet installed at build time.

## Author(s)

Andreas Alfons [aut, cre] (<https://orcid.org/0000-0002-2513-3788>)

Maintainer: Andreas Alfons <alfons@ese.eur.nl>

---

| ANOVA | *One-way and Two-way ANOVA* |
|---|---|

---

## Description

Perform one-way or two-way ANOVA on variables of a data set. The output is printed as a LaTeX table that mimics the look of SPSS output, and a profile plot of the results mimics the look of SPSS graphs.

## Usage

```
ANOVA(data, variable, group, conf.level = 0.95)

## S3 method for class 'ANOVA_SPSS'
to_SPSS(
  object,
  statistics = c("test", "variance", "descriptives"),
  version = r2spss_options$get("version"),
  digits = 3,
  ...
)

## S3 method for class 'ANOVA_SPSS'
print(
  x,
  statistics = c("descriptives", "variance", "test"),
  version = r2spss_options$get("version"),
  ...
)

## S3 method for class 'ANOVA_SPSS'
plot(x, y, which = 1, version = r2spss_options$get("version"), ...)
```

## Arguments

| | |
|---|---|
| data | a data frame containing the variables. |
| variable | a character string specifying the numeric variable of interest. |
| group | a character vector specifying one or two grouping variables. |
| conf.level | a number between 0 and 1 giving the confidence level of the confidence interval. |
| object, x | an object of class "ANOVA_SPSS" as returned by function ANOVA. |

statistics    a character string or vector specifying which SPSS tables to produce. Available options are "descriptives" for descriptive statistics, "variance" for Levene's test on homogeneity of the variances, and "test" for ANOVA results. For the to_SPSS method, only one option is allowed (the default is the table of ANOVA results), but the print method allows several options (the default is to print all tables).

version       a character string specifying whether the table or plot should mimic the content and look of recent SPSS versions ("modern") or older versions (<24; "legacy"). For the table, the main differences in terms of content are that recent versions include different variations of Levene's test, and that small p-values are displayed differently.

digits        an integer giving the number of digits after the comma to be printed in the SPSS tables.

...           for the to_SPSS and print methods, additional arguments to be passed down to [format_SPSS](). For the plot method, additional arguments to be passed down to [linesSPSS](), in particular graphical parameters.

y             ignored (only included because it is defined for the generic function [plot]()).

which         for two-way ANOVA, an integer with possible values 1 or 2 indicating whether the first or the second factor should be used on the $x$-axis. The other factor will then be used for drawing separate lines. For one-way ANOVA, this is not meaningful and ignored.

## Details

The print method first calls the to_SPSS method followed by [to_latex](). Further customization can be done by calling those two functions separately, and modifying the object returned by to_SPSS.

## Value

ANOVA returns an object of class "ANOVA_SPSS" with the following components:

descriptives a data frame containing per-group descriptive statistics.

levene an object as returned by [leveneTest]() (if version = "legacy"); or a list of such objects containing different variations of Levene's test (if version = "modern").

test a data frame containing the ANOVA table.

variable a character string containing the name of the numeric variable of interest.

group a character vector containing the name(s) of the grouping variable(s).

i an integer giving the number of groups in the (first) grouping variable.

j an integer giving the number of groups in the second grouping variable (only two-way ANOVA).

conf.level numeric; the confidence level used.

type a character string giving the type of ANOVA performed ("one-way" or "two-way").

The to_SPSS method returns an object of class "SPSS_table" which contains all relevant information in the required format to produce the LaTeX table. See [to_latex]() for possible components and how to further customize the LaTeX table based on the returned object.

The `print` method produces a LaTeX table that mimics the look of SPSS output.

The `plot` method returns an object of class `"ggplot"`, which produces a profile plot of the ANOVA results when printed.

**Note**

The test statistic and p-value for Levene's test based on the trimmed mean (only returned for `version = "modern"`) differ slightly from those returned by SPSS. Function `trimmed_mean` rounds the number of observations to be trimmed in a different manner than the base R function `mean`, which brings the results closer to those of SPSS, but they are still not identical.

LaTeX tables that mimic recent versions of SPSS (`version = "modern"`) may require several LaTeX compilations to be displayed correctly.

**Author(s)**

Andreas Alfons

**Examples**

```
# load data
data("Eredivisie")
# log-transform market values
Eredivisie$logMarketValue <- log(Eredivisie$MarketValue)

# one-way ANOVA
oneway <- ANOVA(Eredivisie, "logMarketValue",
                group = "Position")
oneway        # print LaTeX table
plot(oneway)  # create profile plot

# two-way ANOVA
twoway <- ANOVA(Eredivisie, "logMarketValue",
                group = c("Position", "Foreign"))
twoway        # print LaTeX table
plot(twoway)  # create profile plot
```

---

box_plot                          *Box Plots*

---

**Description**

Draw box plots of variables in a data frame, including box plots for groups of observations and box plots for separate variables. The plots thereby mimic the look of SPSS graphs.

**Usage**

```
box_plot(
  data,
  variables,
  group = NULL,
  cut.names = NULL,
  style = c("T", "whiskers"),
  coef = c(1.5, 3),
  outlier.shape = c(1, 42),
  version = r2spss_options$get("version"),
  ...
)
```

**Arguments**

| | |
|---|---|
| data | a data frame containing the variables to be plotted. |
| variables | a character vector specifying separate variables to be plotted. If group is not NULL, only the first variable is used and box plots of groups of observations are drawn instead. |
| group | an character string specifying a grouping variable, or NULL for no grouping. |
| cut.names | a logical indicating whether to cut long variable names or group labels to 8 characters. The default is TRUE for box plots of separate variables, but FALSE for box plots of groups of observations (which mimics SPSS behavior). |
| style | a character string specifying the box plot style. Possible values are "T" for T-bars (the default) or "whiskers" for simple whiskers. |
| coef | a numeric vector of length 2 giving the multipliers of the interquartile range for determining intermediate and extreme outliers, respectively. |
| outlier.shape | an integer vector of length 2 giving the plot symbol for intermediate and extreme outliers, respectively. |
| version | a character string specifying whether the plot should mimic the look of recent SPSS versions ("modern") or older versions (<24; "legacy"). |
| ... | additional arguments to be passed down, in particular aesthetics (see [geom_boxplot](#)). |

**Value**

An object of class ["ggplot"](#), which produces a box plot when printed.

**Author(s)**

Andreas Alfons

**Examples**

```
## paired sample
# load data
data("Exams")
```

```
# plot grades on regular and resit exams
box_plot(Exams, c("Regular", "Resit"))


## independent samples

# load data
data("Eredivisie")
# log-transform market values
Eredivisie$logMarketValue <- log(Eredivisie$MarketValue)

# plot log market values of Dutch and Foreign players
box_plot(Eredivisie, "logMarketValue", group = "Foreign")
```

---

chisq_test                       $\chi\hat{}2$ *Tests*

---

#### Description

Perform a $\chi^2$ goodness-of-fit test or a $\chi^2$ test on independence on variables of a data set. The output is printed as a LaTeX table that mimics the look of SPSS output.

#### Usage

```
chisq_test(data, variables, p = NULL)

## S3 method for class 'chisq_test_SPSS'
to_SPSS(
  object,
  statistics = c("test", "frequencies"),
  version = r2spss_options$get("version"),
  digits = c(1, 3),
  ...
)

## S3 method for class 'chisq_test_SPSS'
print(
  x,
  statistics = c("frequencies", "test"),
  version = r2spss_options$get("version"),
  digits = c(1, 3),
  ...
)

chisqTest(data, variables, p = NULL)
```

## Arguments

| | |
|---|---|
| data | a data frame containing the variables. |
| variables | a character vector specifying the categorical variable(s) of interest. If only one variable is specified, a goodness-of-fit test is performed. If two variables are specified, a test on independence is performed (with the first variable used for the rows and the second variable for the columns of the crosstabulation). |
| p | a vector of probabilities for the categories in the goodness-of-fit test. |
| object, x | an object of class `"chisq_test_SPSS"` as returned by function `chisq_test`. |
| statistics | a character string or vector specifying which SPSS tables to produce. Available options are `"frequencies"` for a table of the observed and expected frequencies, and `"test"` for test results. For the `to_SPSS` method, only one option is allowed (the default is the table of test results), but the `print` method allows several options (the default is to print all tables). |
| version | a character string specifying whether the table should mimic the content and look of recent SPSS versions (`"modern"`) or older versions (<24; `"legacy"`). The main difference in terms of content is that small p-values are displayed differently. |
| digits | an integer vector giving the number of digits after the comma to be printed in the SPSS tables. The first element corresponds to the number of digits for the expected frequencies, and the second element corresponds to the number of digits in the table for the test. |
| ... | additional arguments to be passed down to [format_SPSS](). |

## Details

The `print` method first calls the `to_SPSS` method followed by [to_latex](). Further customization can be done by calling those two functions separately, and modifying the object returned by `to_SPSS`.

## Value

An object of class `"chisq_test_SPSS"` with the following components:

chisq   a list containing the results of the $\chi^2$ test.

lr   a list containing the results of a likelihood ratio test (only test on independence).

MH   a list containing the results of a Mantel-Haenszel test of linear association (only test on independence).

observed   a table containing the observed frequencies.

expected   a vector or matrix containing the expected frequencies.

n   an integer giving the number of observations.

k   an integer giving the number of groups (only goodness-of-fit test).

r   an integer giving the number of groups in the first variable corresponding to the rows (only test on independence).

c   an integer giving the number of groups in the second variable corresponding to the columns (only test on independence).

variables  a character vector containing the name(s) of the categorical variable(s) of interest.

type  a character string giving the type of $\chi^2$ test performed ("goodness-of-fit" or "independence").

The to_SPSS method returns an object of class "SPSS_table" which contains all relevant information in the required format to produce the LaTeX table. See [to_latex](#) for possible components and how to further customize the LaTeX table based on the returned object.

The print method produces a LaTeX table that mimics the look of SPSS output.

## Note

The test on independence also reports the results of a likelihood ratio test.

LaTeX tables that mimic recent versions of SPSS (version = "modern") may require several La-TeX compilations to be displayed correctly.

## Author(s)

Andreas Alfons

## Examples

```
# load data
data("Eredivisie")

# test whether the traditional Dutch 4-3-3 (total football)
# is still reflected in player composition
chisq_test(Eredivisie, "Position", p = c(1, 4, 3, 3)/11)

# test whether playing position and dummy variable for
# foreign players are independent
chisq_test(Eredivisie, c("Position", "Foreign"))
```

---

descriptives *Descriptive Statistics*

---

## Description

Compute descriptive statistics of numeric variables of a data set (number of observations, minimum, maximum, mean, standard deviaiton). The output is printed as a LaTeX table that mimics the look of SPSS output.

## Usage

```
descriptives(data, variables)

## S3 method for class 'descriptives_SPSS'
to_SPSS(object, digits = 2, ...)

## S3 method for class 'descriptives_SPSS'
print(x, version = r2spss_options$get("version"), ...)
```

## Arguments

| | |
|---|---|
| `data` | a data frame containing the variables. |
| `variables` | a character vector specifying numeric variables for which to compute descriptive statistics. |
| `object, x` | an object of class `"descriptives_SPSS"` as returned by function `descriptives`. |
| `digits` | an integer giving the number of digits after the comma to be printed in the SPSS table. |
| `...` | additional arguments to be passed down to [format_SPSS](). |
| `version` | a character string specifying whether the table should mimic the look of recent SPSS versions (`"modern"`) or older versions (<24; `"legacy"`). |

## Details

The `print` method first calls the `to_SPSS` method followed by [to_latex](). Further customization can be done by calling those two functions separately, and modifying the object returned by `to_SPSS`.

## Value

An object of class `"descriptives_SPSS"` with the following components:

`classes` a character vector giving the (first) class of the variables of interest.

`descriptives` a data frame containing the descriptive statistics.

`n` an integer giving the number of observations.

The `to_SPSS` method returns an object of class `"SPSS_table"` which contains all relevant information in the required format to produce the LaTeX table. See [to_latex]() for possible components and how to further customize the LaTeX table based on the returned object.

The `print` method produces a LaTeX table that mimics the look of SPSS output.

## Note

LaTeX tables that mimic recent versions of SPSS (`version = "modern"`) may require several LaTeX compilations to be displayed correctly.

## Author(s)

Andreas Alfons

## Examples

```
# load data
data("Eredivisie")

# compute descriptive statistics for market value and age
descriptives(Eredivisie, c("MarketValue", "Age"))
```

| Eredivisie | *Football players of the Dutch Eredivisie season 2013-14* |

### Description

Data on all football players in the Dutch Eredivisie, the highest men's football league in the Netherlands, who played at least one match in the 2013-14 season.

### Usage

```
data("Eredivisie")
```

### Format

A data frame with 417 observations on the following 20 variables.

Player  the player's name.

Team  the team with which the player was under contract at the end of the 2013-14 season.

MarketValue  the player's market value after the 2013-14 season.

Age  the player's age in years.

Height  the player's height in centimeters.

Foreign  a dummy variable with value 0 for Dutch players and value 1 for players without a Dutch nationality.

Position  the primary postion of the player ("Goalkeeper", "Defender", "Midfielder", or "Forward").

BothFeet  a dummy variable with value 0 if the player has one stronger foot and value 1 if the player is equally strong with both feet.

AtClub  the number of years the player is with the current team.

Contract  the number of years remaining on the player's current contract.

Matches  the number of matches played in the 2013-14 season.

Goals  the number of goals scored in the 2013-14 season.

OwnGoals  the number of own goals scored in the 2013-14 season.

Assists  the number of assists given in the 2013-14 season.

Yellow  the number of yellow cards received in the 2013-14 season.

YellowRed  the number of yellow-red cards received in the 2013-14 season.

Red  the number of red cards received in the 2013-14 season.

SubOn  the number of times the player was substituted on the field in the 2013-14 season.

SubOff  the number of times the player was substituted off the field in the 2013-14 season.

Minutes  the number of minutes played in the 2013-14 season.

### Source

<https://www.transfermarkt.de/>

**Examples**

```
data("Eredivisie")
summary(Eredivisie)
```

---

Exams                          *Exam results of an applied statistics course*

---

**Description**

Data on grades for an applied statistics course at Erasmus University Rotterdam for students who took both the regular exam and the resit. Grades in the Netherlands are on a scale from 1 to 10, with a higher grade being better, and a minimum of 5.5 is required to pass.

**Usage**

```
data("Exams")
```

**Format**

A data frame with 45 observations on the following 2 variables.

Regular  the student's grade based on the regular exam at the end of the course.

Resit  the student's grade based on the resit exam at the end of the academic year.

**Examples**

```
data("Exams")
summary(Exams)
```

---

format_SPSS                    *Format Objects*

---

**Description**

Format an object for printing, mostly used to print numeric data in the same way as SPSS. This is mainly for internal use in `to_SPSS` and `print` methods.

## Usage

```
format_SPSS(object, ...)

## Default S3 method:
format_SPSS(object, ...)

## S3 method for class 'integer'
format_SPSS(object, ...)

## S3 method for class 'numeric'
format_SPSS(object, digits = 3, p_value = FALSE, check_int = FALSE, ...)

## S3 method for class 'matrix'
format_SPSS(object, digits = 3, p_value = FALSE, check_int = FALSE, ...)

## S3 method for class 'data.frame'
format_SPSS(object, digits = 3, p_value = FALSE, check_int = FALSE, ...)

formatSPSS(object, ...)
```

## Arguments

| | |
|---|---|
| `object` | an R object. Currently methods are implemented for vectors, matrices, and data frames. The default method calls `as.character`. |
| `...` | additional arguments passed down to methods. |
| `digits` | an integer giving the number of digits after the comma to display. |
| `p_value` | a logical indicating whether small positive values should be indicated as below the threshold defined by `digits`, e.g., `"<.001"` if `digits = 3`. This is used for formatting p-values in LaTeX tables that mimic the look of SPSS. For the `"numeric"` method, a logical vector indicates the behavior for each element of `object`. For the `"matrix"` or `"data.frame"` methods, a logical vector indicates the behavior for each column of `object`. |
| `check_int` | a logical indicating whether to check for integer values and format them as such, e.g., to format the integer 2 as `"2"` instead of `"2.000"` if `digits = 3`. For the `"numeric"` method, a logical vector indicates the behavior for each element of `object`. For the `"matrix"` or `"data.frame"` methods, a logical vector indicates the behavior for each column of `object`. |

## Value

A character vector, matrix, or data frame containing the formatted object.

## Author(s)

Andreas Alfons

**Examples**

```
# note how numbers in the interval (-1, 1) are printed
# without the zero in front of the comma
format_SPSS(c(-1.5, -2/3, 2/3, 1.5))
```

---

histogram                                  *Histogram*

---

**Description**

Draw a histogram of a variable in a data frame. The plot thereby mimics the look of SPSS graphs.

**Usage**

```
histogram(
  data,
  variable,
  bins = NULL,
  normal = FALSE,
  normal.colour = NULL,
  normal.color = NULL,
  normal.linetype = NULL,
  normal.size = NULL,
  normal.alpha = NULL,
  digits = 3,
  limits = NULL,
  expand = 0.05,
  version = r2spss_options$get("version"),
  ...
)
```

**Arguments**

| | |
|---|---|
| data | a data frame containing the variable to be plotted. |
| variable | a character string specifying the variable to be plotted. |
| bins | an integer giving the number of bins for the histogram. |
| normal | a logical indicating whether to add a normal density with the estimated mean and standard deviation (the default is FALSE). |
| normal.colour, normal.alpha | normal.color, normal.linetype, normal.size, |
| | aesthetics for the normal density. In the unlikely event that both US and UK spellings of color are supplied, the US spelling will take precedence. |
| digits | an integer giving the number of digits after the comma to be printed in the summary statistics in the right plot margin. |

| limits | a list of arguments to be passed to expand_limits. Typically, the list would contain components x or y to specify values that should be included in the range of the corresponding axis. |
|---|---|
| expand | a numeric value specifying the percentage of the range to be used for padding the axes. The default is 0.05 to expand the $x$-axis by 5% on both sides and the $y$-axis by 5% on the upper end. Note that there is no padding on lower end of the $y$-axis to mimic SPSS behavior. |
| version | a character string specifying whether the plot should mimic the look of recent SPSS versions ("modern") or older versions (<24; "legacy"). |
| ... | additional arguments to be passed down, in particular aesthetics (see geom_histogram and geom_line). |

## Value

An object of class "ggplot", which produces a histogram when printed.

## Note

Due to the inner workings of this function to mimic the look of histograms in SPSS, it is not expected that the user adds scale_x_continuous or scale_y_continuous to the plot. Instead, axis limits and padding should be modified via the limits and expand arguments.

## Author(s)

Andreas Alfons

## Examples

```
# load data
data("Eredivisie")
# log-transform market values
Eredivisie$logMarketValue <- log(Eredivisie$MarketValue)

# plot histogram of log market values
histogram(Eredivisie, "logMarketValue", normal = TRUE,
          limits = list(x = c(9.5, 17.5)))
```

---

| kruskal_test | *Kruskal-Wallis Test* |
|---|---|

---

## Description

Perform a Kruskal-Wallis test on variables of a data set. The output is printed as a LaTeX table that mimics the look of SPSS output.

**Usage**

```
kruskal_test(data, variable, group)

## S3 method for class 'kruskal_test_SPSS'
to_SPSS(
  object,
  statistics = c("test", "ranks"),
  version = r2spss_options$get("version"),
  digits = NULL,
  ...
)

## S3 method for class 'kruskal_test_SPSS'
print(
  x,
  statistics = c("ranks", "test"),
  version = r2spss_options$get("version"),
  digits = 2:3,
  ...
)

kruskalTest(data, variable, group)
```

**Arguments**

| | |
|---|---|
| data | a data frame containing the variables. |
| variable | a character string specifying the numeric variable of interest. |
| group | a character string specifying a grouping variable. |
| object, x | an object of class "kruskal_test_SPSS" as returned by function kruskal_test. |
| statistics | a character string or vector specifying which SPSS tables to produce. Available options are "ranks" for a summary of the ranks and "test" for test results. For the to_SPSS method, only one option is allowed (the default is the table of test results), but the print method allows several options (the default is to print all tables). |
| version | a character string specifying whether the table should mimic the content and look of recent SPSS versions ("modern") or older versions (<24; "legacy"). The main differences in terms of content are the label of the test statistic and that small p-values are displayed differently. |
| digits | for the to_SPSS method, an integer giving the number of digits after the comma to be printed in the SPSS table. For the print method, this should be an integer vector of length 2, with the first element corresponding to the number of digits in table with the summary of the ranks, and the second element corresponding to the number of digits in the table for the test. |
| ... | additional arguments to be passed down to format_SPSS. |

## Details

The `print` method first calls the `to_SPSS` method followed by [`to_latex`](). Further customization can be done by calling those two functions separately, and modifying the object returned by `to_SPSS`.

## Value

An object of class `"kruskal_test_SPSS"` with the following components:

`statistics` a data frame containing information on the per-group mean ranks.

`test` a list containing the results of the Kruskal-Wallis test.

`variable` a character string containing the name of the numeric variable of interest.

`group` a character string containing the name of the grouping variable.

The `to_SPSS` method returns an object of class `"SPSS_table"` which contains all relevant information in the required format to produce the LaTeX table. See [`to_latex`]() for possible components and how to further customize the LaTeX table based on the returned object.

The `print` method produces a LaTeX table that mimics the look of SPSS output.

## Note

LaTeX tables that mimic recent versions of SPSS (`version = "modern"`) may require several LaTeX compilations to be displayed correctly.

## Author(s)

Andreas Alfons

## Examples

```
# load data
data("Eredivisie")

# test whether market values differ by playing position
kruskal_test(Eredivisie, "MarketValue", group = "Position")
```

---

labels_SPSS                *Format axis tick labels similar to SPSS*

---

## Description

Format axis tick labels in a similar manner to SPSS to mimic the look of SPSS graphs.

**Usage**

```
number_SPSS(x, big.mark = "", ...)

numberSPSS(x, big.mark = "", ...)

substr_SPSS(x, start = 1, stop = 8)

substrSPSS(x, start = 1, stop = 8)
```

**Arguments**

| | |
|---|---|
| x | for number_SPSS, a numeric vector to format. For substr_SPSS, a vector of character strings to be cut. |
| big.mark | a character string to be inserted every 3 digits to separate thousands. The default is an empty string for no separation. |
| ... | additional arguments to be passed to number. |
| start, stop | integers giving the first and last character to remain in the cut string, respectively. The default is to cut strings to the first 8 characters. |

**Details**

number_SPSS is a wrapper for number that by default does not put a separator every 3 digits so separate thousands. It mainly exists to prevent scientific notation in axis tick labels, hence it is typically supplied as the labels argument of scale_x_continuous or scale_y_continuous.

substr_SPSS is a wrapper for substr to cut character strings by default to the first 8 characters, which is SPSS behavior for the tick labels of a discrete axis in some (but not all) plots. It is typically supplied as the labels argument of scale_x_discrete or scale_y_discrete.

**Value**

A character vector of the same length as x

---

| line_plot | *Line Plots* |
|---|---|

---

**Description**

Draw connected lines for variables in a data frame. The plot thereby mimics the look of SPSS graphs.

## Usage

```
line_plot(
  data,
  variables,
  index = NULL,
  version = r2spss_options$get("version"),
  ...
)
```

## Arguments

| | |
|---|---|
| data | a data frame containing the variables to be plotted. |
| variables | a character vector specifying at least one variable to be plotted on the $y$-axis. In case of multiple variables, separate lines are drawn for each variable and a legend is shown. |
| index | a character string specifying a variable to be plotted on the $x$-axis, or NULL to plot the observations against their index. |
| version | a character string specifying whether the plot should mimic the look of recent SPSS versions ("modern") or older versions (<24; "legacy"). |
| ... | additional arguments to be passed down to geom_line. |

## Value

An object of class "ggplot", which produces a line plot when printed.

## Author(s)

Andreas Alfons

## Examples

```
# load data
data("Eredivisie")
# log-transform market values
Eredivisie$logMarketValue <- log(Eredivisie$MarketValue)

# aggregate log market values by position
means <- aggregate(Eredivisie[, "logMarketValue", drop = FALSE],
                   Eredivisie[, "Position", drop = FALSE],
                   FUN = mean)

# create profile plot
line_plot(means, "logMarketValue", "Position")

# easier and fancier as the plot method of ANOVA results
oneway <- ANOVA(Eredivisie, "logMarketValue",
                group = "Position")
plot(oneway)
```

palette_SPSS                    *SPSS Color Palette and Color Scales*

### Description

Color palette used by SPSS, and discrete color scales to be used in plots (e.g., for multiple lines in a plot) to mimic the look of SPSS graphs.

### Usage

```
palette_SPSS(n = NULL, version = r2spss_options$get("version"))

paletteSPSS(n = NULL, version = r2spss_options$get("version"))

SPSS_pal(version = r2spss_options$get("version"), direction = 1)

scale_color_SPSS(
  ...,
  version = r2spss_options$get("version"),
  direction = 1,
  aesthetics = "color"
)

scale_colour_SPSS(
  ...,
  version = r2spss_options$get("version"),
  direction = 1,
  aesthetics = "colour"
)

scale_fill_SPSS(
  ...,
  version = r2spss_options$get("version"),
  direction = 1,
  aesthetics = "fill"
)
```

### Arguments

| | |
|---|---|
| n | an integer giving the number of colors to select from the palette. If NULL (the default), all colors of the palette are returned. |
| version | a character string specifying whether to use the color palette of recent SPSS versions ("modern") or older versions (<24; "legacy"). |
| direction | an integer giving the direction to travel through the palette. Possible values are 1 for forward (the default) and -1 for backward. |
| ... | additional arguments to be passed to [discrete_scale](). |

aesthetics      a character string or vector listing the names of the aesthetics with which the
                scale works. For example, color settings can be applied to the `color` and `fill`
                aesthetics by supplying `c("color", "fill")`.

### Value

`palette_SPSS` returns a character vector specifying up to 30 colors as used by SPSS.

`SPSS_pal` returns a function that generates colors from the specified SPSS color palette, in the specified direction. It is mainly used internally by the discrete color scales.

`scale_color_SPSS`, `scale_colour_SPSS`, and `scale_fill_SPSS` return a discrete color scale to be added to plots.

### Author(s)

Andreas Alfons

### Examples

```
# data to be plotted
df <- data.frame(x = 1:30, y = 0)

# initialize plot
p <- ggplot(aes(x = x, y = y, fill = factor(x)), data = df) +
  geom_point(shape = 21, size = 3, show.legend = FALSE) +
  theme_SPSS()

# colors of modern SPSS versions
p + theme_SPSS() + scale_fill_SPSS()

# colors of legacy SPSS versions
p + theme_SPSS(version = "legacy") +
  scale_fill_SPSS(version = "legacy")
```

---

r2spss-deprecated      *Deprecated plot functions in r2spss*

---

### Description

These plot functions are deprecated and may be removed as soon as the next release of **r2spss**. The functions `plotSPSS`, `linesSPSS`, `boxSPSS`, and `histSPSS` are built around base R graphics and have been superseded by functions built on [ggplot2](#).

## Usage

```
plotSPSS(data, variables, xlab = NULL, ylab = NULL, ...)

linesSPSS(data, variables, index = NULL, xlab = NULL, ylab = NULL, ...)

boxplotSPSS(
  data,
  variables,
  group = NULL,
  xlab = NULL,
  ylab = NULL,
  cut.names = NULL,
  ...
)

histSPSS(data, variable, normal = FALSE, xlab = NULL, ylab = NULL, ...)
```

## Arguments

| | |
|---|---|
| data | a data frame containing the variables to be plotted. |
| variables | For plotSPSS, a character vector specifying at least two variables to be plotted. In case of two variables, a simple scatter plot is produced with the first variable on the $x$-axis and the second variable on the $y$-axis. In case of more than two variables, a scatter plot matrix is produced. |
| | For linesSPSS, a character vector specifying at least one variable to be plotted on the $y$-axis. In case of multiple variables, separate lines are drawn for each variable and a legend is shown. |
| | For boxplotSPSS, a character vector specifying separate variables to be plotted. If group is not NULL, only the first variable is used and box plots of groups of observations are drawn instead. |
| xlab, ylab | the axis labels. |
| ... | additional arguments to be passed down, in particular graphical parameters (see boxplot, hist, and par). |
| index | a character string specifying a variable to be plotted on the $x$-axis, or NULL to plot the observations against their index. |
| group | an character string specifying a grouping variable, or NULL for no grouping. |
| cut.names | a logical indicating whether to cut long variable names or group labels to 8 characters. The default is TRUE for box plots of separate variables, but FALSE for box plots of groups of observations (which mimics SPSS behavior). |
| variable | a character string specifying the variable to be plotted. |
| normal | a logical indicating whether to add a normal density with the estimated mean and standard deviation (the default is FALSE). |

## Details

plotSPSS draws a scatter plot or a scatter plot matrix of variables in a data frame.

linesSPSS draws connected lines for variables in a data frame.

boxplotSPSS draw box plots of variables in a data frame, including box plots for groups of observations and box plots for separate variables.

histSPSS draws a histogram of a variable in a data frame.

The plots thereby mimic the look of graphs in older versions of SPSS (<24).

### Value

plotSPSS and linesSPSS do not return anything but produce a plot.

boxplotSPSS returns a list containing summary statistics invisibly (see [boxplot](#)) and produces a plot.

histSPSS returns an object of class "histogram" invisibly (see [hist](#)) and produces a plot.

### Author(s)

Andreas Alfons

---

r2spss.sty *Create the LaTeX style file for 'r2spss'*

---

### Description

Create the LaTeX style file required to compile LaTeX documents that include tables created by package **r2spss**. You can put the resulting file *r2spss.sty* in the folder containing your LaTeX document, and you should include

```
\usepackage{r2spss}
```

in the preamble of your LaTeX document.

### Usage

```
r2spss.sty(path = NULL)
```

### Arguments

path        a character string specifying the path to the folder in which to put the style file, or NULL (the default) to print the contents of the style file to the standard output connection (usually the R console).

### Value

Nothing is returned, the function is called for its side effects.

### Author(s)

Andreas Alfons

### Examples

```
# print contents of style file
r2spss.sty()

## Not run:

  # put file 'r2spss.sty' in the current working directory
  r2spss.sty(".")

## End(Not run)
```

---

r2spss_options *Options for package r2spss*

---

### Description

Retrieve or set global options for package **r2spss** (within the current R session) via accessor functions.

### Usage

```
r2spss_options
```

### Format

A list with the following two components:

get(which, drop = TRUE) an accessor function to retrieve current options, which are usually returned as a named list. Argument which allows to select which options to retrieve. If a single option is selected, argument drop indicates whether only its value should be returned (TRUE) or a list of length one (FALSE).

set(...) an accessor function to set certain options using name = value pairs.

### Details

The following options are available:

version a character string that controls the default for whether tables and plots should mimic the content and look of recent SPSS versions ("modern") or older versions (<24; "legacy").

minor a logical that overrides whether function [to_latex](#) should include any supplied minor grid lines in SPSS tables. In particular for tables that mimic older SPSS versions, minor grid lines can be somewhat distracting from the content, so setting this option to FALSE provides a quick way to suppress them. The look of the resulting tables still closely mimics SPSS while being somewhat cleaner.

### Author(s)

Andreas Alfons

## Examples

```
# retrieve the list of options
r2spss_options$get()

# retrieve a single option
r2spss_options$get("version")

## Not run:

# set options
r2spss_options$set(version = "legacy", minor = FALSE)

## End(Not run)
```

---

regression                     *Linear Regression*

---

## Description

Perform linear regression on variables of a data set. The output is printed as a LaTeX table that mimics the look of SPSS output, and plots of the results mimic the look of SPSS graphs.

## Usage

```
regression(..., data, labels = NULL)

## S3 method for class 'regression_SPSS'
to_SPSS(
  object,
  statistics = c("estimates", "anova", "summary"),
  change = FALSE,
  version = r2spss_options$get("version"),
  ...
)

## S3 method for class 'regression_SPSS'
print(
  x,
  statistics = c("summary", "anova", "estimates"),
  change = FALSE,
  version = r2spss_options$get("version"),
  ...
)

## S3 method for class 'regression_SPSS'
coef(object, ...)
```

```
## S3 method for class 'regression_SPSS'
df.residual(object, ...)

## S3 method for class 'regression_SPSS'
fitted(object, standardized = FALSE, ...)

## S3 method for class 'regression_SPSS'
residuals(object, standardized = FALSE, ...)

## S3 method for class 'regression_SPSS'
plot(
  x,
  y,
  which = c("histogram", "scatter"),
  version = r2spss_options$get("version"),
  ...
)
```

## Arguments

| | |
|---|---|
| `...` | for `regression`, at least one formula specifying a regression model. Different models can be compared by supplying multiple formulas. For the `to_SPSS` and `print` methods, additional arguments to be passed down to [format_SPSS](#). For the `plot` method, additional arguments to be passed down to [histSPSS](#) or [plotSPSS](#), in particular graphical parameters. For other methods, this is currently ignored. |
| `data` | a data frame containing the variables. |
| `labels` | a character or numeric vector giving labels for the regression models in the output tables. |
| `object, x` | an object of class `"regression_SPSS"` as returned by function `regression`. |
| `statistics` | a character string or vector specifying which SPSS tables to produce. Available options are `"summary"` for model summaries, `"anova"` for ANOVA results, and `"estimates"` for estimated coefficients. For the `to_SPSS` method, only one option is allowed (the default is the table of ANOVA results), but the `print` method allows several options (the default is to print all tables). |
| `change` | a logical indicating whether tests on the $R^2$ change should be included in the table with model summaries (if `statistics = "summary"`). The default is FALSE. |
| `version` | a character string specifying whether the table or plot should mimic the content and look of recent SPSS versions (`"modern"`) or older versions (<24; `"legacy"`). For the table, the main difference in terms of content is that small p-values are displayed differently. |
| `standardized` | a logical indicating whether to return standardized residuals and fitted values (TRUE), or residuals and fitted values on their original scale (FALSE). |
| `y` | ignored (only included because it is defined for the generic function [plot](#)). |
| `which` | a character string specifying which plot to produce. Possible values are `"histogram"` for a histogram of the residuals, or `"scatter"` for a scatterplot of the standardized residuals against the standardized fitted values. |

**Details**

The print method first calls the to_SPSS method followed by to_latex. Further customization can be done by calling those two functions separately, and modifying the object returned by to_SPSS.

**Value**

An object of class "regression_SPSS" with the following components:

models a list in which each component is an ojbect of class "lm" as returned by function lm.

summaries a list in which each component is an ojbect of class "summary.lm" as returned by the summary method for objects of class "lm".

response a character string containing the name of the response variable.

method a character string specifying whether the nested models are increasing in dimension by entering additional variables ("enter") or decreasing in dimension by removing variables ("remove").

The to_SPSS method returns an object of class "SPSS_table" which contains all relevant information in the required format to produce the LaTeX table. See to_latex for possible components and how to further customize the LaTeX table based on the returned object.

The print method produces a LaTeX table that mimics the look of SPSS output.

The coef, df.residual, fitted and residuals methods return the coefficients, residual degrees of freedom, fitted values and residuals, respectively, of the *last* model (to mimic SPSS functionality).

Similarly, the plot method returns the specified plot for the *last* model as an object of class "ggplot", which produces the plot when printed.

**Note**

LaTeX tables that mimic recent versions of SPSS (version = "modern") may require several LaTeX compilations to be displayed correctly.

**Author(s)**

Andreas Alfons

**Examples**

```
# load data
data("Eredivisie")
# log-transform market values
Eredivisie$logMarketValue <- log(Eredivisie$MarketValue)
# squared values of age
Eredivisie$AgeSq <- Eredivisie$Age^2

# simple regression model of log market value on age
fit1 <- regression(logMarketValue ~ Age, data = Eredivisie)
fit1                         # print LaTeX table
plot(fit1, which = "scatter")  # diagnostic plot
```

```
# add a squared effect for age
fit2 <- regression(logMarketValue ~ Age + AgeSq,
                   data = Eredivisie, labels = 2)
fit2                            # print LaTeX table
plot(fit2, which = "scatter")  # diagnostic plot

# more complex models with model comparison
fit3 <- regression(logMarketValue ~ Age + AgeSq,
                   logMarketValue ~ Age + AgeSq + Contract +
                                    Foreign,
                   logMarketValue ~ Age + AgeSq + Contract +
                                    Foreign + Position,
                   data = Eredivisie, labels = 2:4)
print(fit3, change  = TRUE)      # print LaTeX table
plot(fit3, which = "histogram")  # diagnostic plot
```

scatter_plot                *Scatter Plot and Scatter Plot Matrix*

### Description

Draw a scatter plot or a scatter plot matrix of variables in a data frame. The plots thereby mimic the look of SPSS graphs.

### Usage

```
scatter_plot(data, variables, version = r2spss_options$get("version"), ...)
```

### Arguments

| | |
|---|---|
| data | a data frame containing the variables to be plotted. |
| variables | a character vector specifying at least two variables to be plotted. In case of two variables, a simple scatter plot is produced with the first variable on the $x$-axis and the second variable on the $y$-axis. In case of more than two variables, a scatter plot matrix is produced. |
| version | a character string specifying whether the plot should mimic the look of recent SPSS versions ("modern") or older versions (<24; "legacy"). |
| ... | for a simple scatter plot, additional arguments are passed down to geom_point. For a scatter plot matrix, additional arguments to be passed down, in particular base graphics parameters (see par). |

### Value

In case of a simple scatter plot, an object of class "ggplot", which produces the plot when printed.

In case of a scatter plot matrix, nothing is returned but a plot is produced.

## Note

Wile all other plots in **r2spss** are based on **ggplot2** (including the simple scatter plot), the scatter plot matrix is built around base R graphics. This is because **ggplot2** does not provide an implementation of a scatter plot matrix, and an implementation based on separate scatter plots on a matrix layout would be slow.

## Author(s)

Andreas Alfons

## Examples

```
# load data
data("Eredivisie")
# log-transform market values
Eredivisie$logMarketValue <- log(Eredivisie$MarketValue)

# plot log market values against age
scatter_plot(Eredivisie, c("Age", "logMarketValue"))

# scatterplot matrix of age, number of minutes played, and
# log market values
scatter_plot(Eredivisie, c("Age", "Minutes", "logMarketValue"))
```

---

| sign_test | *Sign Test* |
|---|---|

---

## Description

Perform a sign test for a paired sample on variables of a data set. The output is printed as a LaTeX table that mimics the look of SPSS output.

## Usage

```
sign_test(data, variables, exact = FALSE)

## S3 method for class 'sign_test_SPSS'
to_SPSS(
  object,
  statistics = c("test", "frequencies"),
  version = r2spss_options$get("version"),
  ...
)

## S3 method for class 'sign_test_SPSS'
print(
  x,
```

```
  statistics = c("frequencies", "test"),
  version = r2spss_options$get("version"),
  ...
)

signTest(data, variables, exact = FALSE)
```

## Arguments

| | |
|---|---|
| data | a data frame containing the variables. |
| variables | a character vector specifying two numeric variables containing the paired observations. |
| exact | a logical indicating whether or not to include the exact p-value using the binomial distribution. Note that the p-value using the normal approximation is always reported. |
| object, x | an object of class "sign_test_SPSS" as returned by function sign_test. |
| statistics | a character string or vector specifying which SPSS tables to produce. Available options are "frequencies" for a summary of the frequencies and "test" for test results. For the to_SPSS method, only one option is allowed (the default is the table of test results), but the print method allows several options (the default is to print all tables). |
| version | a character string specifying whether the table should mimic the content and look of recent SPSS versions ("modern") or older versions (<24; "legacy"). The main difference in terms of content is that small p-values are displayed differently. |
| ... | additional arguments to be passed down to [format_SPSS](). |

## Details

The print method first calls the to_SPSS method followed by [to_latex](). Further customization can be done by calling those two functions separately, and modifying the object returned by to_SPSS.

## Value

An object of class "sign_test_SPSS" with the following components:

statistics a data frame containing information on the number of observations with negative and positive differences.

asymptotic a list containing the results of the test using the normal approximation.

exact if requested, a numeric vector containing the exact two-sided p-value, one-sided p-value, and point probability using the binomial distribution.

variables a character vector containing the names of the two numeric variables with the paired observations.

n an integer giving the number of observations.

The to_SPSS method returns an object of class "SPSS_table" which contains all relevant information in the required format to produce the LaTeX table. See [to_latex](#) for possible components and how to further customize the LaTeX table based on the returned object.

The print method produces a LaTeX table that mimics the look of SPSS output.

### Note

LaTeX tables that mimic recent versions of SPSS (version = "modern") may require several LaTeX compilations to be displayed correctly.

### Author(s)

Andreas Alfons

### Examples

```
# load data
data("Exams")

# test whether grades differ between the
# regular exam and the resit
sign_test(Exams, c("Regular", "Resit"))
```

---

theme_SPSS                    *Plot theme to mimic the look of SPSS graphs*

---

### Description

Complete theme that controls all non-data display of a plot to mimic the look of SPSS graphs. Use [theme](#) after theme_SPSS to further tweak the display.

### Usage

```
theme_SPSS(
  base_size = 12,
  base_family = "",
  base_line_size = 0.5,
  base_rect_size = 0.5,
  version = r2spss_options$get("version"),
  scales = NULL,
  scale.x = scales,
  scale.y = scales
)
```

## Arguments

base_size        an integer giving the base font size in pts.

base_family      a character string giving the base font family.

base_line_size   base size for line elements.

base_rect_size   base size for borders of rectangle elements.

version          a character string specifying whether to mimic the look of recent SPSS versions
                 ("modern") or older versions (<24; "legacy").

scales, scale.x, scale.y

                 a character string specifying whether both or each of the axes are expected to
                 be continuous ("continuous") or discrete ("discrete"). Note that this only
                 controls the appearance of the tick labels on the axis, as the theme has no control
                 over the information that is displayed in the plot. SPSS displays larger tick labels
                 a for discrete axis than for a continuous axis, hence specifying this information
                 in the theme adjusts the tick label size accordingly. The default (NULL) means
                 that the tick labels are of the same size as for a continuous axis.

## Examples

```
# data to be plotted
df <- data.frame(x = 1:30, y = 0)

# initialize plot
p <- ggplot(aes(x = x, y = y, fill = factor(x)), data = df) +
  geom_point(shape = 21, size = 3, show.legend = FALSE) +
  theme_SPSS()

# colors of modern SPSS versions
p + theme_SPSS() + scale_fill_SPSS()

# colors of legacy SPSS versions
p + theme_SPSS(version = "legacy") +
  scale_fill_SPSS(version = "legacy")
```

---

to_latex                 *Print LaTeX Tables that Mimic the Look of SPSS Output*

---

## Description

Use information from an R object to print a LaTeX table that mimics the look of SPSS output.
Typically, one would first call to_SPSS with an object returned by a function in **r2spss**, and then
call to_latex with the resulting object of class "SPSS_table" to print the LaTeX table. Note that
the print methods in **r2spss** perform these two steps at once, but calling to_SPSS and to_latex
separately can be useful for customization of the LaTeX table.

## Usage

```
to_latex(object, ...)

## S3 method for class 'SPSS_table'
to_latex(object, version = r2spss_options$get("version"), ...)

## S3 method for class 'data.frame'
to_latex(
  object,
  main = NULL,
  sub = NULL,
  header = TRUE,
  label = NULL,
  row_names = TRUE,
  info = NULL,
  alignment = NULL,
  border = NULL,
  footnotes = NULL,
  major = NULL,
  minor = NULL,
  version = r2spss_options$get("version"),
  ...
)
```

## Arguments

| | |
|---|---|
| object | an object of class "SPSS_table" as returned by to_SPSS methods, or a data.frame. |
| ... | for the "data.frame" method, additional arguments to be passed to [format_SPSS](). For the "SPSS_table" method, additional arguments are currently ignored. |
| version | a character string specifying whether the table should mimic the look of recent SPSS versions ("modern") or older versions (<24; "legacy"). For the "SPSS_table" method, note that also the *content* of some tables generated by functions in **r2spss** is different for current and older SPSS versions. These objects contain a component "version" which will passed to the "data.frame" method to ensure that the content and look of the table match. Other tables have the same content irrespective of the SPSS version, and this argument controls the look of those tables. The default is to inherit from the global option within the current R session (see [r2spss_options]()). |
| main | a single character string defining the main title of the SPSS table, or NULL to suppress the main title. |
| sub | a single character string defining the sub-title of the SPSS table, or NULL to suppress the sub-title. |
| header | a logical indicating whether to include a header in the SPSS table based on the column names of object (defaults to TRUE). Alternatively, it is possible to supply a character vector giving the header of each column, or a list defining a complex header layout with merged header cells. In the latter case, the list can have up to three components, with each component defining one level of the |

header. The last list component should be a character vector giving the bottom-level header of each column. The other list components should be data frames with the following columns:

first an integer vecot giving the first column of each (merged) header cell.

last an integer vector giving the last column of each (merged) header cell.

text a character vector containing the text of each (merged) header cell.

Line breaks (character \n) can be included to wrap the text of a header cell over several rows.

label            a character string giving a label to be added as the first column of the table, or NULL to suppress such a column. In many SPSS tables, this contains the name of a variable used in the analysis.

row_names        a logical indicating whether to add the row names of object as a column in the SPSS table (defaults to TRUE). Alternatively, it is possible to supply a character vector giving the row labels to be added as a column. Line breaks (character \n) can be included to wrap the text of a row label over several rows.

info             an integer giving the number of columns in the SPSS table that contain auxiliary information on the results. This has an effect of the default formatting, alignment, and borders. The default is 0 if row_names is FALSE and 1 otherwise. Note that a column defined by label and a column defined by row_names are always added to info if the former are supplied.

alignment        a list with components header and table, with each component being a character vector that contains the LaTeX alignment specifiers of the header and table body, respectively, of each column. Permissible alignment specifiers are "l" for left-aligned, "c" for centered, and "r" for right aligned. The default is left-aligned for the header and table body of the columns containing auxiliary information, and centered and right-aligned, respectively, for the header and table body of the columns containing the actual results. It should not be necessary to set the column alignment manually.

border           a logical vector indicating which (outer and inner) vertical borders should be drawn. The default is that tables that mimic recent versions of SPSS (version = "modern") draw only borders in between columns that contain the actual results, whereas tables that mimic older versions of SPSS (version = "legacy") draw all borders except in between columns containing auxiliary information. It should not be necessary to set the vertical borders manually.

footnotes        a character vector giving footnotes to be added below the SPSS table, or NULL to suppress footnotes. Alternatively, it is possible to supply a data frame with the following columns:

marker character vector giving footnote markers to be included in a cell of the SPSS table. For footnotes without a marker, an empty character string can be used.

row an integer vector specifying the row of the SPSS table in which to include each footnote marker, or NA for footnotes without a marker. In addition, the character strings "main" and "sub" can be used to include footnote markers in the main title and sub-title, respectively.

column an integer vector specifying the column of the SPSS table in which to include each footnote marker, or NA for footnotes without a marker or footnote markers in the main title or sub-title.

text a character vector containing the text of each footnote.

major, minor an integer vector specifying the rows of the SPSS table after which to draw major or minor grid lines that stretch across all columns of the table, or NULL to suppress the respective grid lines. Alternatively, each of these arguments can be a data frame with the following columns defining partial grid lines:

row an integer vector specifying the rows of the SPSS table after which to draw grid lines.

first an integer vector specifying the first column of each partial line.

last an integer vector specifying the last column of each partial line.

The only difference between the two type of grid lines is that minor grid lines can also be suppressed globally within the current R session by setting r2spss_options$set(minor = FALSE), see [r2spss_options](#). Also note that these arguments only control the grid lines in between rows of the table body. Horizontal table borders are always drawn.

## Details

The "SPSS_table" method takes component table of the object and supplies it to the data.frame method, with additional components in the object being passed as additional arguments.

The "data.frame" method allows to extend the functionality of **r2spss** with additional LaTeX tables that mimic the look of SPSS output.

## Value

Nothing is returned, the function is called for its side effects.

## Note

LaTeX tables that mimic recent versions of SPSS (version = "modern") may require several LaTeX compilations to be displayed correctly.

## Author(s)

Andreas Alfons

## Examples

```
## Kruskal-Wallis test example

# load data
data("Eredivisie")

# compute a Kruskual-Wallis test to investigate whether
# market values differ by playing position
kw <- kruskal_test(Eredivisie, "MarketValue",
                   group = "Position")
```

```
# convert to an object of class "SPSS_table" that
# contains the table with the test results
kw_spss <- to_SPSS(kw, statistics = "test")
kw_spss

# blank out the number of degrees of freedom to ask
# an assignment question about it
kw_spss$table[2, 1] <- "???"

# print the LaTeX table to be included in the assignment
to_latex(kw_spss)


## t test example

# load data
data("Exams")

# test whether the average grade on the resit
# differs from 5.5 (minimum passing grade)
t <- t_test(Exams, "Resit", mu = 5.5)

# convert to an object of class "SPSS_table" that
# contains the table with the test results
t_spss <- to_SPSS(t, statistics = "test")

# this is an example of a complex header layout
t_spss$header

# add additional line breaks in bottom-level header
t_spss$header[[3]] <- gsub("-", "-\n", t_spss$header[[3]],
                           fixed = TRUE)

# print the LaTeX table
to_latex(t_spss)
```

---

to_SPSS                          *Convert R Objects to SPSS-Style Tables*

---

### Description

Generic function to convert an R object into an object that contains all necessary information for printing a LaTeX table that mimics the look of SPSS output.

### Usage

```
to_SPSS(object, ...)

to_spss(object, ...)
```

## Arguments

| | |
|---|---|
| `object` | an R object for which a `to_SPSS` method exists, such as objects returned by functions in **r2spss**. |
| `...` | additional arguments passed down to methods. |

## Value

In order to work as expected, methods of `to_SPSS` should return an object of class `"SPSS_table"`. It should include a component `table` that contains a data frame, which can be supplied as the first argument to `to_latex` to print a LaTeX table that mimics the look of SPSS output. Additional components of the returned object define additional arguments to be passed to the `"data.frame"` method of `to_latex`.

## Note

`to_spss` is a simple wrapper for `to_SPSS`, which exists for convenience.

## Author(s)

Andreas Alfons

## Examples

```
# load data
data("Eredivisie")

# compute a Kruskual-Wallis test to investigate whether
# market values differ by playing position
kw <- kruskal_test(Eredivisie, "MarketValue",
                   group = "Position")

# convert to an object of class "SPSS_table" that
# contains the table with the test results
kw_spss <- to_SPSS(kw, statistics = "test")
kw_spss

# blank out the number of degrees of freedom to ask
# an assignment question about it
kw_spss$table[2, 1] <- "???"

# print the LaTeX table to be included in the assignment
to_latex(kw_spss)
```

---

trimmed_mean *Trimmed mean*

---

#### Description

Compute the trimmed mean. This function differs from the implementation of the trimmed mean in the base R function mean in the following ways. While mean always rounds down the number of observations to be trimmed, this function rounds to the nearest integer. In addition, mean implements proper NA handling, whereas this function assumes that there are no missing values and may fail in their presence.

#### Usage

```
trimmed_mean(x, trim = 0.05)
```

#### Arguments

| | |
|---|---|
| x | a numeric vector. |
| trim | numeric; the fraction of observations to be trimmed from each tail of x before computing the mean (defaults to 0.05). |

#### Details

The main purpose of this function is to reproduce SPSS results for Levene's test on homogeneity of the variances based on the trimmed mean (see ANOVA), which are slightly too far off when using the base R function mean. Rounding the number of observations to be trimmed to the nearest integer brings the results closer to those of SPSS, but they are still not identical.

#### Value

The trimmed mean of the values in x as a single numeric value.

#### Author(s)

Andreas Alfons

#### See Also

mean

#### Examples

```
x <- c(0:10, 50)

# trimmed_mean() rounds number of observations
# to be trimmed to the nearest integer
trimmed_mean(x, trim = 0.05)
```

```
# base R function mean() rounds down number of
# observations to be trimmed
mean(x, trim = 0.05)
mean(x)
```

---

t_test                          *t Tests*

---

### Description

Perform a one-sample t test, a paired-sample t test or an independent-samples t test on variables of a data set. The output is printed as a LaTeX table that mimics the look of SPSS output.

### Usage

```
t_test(data, variables, group = NULL, mu = 0, conf.level = 0.95)

## S3 method for class 't_test_SPSS'
to_SPSS(
  object,
  statistics = c("test", "statistics"),
  version = r2spss_options$get("version"),
  digits = 3,
  ...
)

## S3 method for class 't_test_SPSS'
print(
  x,
  statistics = c("statistics", "test"),
  version = r2spss_options$get("version"),
  digits = 3,
  ...
)

tTest(data, variables, group = NULL, mu = 0, conf.level = 0.95)
```

### Arguments

data            a data frame containing the variables.

variables       a character vector specifying numeric variable(s) to be used for testing the mean(s).
                If group is NULL, a one-sample t test is performed if only one variable is spec-
                ified, and a paired-sample t test is performed if two variables are specified. If
                a grouping variable is specified in group, an independent-samples t-test is per-
                formed and this should be a character string specifying the numeric variable of
                interest.

| group | a character string specifying a grouping variable for an independent-samples t-test, or NULL. |
|---|---|
| mu | a number indicating the true value of the mean for a one-sample t test. |
| conf.level | a number between 0 and 1 giving the confidence level of the confidence interval. |
| object, x | an object of class "t_test_SPSS" as returned by function t_test. |
| statistics | a character string or vector specifying which SPSS tables to produce. Available options are "statistics" for descriptive statistics and "test" for test results. For the to_SPSS method, only one option is allowed (the default is the table of test results), but the print method allows several options (the default is to print all tables). |
| version | a character string specifying whether the table should mimic the content and look of recent SPSS versions ("modern") or older versions (<24; "legacy"). The main differences in terms of content are that recent SPSS versions show a one-sided p-value in addition to the two-sided p-value, and that small p-values are displayed differently. For the paired-sample test, recent versions of SPSS also display a label *Pair 1* for the selected pair of variables. |
| digits | an integer giving the number of digits after the comma to be printed in the SPSS tables. |
| ... | additional arguments to be passed down to [format_SPSS](). |

## Details

The print method first calls the to_SPSS method followed by [to_latex](). Further customization can be done by calling those two functions separately, and modifying the object returned by to_SPSS.

## Value

An object of class "t_test_SPSS" with the following components:

statistics a data frame containing the relevant descriptive statistics.

test an object of class "htest" as returned by [t.test]() (only one-sample and paired-sample tests).

variables a character vector containing the name(s) of the relevant numeric variable(s).

n an integer giving the number of observations (only paired-sample test).

levene an object as returned by [leveneTest]() (only independent-samples test).

pooled an object of class "htest" as returned by [t.test]() assuming equal variances (only independent-samples test).

satterthwaite an object of class "htest" as returned by [t.test]() not assuming equal variance (only independent-samples test).

group a character string containing the name of the grouping variable (only independent-samples test).

type a character string giving the type of t test performed ("one-sample", "paired", or "independent").

The to_SPSS method returns an object of class "SPSS_table" which contains all relevant information in the required format to produce the LaTeX table. See [to_latex]() for possible components and how to further customize the LaTeX table based on the returned object.

The print method produces a LaTeX table that mimics the look of SPSS output.

## Note

LaTeX tables that mimic recent versions of SPSS (version = "modern") may require several La-TeX compilations to be displayed correctly.

## Author(s)

Andreas Alfons

## Examples

```
## one-sample and paired-sample t test

# load data
data("Exams")

# test whether the average grade on the resit
# differs from 5.5 (minimum passing grade)
t_test(Exams, "Resit", mu = 5.5)

# test whether average grades differ between the
# regular exam and the resit
t_test(Exams, c("Resit", "Regular"))


## independent-samples t test

# load data
data("Eredivisie")
# log-transform market values
Eredivisie$logMarketValue <- log(Eredivisie$MarketValue)

# test whether average log market values differ between
# Dutch and foreign players
t_test(Eredivisie, "logMarketValue", group = "Foreign")
```

---

wilcoxon_test *Wilcoxon Signed Rank and Rank Sum Tests*

---

## Description

Perform a Wilcoxon signed rank test for a paired sample or a Wilcoxon rank sum test for independent samples on variables of a data set. The output is printed as a LaTeX table that mimics the look of SPSS output.

**Usage**

```
wilcoxon_test(data, variables, group = NULL, exact = FALSE)

## S3 method for class 'wilcoxon_test_SPSS'
to_SPSS(
  object,
  statistics = c("test", "ranks"),
  version = r2spss_options$get("version"),
  digits = NULL,
  ...
)

## S3 method for class 'wilcoxon_test_SPSS'
print(
  x,
  statistics = c("ranks", "test"),
  version = r2spss_options$get("version"),
  digits = 2:3,
  ...
)

wilcoxonTest(data, variables, group = NULL, exact = FALSE)
```

**Arguments**

| | |
|---|---|
| data | a data frame containing the variables. |
| variables | a character vector specifying numeric variable(s) to be used. If group is NULL, the Wilcoxon signed rank test is performed and this should be a character vector specifying two numeric variables which contain the paired observations. If a grouping variable is specified in group, the Wilcoxon rank sum test is performed and this should be a character string specifying the numeric variable of interest. |
| group | a character string specifying a grouping variable for the Wilcoxon rank sum test, or NULL. |
| exact | a logical indicating whether the Wilcoxon rank sum test should also return the p-value of the exact test. The default is FALSE. Note that the p-value of the asymptotic test is always returned. |
| object, x | an object of class "wilcoxon_test_SPSS" as returned by function wilcoxon_test. |
| statistics | a character string or vector specifying which SPSS tables to produce. Available options are "ranks" for a summary of the ranks and "test" for test results. For the to_SPSS method, only one option is allowed (the default is the table of test results), but the print method allows several options (the default is to print all tables). |
| version | a character string specifying whether the table should mimic the content and look of recent SPSS versions ("modern") or older versions (<24; "legacy"). The main difference in terms of content is that small p-values are displayed differently. |

digits            for the to_SPSS method, an integer giving the number of digits after the comma to be printed in the SPSS table. For the print method, this should be an integer vector of length 2, with the first element corresponding to the number of digits in table with the summary of the ranks, and the second element corresponding to the number of digits in the table for the test.

...              additional arguments to be passed down to [format_SPSS](format_SPSS).

## Details

The print method first calls the to_SPSS method followed by [to_latex](to_latex). Further customization can be done by calling those two functions separately, and modifying the object returned by to_SPSS.

## Value

An object of class "wilcoxon_test_SPSS" with the following components:

statistics a data frame containing the relevant information on the ranks.

test a list containing the results of the Wilcoxon signed rank test (only paired-sample test).

variables a character vector containing the name(s) of the relevant numeric variable(s).

n an integer giving the number of observations (only paired-sample test).

u numeric; the Mann-Whitney U test statistic (only independent-samples test).

w numeric; the Wilcoxon rank sum test statistic (only independent-samples test).

asymptotic a list containing the results of the Wilcoxon rank sum test using the normal approximation (only independent-samples test).

exact if requested, the corresponding p-value of the exact Wilcoxon rank sum test test (only independent-samples test).

group a character string containing the name of the grouping variable (only independent-samples test).

type a character string giving the type of Wilcoxon test performed "paired" or "independent").

The to_SPSS method returns an object of class "SPSS_table" which contains all relevant information in the required format to produce the LaTeX table. See [to_latex](to_latex) for possible components and how to further customize the LaTeX table based on the returned object.

The print method produces a LaTeX table that mimics the look of SPSS output.

## Note

The Wilcoxon rank sum test also reports the value of the equivalent Mann-Whitney U test statistic.

LaTeX tables that mimic recent versions of SPSS (version = "modern") may require several LaTeX compilations to be displayed correctly.

## Author(s)

Andreas Alfons

## Examples

```
## paired sample

# load data
data("Exams")

# test whether grades differ between the
# regular exam and the resit
wilcoxon_test(Exams, c("Regular", "Resit"))


## independent samples

# load data
data("Eredivisie")

# test whether market values differ between Dutch and foreign
# players
wilcoxon_test(Eredivisie, "MarketValue", group = "Foreign")
```

# Index