

Package ‘spicy’

April 11, 2026

Title Descriptive Statistics, Summary Tables, and Data Management Tools

Version 0.8.0

Description Provides tools for descriptive data analysis, variable inspection, data management, and tabulation workflows in 'R'. Summarizes variable metadata, labels, classes, missing values, and representative values, with support for readable frequency tables, cross-tabulations, association measures for contingency tables (Cramer's V, Phi, Goodman-Kruskal Gamma, Kendall's Tau-b, Somers' D, and others), categorical and continuous summary tables, and model-based bivariate tables for continuous outcomes, including APA-style reporting outputs. Includes helpers for interactive codebooks, variable label extraction, clipboard export, and row-wise descriptive summaries. Designed to make descriptive analysis and reporting faster, clearer, and easier to work with in practice.

License MIT + file LICENSE

URL <https://github.com/amaltawfik/spicy/>,
<https://amaltawfik.github.io/spicy/>

BugReports <https://github.com/amaltawfik/spicy/issues>

Encoding UTF-8

Language en-US

RoxygenNote 7.3.3

Imports crayon, dplyr, haven, labelled, rlang, stats, stringr, tibble, tidyselect, utils

Suggests clipr, DT, flextable, gt, knitr, officer, openxlsx2, rmarkdown, testthat (>= 3.0.0), tinytable, withr

VignetteBuilder knitr

Depends R (>= 4.1.0)

Config/testthat/edition 3

LazyData true

NeedsCompilation no

Author Amal Tawfik [aut, cre, cph] (ORCID:
<https://orcid.org/0009-0006-2422-1555>), ROR:
<https://ror.org/04j47fz63>)

Maintainer Amal Tawfik <amal.tawfik@hesav.ch>

Repository CRAN

Date/Publication 2026-04-10 22:30:02 UTC

Contents

assoc_measures	3
code_book	4
contingency_coef	5
copy_clipboard	6
count_n	8
cramer_v	12
cross_tab	13
freq	16
gamma_gk	19
goodman_kruskal_tau	20
kendall_tau_b	21
kendall_tau_c	22
label_from_names	24
lambda_gk	25
mean_n	26
phi	28
sochealth	29
somers_d	31
spicy_print_table	32
sum_n	34
table_categorical	37
table_continuous	41
table_continuous_lm	46
uncertainty_coef	51
varlist	53
yule_q	54
Index	56

assoc_measures *Association measures summary table*

Description

assoc_measures() computes a range of association measures for a two-way contingency table and returns them in a tidy data frame.

Usage

```
assoc_measures(
  x,
  type = c("all", "nominal", "ordinal"),
  conf_level = 0.95,
  digits = 3L
)
```

Arguments

x	A contingency table (of class table).
type	Which family of measures to compute: "all" (default), "nominal", or "ordinal".
conf_level	A number between 0 and 1 giving the confidence level (default 0.95). Set to NULL to omit the confidence interval.
digits	Number of decimal places used when printing the result (default 3).

Details

type = "all" (the default) returns all nominal and ordinal measures. Use type = "nominal" or type = "ordinal" to restrict the output to a single family.

The nominal family includes [cramer_v\(\)](#), [contingency_coef\(\)](#), [lambda_gk\(\)](#), [goodman_kruskal_tau\(\)](#), [uncertainty_coef\(\)](#), and (for 2x2 tables) [phi\(\)](#) and [yule_q\(\)](#).

The ordinal family includes [gamma_gk\(\)](#), [kendall_tau_b\(\)](#), [kendall_tau_c\(\)](#), and [somers_d\(\)](#).

Standard error formulas follow the DescTools implementations (Signorell et al., 2024).

Value

A data frame with columns measure, estimate, se, ci_lower, ci_upper, and p_value. For nominal measures (Cramer's V, Phi, Contingency Coef.), the p-value comes from the Pearson chi-squared test of independence. For all other measures, it is a Wald z-test of $H_0: \text{measure} = 0$.

References

- Agresti, A. (2002). *Categorical Data Analysis* (2nd ed.). Wiley.
- Liebetrau, A. M. (1983). *Measures of Association*. Sage.
- Signorell, A. et al. (2024). *DescTools: Tools for Descriptive Statistics*. R package.

See Also

[cramer_v\(\)](#), [gamma_gk\(\)](#), [kendall_tau_b\(\)](#)

Examples

```
tab <- table(sochealth$smoking, sochealth$education)
assoc_measures(tab)
assoc_measures(tab, type = "nominal")
assoc_measures(tab, type = "ordinal")
```

code_book

Generate an interactive variable codebook

Description

`code_book()` creates an interactive and exportable codebook summarizing all variables of a data frame. It builds upon [varlist\(\)](#) to provide an overview of variable names, labels, classes, and representative values in a sortable, searchable table.

The output is displayed as an interactive DT::datatable() in the Viewer pane, allowing filtering, column reordering, and export (copy, print, CSV, Excel, PDF) directly.

Usage

```
code_book(x, values = FALSE, include_na = FALSE, title = "Codebook", ...)
```

Arguments

<code>x</code>	A data frame or tibble.
<code>values</code>	Logical. If FALSE (the default), displays a compact summary of the variable's values. For numeric, character, date/time, labelled, and factor variables, up to four unique non-missing values are shown: the first three values, followed by an ellipsis (...), and the last value. Values are sorted when appropriate (e.g., numeric, character, date) For factors, the levels are used directly and are not sorted. For labelled variables, prefixed labels are displayed via <code>labelled::to_factor(levels = "prefixed")</code> . If TRUE, all unique non-missing values are displayed.
<code>include_na</code>	Logical. If TRUE, unique missing values (NA, NaN) are explicitly appended at the end of the Values summary when present in the variable. This applies to all variable types. If FALSE (the default), missing values are omitted from Values but still counted in the NAs column.
<code>title</code>	Optional character string displayed as the table title in the Viewer. Defaults to "Codebook". Set to NULL to remove the title completely.
<code>...</code>	Additional arguments (currently unused).

Details

- The interactive datatable supports column sorting, searching, and client-side export to various formats.
- All exports occur client-side through the Viewer or Tab.

Value

A `DT::datatable` object.

Dependencies

Requires the following package:

- **DT**

See Also

[varlist\(\)](#) for generating the underlying variable summaries.

Examples

```
## Not run:  
# Launch the interactive codebook (opens in Viewer)  
code_book(sochealth)  
  
## End(Not run)
```

contingency_coef	<i>Pearson's contingency coefficient</i>
------------------	------------------------------------------

Description

`contingency_coef()` computes Pearson's contingency coefficient C for a two-way contingency table.

Usage

```
contingency_coef(  
  x,  
  detail = FALSE,  
  conf_level = 0.95,  
  digits = 3L,  
  .include_se = FALSE  
)
```

Arguments

<code>x</code>	A contingency table (of class <code>table</code>).
<code>detail</code>	Logical. If <code>FALSE</code> (default), return the estimate as a numeric scalar. If <code>TRUE</code> , return a named numeric vector including confidence interval and p-value.
<code>conf_level</code>	A number between 0 and 1 giving the confidence level (default 0.95). Only used when <code>detail = TRUE</code> . Set to <code>NULL</code> to omit the confidence interval.
<code>digits</code>	Number of decimal places used when printing the result (default 3). Only affects the <code>detail = TRUE</code> output.
<code>.include_se</code>	Internal parameter; do not use.

Details

The contingency coefficient is $C = \sqrt{\chi^2/(\chi^2 + n)}$. It ranges from 0 (independence) to a maximum that depends on the table dimensions. No standard asymptotic standard error exists, so the confidence interval is not computed.

Value

Same structure as `cramer_v()`: a scalar when `detail = FALSE`, a named vector when `detail = TRUE`. The p-value tests the null hypothesis of no association (Pearson chi-squared test). CI values are NA because no standard asymptotic SE exists for C.

See Also

[cramer_v\(\)](#), [assoc_measures\(\)](#)

Examples

```
tab <- table(sochealth$smoking, sochealth$education)
contingency_coef(tab)
```

copy_clipboard

Copy data to the clipboard

Description

`copy_clipboard()` copies a data frame, matrix, array (2D or higher), table or vector to the clipboard. You can paste the result into a text editor (e.g. Notepad++, Sublime Text), a spreadsheet (e.g. Excel, LibreOffice Calc), or a word processor (e.g. Word).

Usage

```
copy_clipboard(
  x,
  row.names.as.col = FALSE,
  row.names = TRUE,
  col.names = TRUE,
  show_message = TRUE,
  quiet = FALSE,
  ...
)
```

Arguments

<code>x</code>	A data frame, matrix, 2D array, 3D array, table, or atomic vector to be copied.
<code>row.names.as.col</code>	Logical or character. If FALSE (the default), row names are not added as a column. If TRUE, a column named "rownames" is prepended. If a character string is supplied, it is used as the column name for row names.
<code>row.names</code>	Logical. If TRUE (the default), includes row names in the clipboard output. If FALSE, row names are omitted.
<code>col.names</code>	Logical. If TRUE (the default), includes column names in the clipboard output. If FALSE, column names are omitted.
<code>show_message</code>	Logical. If TRUE (the default), displays a success message after copying. If FALSE, no success message is printed.
<code>quiet</code>	Logical. If FALSE (the default), messages are shown. If TRUE, suppresses all messages, including success, coercion notices, and warnings.
<code>...</code>	Additional arguments passed to <code>clipr::write_clip()</code> .

Details

Note: Objects that are not data frames or 2D matrices (e.g. atomic vectors, arrays, tables) are automatically converted to character when copied to the clipboard, as required by `clipr::write_clip()`. The original object in R remains unchanged.

For multidimensional arrays (e.g. 3D arrays), the entire array is flattened into a 1D character vector, with each element on a new line. To preserve a tabular structure, you should extract a 2D slice before copying. For example: `copy_clipboard(my_array[, , 1])`.

Value

Invisibly returns the object `x`. The main purpose is the side effect of copying data to the clipboard.

Examples

```
if (clipr::clipr_available()) {
  # Data frame
  copy_clipboard(sochealth)
```

```

# Data frame with row names as column
copy_clipboard(head(sochealth), row.names.as.col = "id")

# Matrix
mat <- matrix(1:6, nrow = 2)
copy_clipboard(mat)

# Table
tbl <- table(sochealth$education)
copy_clipboard(tbl)

# Array (3D) - flattened to character
arr <- array(1:8, dim = c(2, 2, 2))
copy_clipboard(arr)

# Recommended: copy 2D slice for tabular layout
copy_clipboard(arr[, , 1])

# Numeric vector
copy_clipboard(c(3.14, 2.71, 1.618))

# Character vector
copy_clipboard(c("apple", "banana", "cherry"))

# Quiet mode (no messages shown)
copy_clipboard(sochealth, quiet = TRUE)
}

```

count_n

Row-wise Count of Specific or Special Values

Description

count_n() counts, for each row of a data frame or matrix, how many times one or more values appear across selected columns. It supports type-safe comparison, case-insensitive string matching, and detection of special values such as NA, NaN, Inf, and -Inf.

Usage

```

count_n(
  data = NULL,
  select = tidyselect::everything(),
  exclude = NULL,
  count = NULL,
  special = NULL,
  allow_coercion = TRUE,
  ignore_case = FALSE,
  regex = FALSE,

```

```

    verbose = FALSE
  )

```

Arguments

data	A data frame or matrix. Optional inside <code>mutate()</code> .
select	Columns to include. Defaults to <code>tidyselect::everything()</code> . Uses tidyselect helpers like <code>tidyselect::starts_with()</code> , etc. If <code>regex = TRUE</code> , select is treated as a regex string.
exclude	Character vector of column names to exclude after selection. Defaults to NULL (no exclusion).
count	Value(s) to count. Defaults to NULL. Ignored if <code>special</code> is used. Multiple values are allowed (e.g., <code>count = c(1, 2, 3)</code> or <code>count = c("yes", "no")</code>). R automatically coerces all values in <code>count</code> to a common type (e.g., <code>c(2, "2")</code> becomes <code>c("2", "2")</code>), so all values are expected to be of the same final type. If <code>allow_coercion = FALSE</code> , matching is type-safe using <code>identical()</code> , and the type of <code>count</code> must match that of the values in the data.
special	Character vector of special values to count: "NA", "NaN", "Inf", "-Inf", or "all". Defaults to NULL. "NA" uses <code>is.na()</code> , and therefore includes both NA and NaN values. "NaN" uses <code>is.nan()</code> to match only actual NaN values.
allow_coercion	Logical. If TRUE (the default), values are compared after coercion. If FALSE, uses strict matching via <code>identical()</code> .
ignore_case	Logical. If FALSE (the default), comparisons are case-sensitive. If TRUE, performs case-insensitive string comparisons.
regex	Logical. If FALSE (the default), uses tidyselect helpers. If TRUE, interprets select as a regular expression pattern.
verbose	Logical. If FALSE (the default), messages are suppressed. If TRUE, prints processing messages.

Details

This function is particularly useful for summarizing data quality or patterns in row-wise structures, and is designed to work fluently inside `dplyr::mutate()` pipelines.

Internally, `count_n()` wraps the stable and dependency-free base function `base_count_n()`, allowing high flexibility and testability.

Value

A numeric vector of row-wise counts (unnamed).

Note

This function is inspired by `datawizard::row_count()`, but provides additional flexibility:

- **Element-wise type-safe matching** using `identical()` when `allow_coercion = FALSE`. This ensures that both the value and its type match exactly, enabling precise comparisons in mixed-type columns.

- **Support for multiple values in count**, allowing queries like `count = c(2, 3)` or `count = c("yes", "no")` to count any of several values per row.
- **Detection of special values** such as NA, NaN, Inf, and -Inf through the special argument — a feature not available in `row_count()`.
- **Tidyverse-native behavior**: can be used inside `mutate()` without explicitly passing a data argument.

Value coercion behavior:

R automatically coerces mixed-type vectors passed to `count` into a common type. For example, `count = c(2, "2")` becomes `c("2", "2")`, because R converts numeric and character values to a unified type. This means that mixed-type checks are not possible at runtime once `count` is passed to the function. To ensure accurate type-sensitive matching, users should avoid mixing types in `count` explicitly.

Strict matching mode (`allow_coercion = FALSE`):

When strict matching is enabled, each value in `count` must match the type of the target column exactly.

For factor columns, this means that `count` must also be a factor. Supplying `count = "b"` (a character string) will not match a factor value, even if the label appears identical.

A common and intuitive approach is to use `count = factor("b")`, which works in many cases. However, `identical()` — used internally for strict comparisons — also checks the internal structure of the factor, including the order and content of its levels. As a result, comparisons may still fail if the levels differ, even when the label is the same.

To ensure a perfect match (label **and** levels), you can reuse a value taken directly from the data (e.g., `df$x[2]`). This guarantees that both the class and the factor levels align. However, this approach only works reliably if all selected columns have the same factor structure.

Case-insensitive matching (`ignore_case = TRUE`):

When `ignore_case = TRUE`, all values involved in the comparison are converted to lowercase using `tolower()` before matching. This behavior applies to both character and factor columns. Factors are first converted to character internally.

Importantly, this case-insensitive mode takes precedence over strict type comparison: values are no longer compared using `identical()`, but rather using lowercase string equality. This enables more flexible matching — for example, "b" and "B" will match even when `allow_coercion = FALSE`.

Example: strict vs. case-insensitive matching with factors:

```
df <- tibble::tibble(
  x = factor(c("a", "b", "c")),
  y = factor(c("b", "B", "a"))
)

# Strict match fails with character input
count_n(df, count = "b", allow_coercion = FALSE)
#> [1] 0 0 0

# Match works only where factor levels match exactly
count_n(df, count = factor("b", levels = levels(df$x)), allow_coercion = FALSE)
```

```
#> [1] 0 1 0

# Case-insensitive match succeeds for both "b" and "B"
count_n(df, count = "b", ignore_case = TRUE)
#> [1] 1 2 0
```

Like `datawizard::row_count()`, this function also supports regex-based column selection, case-insensitive string comparison, and column exclusion.

Examples

```
library(dplyr)
library(tibble)
library(haven)

# Basic usage
df <- tibble(
  x = c(1, 2, 2, 3, NA),
  y = c(2, 2, NA, 3, 2),
  z = c("2", "2", "2", "3", "2")
)
count_n(df, count = 2)
count_n(df, count = 2, allow_coercion = FALSE)
df |> mutate(num_twos = count_n(count = 2))

# Mixed types and special values
df <- tibble(
  num = c(1, 2, NA, -Inf, NaN),
  char = c("a", "B", "b", "a", NA),
  fact = factor(c("a", "b", "b", "a", "c")),
  date = as.Date(c("2023-01-01", "2023-01-01", NA, "2023-01-02", "2023-01-01")),
  lab = labelled(c(1, 2, 1, 2, NA), labels = c(No = 1, Yes = 2)),
  logic = c(TRUE, FALSE, NA, TRUE, FALSE)
)
count_n(df, count = 2)
count_n(df, count = "b", ignore_case = TRUE)
count_n(df, count = "a", select = fact)
count_n(df, count = as.Date("2023-01-01"), select = date)

# Count special values
count_n(df, special = "NA")

# Column selection strategies
df <- tibble(
  score_math = c(1, 2, 2, 3, NA),
  score_science = c(2, 2, NA, 3, 2),
  score_lang = c("2", "2", "2", "3", "2"),
  name = c("Jean", "Marie", "Ali", "Zoe", "Nina")
)
count_n(df, select = c(score_math, score_science), count = 2)
count_n(df, select = starts_with("score_"), exclude = "score_lang", count = 2)
count_n(df, select = "^score_", regex = TRUE, count = 2)
df |> mutate(nb_two = count_n(count = 2))
```

```

# Strict type-safe matching with factor columns
df <- tibble(
  x = factor(c("a", "b", "c")),
  y = factor(c("b", "B", "a"))
)

# Coercion: character "b" matches both x and y
count_n(df, count = "b")

# Strict match: fails because "b" is character, not factor (returns only 0s)
count_n(df, count = "b", allow_coercion = FALSE)

# Strict match with factor value: works only where levels match
count_n(df, count = factor("b", levels = levels(df$x)), allow_coercion = FALSE)

```

cramer_v

Cramer's V

Description

cramer_v() computes Cramer's V for a two-way contingency table, measuring the strength of association between two categorical variables.

Usage

```

cramer_v(
  x,
  detail = FALSE,
  conf_level = 0.95,
  digits = 3L,
  .include_se = FALSE
)

```

Arguments

x	A contingency table (of class table).
detail	Logical. If FALSE (default), return the estimate as a numeric scalar. If TRUE, return a named numeric vector including confidence interval and p-value.
conf_level	A number between 0 and 1 giving the confidence level (default 0.95). Only used when detail = TRUE. Set to NULL to omit the confidence interval.
digits	Number of decimal places used when printing the result (default 3). Only affects the detail = TRUE output.
.include_se	Internal parameter; do not use.

Details

Cramer's V is computed as $V = \sqrt{\chi^2 / (n \cdot (k - 1))}$, where χ^2 is the Pearson chi-squared statistic, n is the total count, and $k = \min(r, c)$. The confidence interval uses the Fisher z-transformation. Standard error formulas follow the DescTools implementations (Signorell et al., 2024).

Value

When `detail = FALSE`: a single numeric value (the estimate). When `detail = TRUE` and `conf_level` is non-NULL: `c(estimate, ci_lower, ci_upper, p_value)`. When `detail = TRUE` and `conf_level = NULL`: `c(estimate, p_value)`. The p-value tests the null hypothesis of no association (Pearson chi-squared test).

References

- Agresti, A. (2002). *Categorical Data Analysis* (2nd ed.). Wiley.
- Liebetrau, A. M. (1983). *Measures of Association*. Sage.
- Signorell, A. et al. (2024). *DescTools: Tools for Descriptive Statistics*. R package.

See Also

[phi\(\)](#), [contingency_coef\(\)](#), [assoc_measures\(\)](#)

Examples

```
tab <- table(sohealth$smoking, sohealth$education)
cramer_v(tab)
cramer_v(tab, detail = TRUE)
cramer_v(tab, detail = TRUE, conf_level = NULL)
```

cross_tab

Cross-tabulation

Description

Computes a two-way cross-tabulation with optional weights, grouping (including combinations of multiple variables), percentage displays, and inferential statistics.

`cross_tab()` produces weighted or unweighted contingency tables with row or column percentages, optional grouping via `by`, and associated Chi-squared tests with an association measure and diagnostic information.

Both `x` and `y` variables are required. For one-way frequency tables, use [freq\(\)](#) instead.

Usage

```

cross_tab(
  data,
  x,
  y = NULL,
  by = NULL,
  weights = NULL,
  rescale = FALSE,
  percent = c("none", "column", "row"),
  include_stats = TRUE,
  assoc_measure = c("auto", "cramer_v", "phi", "gamma", "tau_b", "tau_c", "somers_d",
    "lambda", "none"),
  assoc_ci = FALSE,
  correct = FALSE,
  simulate_p = FALSE,
  simulate_B = 2000,
  digits = NULL,
  styled = TRUE,
  show_n = TRUE
)

```

Arguments

<code>data</code>	A data frame. Alternatively, a vector when using the vector-based interface.
<code>x</code>	Row variable (unquoted).
<code>y</code>	Column variable (unquoted). Mandatory; for one-way tables, use freq() .
<code>by</code>	Optional grouping variable or expression. Can be a single variable or a combination of multiple variables (e.g. <code>interaction(vs, am)</code>).
<code>weights</code>	Optional numeric weights.
<code>rescale</code>	Logical. If FALSE (the default), weights are used as-is. If TRUE, rescales weights so total weighted N matches raw N.
<code>percent</code>	One of "none" (the default), "row", "column". Unique abbreviations are accepted (e.g. "n", "r", "c").
<code>include_stats</code>	Logical. If TRUE (the default), computes Chi-squared and an association measure (see <code>assoc_measure</code>).
<code>assoc_measure</code>	Character. Which association measure to report. "auto" (default) selects Kendall's Tau-b when both variables are ordered factors and Cramer's V otherwise. Other choices: "cramer_v", "phi", "gamma", "tau_b", "tau_c", "somers_d", "lambda", "none".
<code>assoc_ci</code>	Logical. If TRUE, includes the 95 percent confidence interval of the association measure in the note. Defaults to FALSE.
<code>correct</code>	Logical. If FALSE (the default), no continuity correction is applied. If TRUE, applies Yates correction (only for 2x2 tables).
<code>simulate_p</code>	Logical. If FALSE (the default), uses asymptotic p-values. If TRUE, uses Monte Carlo simulation.

simulate_B	Integer. Number of replicates for Monte Carlo simulation. Defaults to 2000.
digits	Number of decimals. Defaults to 1 for percentages, 0 for counts.
styled	Logical. If TRUE (the default), returns a <code>spicy_cross_table</code> object (for formatted printing). If FALSE, returns a plain <code>data.frame</code> .
show_n	Logical. If TRUE (the default), adds marginal N totals when percent != "none".

Value

A `data.frame`, list of `data.frames`, or `spicy_cross_table` object. When `by` is used, returns a `spicy_cross_table_list`.

Global Options

The function recognizes the following global options that modify its default behavior:

- `options(spicy.percent = "column")` Sets the default percentage mode for all calls to `cross_tab()`. Valid values are "none", "row", and "column". Equivalent to setting `percent = "column"` (or another choice) in each call.
- `options(spicy.simulate_p = TRUE)` Enables Monte Carlo simulation for all Chi-squared tests by default. Equivalent to setting `simulate_p = TRUE` in every call.
- `options(spicy.rescale = TRUE)` Automatically rescales weights so that total weighted N equals the raw N. Equivalent to setting `rescale = TRUE` in each call.

These options are convenient for users who wish to enforce consistent behavior across multiple calls to `cross_tab()` and other `spicy` table functions. They can be disabled or reset by setting them to `NULL`: `options(spicy.percent = NULL, spicy.simulate_p = NULL, spicy.rescale = NULL)`.

Example:

```
options(spicy.simulate_p = TRUE, spicy.rescale = TRUE)
cross_tab(sochealth, smoking, education, weights = weight)
```

Examples

```
# Basic crosstab
cross_tab(sochealth, smoking, education)

# Column percentages
cross_tab(sochealth, smoking, education, percent = "column")

# Weighted (rescaled)
cross_tab(sochealth, smoking, education, weights = weight, rescale = TRUE)

# Grouped by sex
cross_tab(sochealth, smoking, education, by = sex)

# Grouped by combination of variables
cross_tab(sochealth, smoking, education, by = interaction(sex, age_group))

# Ordinal variables: auto-selects Kendall's Tau-b
```

```

cross_tab(sochealth, education, self_rated_health)

# 2x2 table with Yates correction
cross_tab(sochealth, smoking, physical_activity, correct = TRUE)

```

freq

Frequency Table

Description

Creates a frequency table for a vector or variable from a data frame, with options for weighting, sorting, handling *labelled* data, defining custom missing values, and displaying cumulative percentages.

When `styled = TRUE`, the function prints a spicy-formatted ASCII table using `print.spicy_freq_table()` and `spicy_print_table()`; otherwise, it returns a `data.frame` containing frequencies and proportions.

Usage

```

freq(
  data,
  x = NULL,
  weights = NULL,
  digits = 1,
  valid = TRUE,
  cum = FALSE,
  sort = "",
  na_val = NULL,
  labelled_levels = c("prefixed", "labels", "values", "p", "l", "v"),
  rescale = TRUE,
  styled = TRUE,
  ...
)

```

Arguments

<code>data</code>	A <code>data.frame</code> , vector, or factor. If a data frame is provided, specify the target variable <code>x</code> .
<code>x</code>	A variable from <code>data</code> (unquoted).
<code>weights</code>	Optional numeric vector of weights (same length as <code>x</code>). The variable may be referenced as a bare name when it belongs to <code>data</code> .
<code>digits</code>	Number of decimal digits to display for percentages (default: 1).
<code>valid</code>	Logical. If <code>TRUE</code> (default), display valid percentages (excluding missing values).
<code>cum</code>	Logical. If <code>FALSE</code> (the default), cumulative percentages are omitted. If <code>TRUE</code> , adds cumulative percentages.

sort	<p>Sorting method for values:</p> <ul style="list-style-type: none"> • "" - no sorting (default) • "+" - increasing frequency • "-" - decreasing frequency • "name+" - alphabetical A-Z • "name-" - alphabetical Z-A
na_val	<p>Vector of numeric or character values to be treated as missing (NA). For <i>labelled</i> variables (from haven or labelled), this argument must refer to the underlying coded values, not the visible labels. Example:</p> <pre>x <- labelled(c(1, 2, 3, 1, 2, 3), c("Low" = 1, "Medium" = 2, "High" = 3)) freq(x, na_val = 1) # Treat all "Low" as missing</pre>
labelled_levels	<p>For labelled variables, defines how labels and values are displayed:</p> <ul style="list-style-type: none"> • "prefixed" or "p" - show labels as [value] label (default) • "labels" or "l" - show only labels • "values" or "v" - show only numeric codes
rescale	<p>Logical. If TRUE (default), rescale weights so that their total equals the unweighted sample size.</p>
styled	<p>Logical. If TRUE (default), print the formatted spicky table. If FALSE, return a plain data.frame with frequency values.</p>
...	<p>Additional arguments passed to <code>print.spicky_freq_table()</code>.</p>

Details

This function is designed to mimic common frequency procedures from statistical software such as SPSS or Stata, while integrating the flexibility of R's data structures.

It automatically detects the type of input (vector, factor, or labelled) and applies appropriate transformations, including:

- Handling of labelled variables via **labelled** or **haven**
- Optional recoding of specific values as missing (`na_val`)
- Optional weighting with a rescaling mechanism
- Support for cumulative percentages (`cum = TRUE`)
- Multiple display modes for labels via `labelled_levels`

When weighting is applied (`weights`), the frequencies and percentages are computed proportionally to the weights. The argument `rescale = TRUE` normalizes weights so their sum equals the unweighted sample size.

Value

A data.frame with columns:

- value - unique values or factor levels
- n - frequency count (weighted if applicable)
- prop - proportion of total
- valid_prop - proportion of valid responses (if valid = TRUE)
- cum_prop, cum_valid_prop - cumulative percentages (if cum = TRUE)

If styled = TRUE, prints the formatted table to the console and returns it invisibly.

See Also

[print.spicy_freq_table\(\)](#) for formatted printing. [spicy_print_table\(\)](#) for the underlying ASCII rendering engine.

Examples

```
# Frequency table with labelled ordered factor
freq(sochealth, education)
freq(sochealth, self_rated_health, sort = "-")

library(labelled)

# Simple numeric vector
x <- c(1, 2, 2, 3, 3, 3, NA)
freq(x)

# Labelled variable (haven-style)
x_lbl <- labelled(
  c(1, 2, 3, 1, 2, 3, 1, 2, NA),
  labels = c("Low" = 1, "Medium" = 2, "High" = 3)
)
var_label(x_lbl) <- "Satisfaction level"

# Treat value 1 ("Low") as missing
freq(x_lbl, na_val = 1)

# Display only labels, add cumulative %
freq(x_lbl, labelled_levels = "labels", cum = TRUE)

# Display values only, sorted descending
freq(x_lbl, labelled_levels = "values", sort = "-")

# With weighting
df <- data.frame(
  sex = factor(c("Male", "Female", "Female", "Male", NA, "Female")),
  weight = c(12, 8, 10, 15, 7, 9)
)

# Weighted frequencies (normalized)
```

```

freq(df, sex, weights = weight, rescale = TRUE)

# Weighted frequencies (without rescaling)
freq(df, sex, weights = weight, rescale = FALSE)

# Base R style, with weights and cumulative percentages
freq(df$sex, weights = df$weight, cum = TRUE)

# Piped version (tidy syntax) and sort alphabetically descending ("name-")
df |> freq(sex, sort = "name-")

# Non-styled return (for programmatic use)
f <- freq(df, sex, styled = FALSE)
head(f)

```

gamma_gk

Goodman-Kruskal Gamma

Description

gamma_gk() computes the Goodman-Kruskal Gamma statistic for a two-way contingency table of ordinal variables.

Usage

```

gamma_gk(
  x,
  detail = FALSE,
  conf_level = 0.95,
  digits = 3L,
  .include_se = FALSE
)

```

Arguments

x	A contingency table (of class table).
detail	Logical. If FALSE (default), return the estimate as a numeric scalar. If TRUE, return a named numeric vector including confidence interval and p-value.
conf_level	A number between 0 and 1 giving the confidence level (default 0.95). Only used when detail = TRUE. Set to NULL to omit the confidence interval.
digits	Number of decimal places used when printing the result (default 3). Only affects the detail = TRUE output.
.include_se	Internal parameter; do not use.

Details

Gamma is computed as $\gamma = (C - D)/(C + D)$, where C and D are the numbers of concordant and discordant pairs. It ignores tied pairs, making it appropriate for ordinal variables with many ties. Standard error formulas follow the DescTools implementations (Signorell et al., 2024); see [cramer_v\(\)](#) for full references.

Value

Same structure as [cramer_v\(\)](#): a scalar when `detail = FALSE`, a named vector when `detail = TRUE`. The p-value tests $H_0: \text{gamma} = 0$ (Wald z-test).

See Also

[kendall_tau_b\(\)](#), [kendall_tau_c\(\)](#), [somers_d\(\)](#), [assoc_measures\(\)](#)

Examples

```
tab <- table(sochealth$education, sochealth$self_rated_health)
gamma_gk(tab)
gamma_gk(tab, detail = TRUE)
```

goodman_kruskal_tau *Goodman-Kruskal's Tau*

Description

`goodman_kruskal_tau()` computes Goodman-Kruskal's Tau, a proportional reduction in error (PRE) measure for nominal variables.

Usage

```
goodman_kruskal_tau(
  x,
  direction = c("row", "column"),
  detail = FALSE,
  conf_level = 0.95,
  digits = 3L,
  .include_se = FALSE
)
```

Arguments

<code>x</code>	A contingency table (of class <code>table</code>).
<code>direction</code>	Direction of prediction: "row" (default, column predicts row) or "column" (row predicts column).

detail	Logical. If FALSE (default), return the estimate as a numeric scalar. If TRUE, return a named numeric vector including confidence interval and p-value.
conf_level	A number between 0 and 1 giving the confidence level (default 0.95). Only used when detail = TRUE. Set to NULL to omit the confidence interval.
digits	Number of decimal places used when printing the result (default 3). Only affects the detail = TRUE output.
.include_se	Internal parameter; do not use.

Details

Unlike [lambda_gk\(\)](#), Goodman-Kruskal's Tau uses all cell frequencies rather than only the modal categories, making it more sensitive to association patterns where lambda may be zero. Standard error formulas follow the DescTools implementations (Signorell et al., 2024); see [cramer_v\(\)](#) for full references.

Value

Same structure as [cramer_v\(\)](#): a scalar when detail = FALSE, a named vector when detail = TRUE. The p-value tests $H_0: \tau = 0$ (Wald z-test).

See Also

[lambda_gk\(\)](#), [uncertainty_coef\(\)](#), [assoc_measures\(\)](#)

Examples

```
tab <- table(sochealth$smoking, sochealth$education)
goodman_kruskal_tau(tab)
goodman_kruskal_tau(tab, direction = "column", detail = TRUE)
```

kendall_tau_b	<i>Kendall's Tau-b</i>
---------------	------------------------

Description

`kendall_tau_b()` computes Kendall's Tau-b for a two-way contingency table of ordinal variables.

Usage

```
kendall_tau_b(
  x,
  detail = FALSE,
  conf_level = 0.95,
  digits = 3L,
  .include_se = FALSE
)
```

Arguments

<code>x</code>	A contingency table (of class <code>table</code>).
<code>detail</code>	Logical. If <code>FALSE</code> (default), return the estimate as a numeric scalar. If <code>TRUE</code> , return a named numeric vector including confidence interval and p-value.
<code>conf_level</code>	A number between 0 and 1 giving the confidence level (default 0.95). Only used when <code>detail = TRUE</code> . Set to <code>NULL</code> to omit the confidence interval.
<code>digits</code>	Number of decimal places used when printing the result (default 3). Only affects the <code>detail = TRUE</code> output.
<code>.include_se</code>	Internal parameter; do not use.

Details

Kendall's Tau-b is computed as $\tau_b = (C - D) / \sqrt{(n_0 - n_1)(n_0 - n_2)}$, where $n_0 = n(n-1)/2$, n_1 is the number of pairs tied on the row variable, and n_2 is the number tied on the column variable. Tau-b corrects for ties and is appropriate for square tables. Standard error formulas follow the DescTools implementations (Signorell et al., 2024); see [cramer_v\(\)](#) for full references.

Value

Same structure as [cramer_v\(\)](#): a scalar when `detail = FALSE`, a named vector when `detail = TRUE`. The p-value tests $H_0: \tau\text{-}b = 0$ (Wald z-test).

See Also

[kendall_tau_c\(\)](#), [gamma_gk\(\)](#), [somers_d\(\)](#), [assoc_measures\(\)](#)

Examples

```
tab <- table(sochealth$education, sochealth$self_rated_health)
kendall_tau_b(tab)
```

<code>kendall_tau_c</code>	<i>Kendall's Tau-c (Stuart's Tau-c)</i>
----------------------------	-----------------------------------------

Description

`kendall_tau_c()` computes Stuart's Tau-c (also known as Kendall's Tau-c) for a two-way contingency table of ordinal variables.

Usage

```
kendall_tau_c(
  x,
  detail = FALSE,
  conf_level = 0.95,
  digits = 3L,
  .include_se = FALSE
)
```

Arguments

<code>x</code>	A contingency table (of class <code>table</code>).
<code>detail</code>	Logical. If <code>FALSE</code> (default), return the estimate as a numeric scalar. If <code>TRUE</code> , return a named numeric vector including confidence interval and p-value.
<code>conf_level</code>	A number between 0 and 1 giving the confidence level (default 0.95). Only used when <code>detail = TRUE</code> . Set to <code>NULL</code> to omit the confidence interval.
<code>digits</code>	Number of decimal places used when printing the result (default 3). Only affects the <code>detail = TRUE</code> output.
<code>.include_se</code>	Internal parameter; do not use.

Details

Stuart's Tau-c is computed as $\tau_c = 2m(C-D)/(n^2(m-1))$, where $m = \min(r, c)$. It is appropriate for rectangular tables and is not restricted to the range $[-1, 1]$ only for square tables. Standard error formulas follow the DescTools implementations (Signorell et al., 2024); see [cramer_v\(\)](#) for full references.

Value

Same structure as [cramer_v\(\)](#): a scalar when `detail = FALSE`, a named vector when `detail = TRUE`. The p-value tests $H_0: \tau_c = 0$ (Wald z-test).

See Also

[kendall_tau_b\(\)](#), [gamma_gk\(\)](#), [somers_d\(\)](#), [assoc_measures\(\)](#)

Examples

```
tab <- table(sochealth$education, sochealth$self_rated_health)
kendall_tau_c(tab)
```

label_from_names	Derive variable labels from column names name<sep>label
------------------	---------------------------------------------------------

Description

Splits each column name at the **first** occurrence of `sep`, renames the column to the part before `sep` (the *name*), and assigns the part after `sep` as a `labelled::var_label()`. This works even if the label itself contains the separator.

Usage

```
label_from_names(df, sep = ". ")
```

Arguments

<code>df</code>	A <code>data.frame</code> or tibble with column names of the form "name<sep>label" (e.g. "name.label"). (by default from LimeSurvey).
<code>sep</code>	Character string used as separator between name and label. Default is ". " (LimeSurvey's default), but any literal string can be used.

Details

This function is especially useful for **LimeSurvey CSV exports** when using *Export results → Export format: CSV → Headings: Question code & question text*, where column names look like "code. question text". In this case the default separator is ". ".

Value

A base tibble with column names equal to the *names* (before `sep`) and `var_label` attributes equal to the *labels* (after `sep`).

Examples

```
# Example with LimeSurvey-style column names
df <- data.frame(
  "age. Age of respondent" = c(25, 30),
  "score. Total score. Manually computed." = c(12, 14),
  check.names = FALSE
)

# sep = ". " by default (LimeSurvey)
out <- label_from_names(df)
labelled::var_label(out)

# Example with a custom separator ("|")
df2 <- data.frame(
  "id|Identifier" = 1:3,
  "score|Total score" = c(10, 20, 30),
)
```

```

    check.names = FALSE
  )
  out2 <- label_from_names(df2, sep = "|")
  labelled::var_label(out2)

```

lambda_gk

Goodman-Kruskal's Lambda

Description

lambda_gk() computes Goodman-Kruskal's Lambda, a proportional reduction in error (PRE) measure for nominal variables.

Usage

```

lambda_gk(
  x,
  direction = c("symmetric", "row", "column"),
  detail = FALSE,
  conf_level = 0.95,
  digits = 3L,
  .include_se = FALSE
)

```

Arguments

x	A contingency table (of class table).
direction	Direction of prediction: "symmetric" (default), "row" (column predicts row), or "column" (row predicts column).
detail	Logical. If FALSE (default), return the estimate as a numeric scalar. If TRUE, return a named numeric vector including confidence interval and p-value.
conf_level	A number between 0 and 1 giving the confidence level (default 0.95). Only used when detail = TRUE. Set to NULL to omit the confidence interval.
digits	Number of decimal places used when printing the result (default 3). Only affects the detail = TRUE output.
.include_se	Internal parameter; do not use.

Details

Lambda measures how much prediction error is reduced when the independent variable is used to predict the dependent variable. It ranges from 0 (no reduction) to 1 (perfect prediction). Lambda can equal zero even when variables are associated if the modal category dominates in every column (or row). Standard error formulas follow the DescTools implementations (Signorell et al., 2024); see [cramer_v\(\)](#) for full references.

Value

Same structure as `cramer_v()`: a scalar when `detail = FALSE`, a named vector when `detail = TRUE`. The p-value tests $H_0: \lambda = 0$ (Wald z-test).

See Also

`goodman_kruskal_tau()`, `uncertainty_coef()`, `assoc_measures()`

Examples

```
tab <- table(sochealth$smoking, sochealth$education)
lambda_gk(tab)
lambda_gk(tab, direction = "row")
lambda_gk(tab, direction = "column", detail = TRUE)
```

mean_n

Row Means with Optional Minimum Valid Values

Description

`mean_n()` computes row means from a `data.frame` or `matrix`, handling missing values (NAs) automatically. Row-wise means are calculated across selected numeric columns, with an optional condition on the minimum number (or proportion) of valid (non-missing) values required for a row to be included. Non-numeric columns are excluded automatically and reported.

Usage

```
mean_n(
  data = NULL,
  select = dplyr::everything(),
  exclude = NULL,
  min_valid = NULL,
  digits = NULL,
  regex = FALSE,
  verbose = FALSE
)
```

Arguments

<code>data</code>	A <code>data.frame</code> or <code>matrix</code> .
<code>select</code>	Columns to include. If <code>regex = FALSE</code> , use <code>tidyselect</code> syntax (default: <code>dplyr::everything()</code>). If <code>regex = TRUE</code> , provide a regular expression pattern (character string).
<code>exclude</code>	Columns to exclude (default: <code>NULL</code>).
<code>min_valid</code>	Minimum number of valid (non-NA) values required per row. If a proportion, it's applied to the number of selected columns. Defaults to <code>NULL</code> (all values must be valid).

digits	Optional number of decimal places to round the result. Defaults to NULL (no rounding).
regex	Logical. If FALSE (the default), uses tidysselect helpers. If TRUE, the select argument is treated as a regular expression.
verbose	Logical. If FALSE (the default), messages are suppressed. If TRUE, prints a message about non-numeric columns excluded.

Value

A numeric vector of row-wise means.

Examples

```
library(dplyr)

# Create a simple numeric data frame
df <- tibble(
  var1 = c(10, NA, 30, 40, 50),
  var2 = c(5, NA, 15, NA, 25),
  var3 = c(NA, 30, 20, 50, 10)
)

# Compute row-wise mean (all values must be valid by default)
mean_n(df)

# Require at least 2 valid (non-NA) values per row
mean_n(df, min_valid = 2)

# Require at least 50% valid (non-NA) values per row
mean_n(df, min_valid = 0.5)

# Round the result to 1 decimal
mean_n(df, digits = 1)

# Select specific columns
mean_n(df, select = c(var1, var2))

# Select specific columns using a pipe
df |>
  select(var1, var2) |>
  mean_n()

# Exclude a column
mean_n(df, exclude = "var3")

# Select columns ending with "1"
mean_n(df, select = ends_with("1"))

# Use with native pipe
df |> mean_n(select = starts_with("var"))

# Use inside dplyr::mutate()
```

```

df |> mutate(mean_score = mean_n(min_valid = 2))

# Select columns directly inside mutate()
df |> mutate(mean_score = mean_n(select = c(var1, var2), min_valid = 1))

# Select columns before mutate
df |>
  select(var1, var2) |>
  mutate(mean_score = mean_n(min_valid = 1))

# Show verbose processing info
df |> mutate(mean_score = mean_n(min_valid = 2, digits = 1, verbose = TRUE))

# Add character and grouping columns
df_mixed <- mutate(df,
  name = letters[1:5],
  group = c("A", "A", "B", "B", "A")
)
df_mixed

# Non-numeric columns are ignored
mean_n(df_mixed)

# Use within mutate() on mixed data
df_mixed |> mutate(mean_score = mean_n(select = starts_with("var")))

# Use everything() but exclude non-numeric columns manually
mean_n(df_mixed, select = everything(), exclude = "group")

# Select columns using regex
mean_n(df_mixed, select = "^var", regex = TRUE)
mean_n(df_mixed, select = "ar", regex = TRUE)

# Apply to a subset of rows (first 3)
df_mixed[1:3, ] |> mean_n(select = starts_with("var"))

# Store the result in a new column
df_mixed$mean_score <- mean_n(df_mixed, select = starts_with("var"))
df_mixed

# With a numeric matrix
mat <- matrix(c(1, 2, NA, 4, 5, NA, 7, 8, 9), nrow = 3, byrow = TRUE)
mat
mat |> mean_n(min_valid = 2)

```

 phi

Phi coefficient

Description

phi() computes the phi coefficient for a 2x2 contingency table.

Usage

```
phi(x, detail = FALSE, conf_level = 0.95, digits = 3L, .include_se = FALSE)
```

Arguments

<code>x</code>	A contingency table (of class <code>table</code>).
<code>detail</code>	Logical. If <code>FALSE</code> (default), return the estimate as a numeric scalar. If <code>TRUE</code> , return a named numeric vector including confidence interval and p-value.
<code>conf_level</code>	A number between 0 and 1 giving the confidence level (default 0.95). Only used when <code>detail = TRUE</code> . Set to <code>NULL</code> to omit the confidence interval.
<code>digits</code>	Number of decimal places used when printing the result (default 3). Only affects the <code>detail = TRUE</code> output.
<code>.include_se</code>	Internal parameter; do not use.

Details

The phi coefficient is $\phi = \sqrt{\chi^2/n}$. It is equivalent to Cramer's V for 2x2 tables and equals the Pearson correlation between the two binary variables. The confidence interval uses the Fisher z-transformation. Standard error formulas follow the DescTools implementations (Signorell et al., 2024); see [cramer_v\(\)](#) for full references.

Value

Same structure as [cramer_v\(\)](#): a scalar when `detail = FALSE`, a named vector when `detail = TRUE`. The p-value tests the null hypothesis of no association (Pearson chi-squared test).

See Also

[cramer_v\(\)](#), [yule_q\(\)](#), [assoc_measures\(\)](#)

Examples

```
tab <- table(sochealth$smoking, sochealth$sex)
phi(tab)
phi(tab, detail = TRUE)
```

sochealth

Simulated social-health survey

Description

A simulated dataset of 1200 respondents from a fictional social-health survey, designed to illustrate the main features of the `spicy` package: variable labels, ordered factors, survey weights, association measures, and APA-style reporting.

Usage

sochealth

Format

A tibble with 1200 rows and 24 variables:

sex Factor. Sex of the respondent.

age Numeric. Age in years (25–75).

age_group Ordered factor. Age group (25–34, 35–49, 50–64, 65–75).

education Ordered factor. Highest education level (Lower secondary, Upper secondary, Tertiary).

social_class Ordered factor. Subjective social class (Lower, Working, Lower middle, Middle, Upper middle).

region Factor. Region of residence (6 regions).

employment_status Factor. Employment status (Employed, Student, Unemployed, Inactive).

income_group Ordered factor. Household income group (Low, Lower middle, Upper middle, High). Contains missing values.

income Numeric. Monthly household income in CHF.

smoking Factor. Current smoker (No, Yes). Contains missing values.

physical_activity Factor. Regular physical activity (No, Yes).

dentist_12m Factor. Dentist visit in the last 12 months (No, Yes).

self_rated_health Ordered factor. Self-rated health (Poor, Fair, Good, Very good). Contains missing values.

wellbeing_score Numeric. WHO-5 wellbeing index (0–100).

bmi Numeric. Body mass index. Contains missing values.

bmi_category Ordered factor. BMI category (Normal weight, Overweight, Obesity). Contains missing values.

institutional_trust Ordered factor. Trust in institutions (Very low, Low, High, Very high).

political_position Numeric. Political position on a 0 (left) to 10 (right) scale. Contains missing values.

life_sat_health Integer. Satisfaction with own health (1–5 Likert scale). Contains missing values.

life_sat_work Integer. Satisfaction with work or main activity (1–5 Likert scale). Contains missing values.

life_sat_relationships Integer. Satisfaction with personal relationships (1–5 Likert scale). Contains missing values.

life_sat_standard Integer. Satisfaction with standard of living (1–5 Likert scale). Contains missing values.

response_date POSIXct. Date and time of survey response (September–November 2024).

weight Numeric. Survey design weight.

Details

All variables carry labels (accessible via `labelled::var_label()` and displayed by `varlist()`). Several ordered factors are included so that `cross_tab()` can demonstrate automatic ordinal measure selection.

Source

Simulated data for illustration purposes.

Examples

```
data(sochealth)
varlist(sochealth)
freq(sochealth, education)
cross_tab(sochealth, education, self_rated_health)
```

somers_d	<i>Somers' D</i>
----------	------------------

Description

`somers_d()` computes Somers' D for a two-way contingency table of ordinal variables.

Usage

```
somers_d(
  x,
  direction = c("row", "column", "symmetric"),
  detail = FALSE,
  conf_level = 0.95,
  digits = 3L,
  .include_se = FALSE
)
```

Arguments

<code>x</code>	A contingency table (of class <code>table</code>).
<code>direction</code>	Direction of prediction: "row" (default, column predicts row), "column" (row predicts column), or "symmetric" (average of both directions).
<code>detail</code>	Logical. If FALSE (default), return the estimate as a numeric scalar. If TRUE, return a named numeric vector including confidence interval and p-value.
<code>conf_level</code>	A number between 0 and 1 giving the confidence level (default 0.95). Only used when <code>detail = TRUE</code> . Set to NULL to omit the confidence interval.
<code>digits</code>	Number of decimal places used when printing the result (default 3). Only affects the <code>detail = TRUE</code> output.
<code>.include_se</code>	Internal parameter; do not use.

Details

Somers' D is an asymmetric ordinal measure defined as $d = (C - D)/(C + D + T)$, where T is the number of pairs tied on the independent variable. The symmetric version is the harmonic mean of the two asymmetric values. Standard error formulas follow the DescTools implementations (Signorell et al., 2024); see [cramer_v\(\)](#) for full references.

Value

Same structure as [cramer_v\(\)](#): a scalar when `detail = FALSE`, a named vector when `detail = TRUE`. The p-value tests $H_0: D = 0$ (Wald z-test).

See Also

[kendall_tau_b\(\)](#), [gamma_gk\(\)](#), [assoc_measures\(\)](#)

Examples

```
tab <- table(sochealth$education, sochealth$self_rated_health)
somers_d(tab, direction = "row")
somers_d(tab, direction = "column", detail = TRUE)
```

spicy_print_table

Print a spicy-formatted ASCII table

Description

User-facing helper that prints a visually aligned, spicy-styled ASCII table created by functions such as [freq\(\)](#) or [cross_tab\(\)](#). It automatically adjusts column alignment, spacing, and separators for improved readability in console outputs.

This function wraps the internal renderer [build_ascii_table\(\)](#), adding optional titles, notes, and automatic alignment rules depending on the type of table.

Usage

```
spicy_print_table(
  x,
  title = attr(x, "title"),
  note = attr(x, "note"),
  padding = c("compact", "normal", "wide"),
  first_column_line = TRUE,
  row_total_line = TRUE,
  column_total_line = TRUE,
  bottom_line = FALSE,
  lines_color = "darkgrey",
  align_left_cols = NULL,
  align_center_cols = integer(0),
```

```

    group_sep_rows = integer(0),
    ...
)

```

Arguments

x	A <code>spicy_table</code> or <code>data.frame</code> to be printed.
title	Optional title displayed above the table. Defaults to the "title" attribute of x if present.
note	Optional note displayed below the table. Defaults to the "note" attribute of x if present.
padding	Character string controlling horizontal spacing between columns: <ul style="list-style-type: none"> "compact" - minimal spacing "normal" - moderate spacing (default) "wide" - extra spacing (for wide displays)
first_column_line	Logical. If TRUE (the default), adds a vertical separator after the first column.
row_total_line, column_total_line, bottom_line	Logical flags controlling the presence of horizontal lines before total rows/columns or at the bottom of the table. Both <code>row_total_line</code> and <code>column_total_line</code> default to TRUE; <code>bottom_line</code> defaults to FALSE.
lines_color	Character. Color for table separators. Defaults to "darkgrey". Only applied if the output supports ANSI colors (see <code>crayon::has_color()</code>).
align_left_cols	Integer vector of column indices to left-align. If NULL (the default), alignment is auto-detected based on x: <ul style="list-style-type: none"> For freq tables -> c(1, 2) For cross tables -> 1
align_center_cols	Integer vector of column indices to center-align. Defaults to <code>integer(0)</code> .
group_sep_rows	Integer vector of row indices before which a light dashed separator line is drawn. Defaults to <code>integer(0)</code> .
...	Additional arguments passed to <code>build_ascii_table()</code> .

Details

`spicy_print_table()` detects whether the table represents frequencies (freq-style) or cross-tabulations (cross-style) and adjusts formatting accordingly:

- For **frequency tables**, the first two columns (*Category* and *Values*) are left-aligned.
- For **cross tables**, only the first column (row variable) is left-aligned.

The function supports Unicode line-drawing characters and colored separators using the **crayon** package, with graceful fallback to monochrome output when color is not supported. If the table exceeds the console width, it is split into stacked horizontal panels while repeating the left-most identifier columns.

Value

Invisibly returns `x`, after printing the formatted ASCII table to the console.

See Also

[build_ascii_table\(\)](#) for the underlying text rendering engine. [print.spicy_freq_table\(\)](#) for the specialized printing method used by [freq\(\)](#).

Examples

```
# Simple demonstration
df <- data.frame(
  Category = c("Valid", "", "Missing", "Total"),
  Values = c("Yes", "No", "NA", ""),
  Freq. = c(12, 8, 1, 21),
  Percent = c(57.1, 38.1, 4.8, 100.0)
)

spicy_print_table(df,
  title = "Frequency table: Example",
  note = "Class: data.frame\nData: demo"
)
```

`sum_n`*Row Sums with Optional Minimum Valid Values*

Description

`sum_n()` computes row sums from a `data.frame` or `matrix`, handling missing values (NAs) automatically. Row-wise sums are calculated across selected numeric columns, with an optional condition on the minimum number (or proportion) of valid (non-missing) values required for a row to be included. Non-numeric columns are excluded automatically and reported.

Usage

```
sum_n(
  data = NULL,
  select = dplyr::everything(),
  exclude = NULL,
  min_valid = NULL,
  digits = NULL,
  regex = FALSE,
  verbose = FALSE
)
```

Arguments

data	A <code>data.frame</code> or <code>matrix</code> .
select	Columns to include. If <code>regex = FALSE</code> , use <code>tidyselect</code> syntax (default: <code>dplyr::everything()</code>). If <code>regex = TRUE</code> , provide a regular expression pattern (character string).
exclude	Columns to exclude (default: <code>NULL</code>).
min_valid	Minimum number of valid (non-NA) values required per row. If a proportion, it's applied to the number of selected columns. Defaults to <code>NULL</code> (all values must be valid).
digits	Optional number of decimal places to round the result. Defaults to <code>NULL</code> (no rounding).
regex	Logical. If <code>FALSE</code> (the default), uses <code>tidyselect</code> helpers. If <code>TRUE</code> , the <code>select</code> argument is treated as a regular expression.
verbose	Logical. If <code>FALSE</code> (the default), messages are suppressed. If <code>TRUE</code> , prints a message about non-numeric columns excluded.

Value

A numeric vector of row-wise sums

Examples

```
library(dplyr)

# Create a simple numeric data frame
df <- tibble(
  var1 = c(10, NA, 30, 40, 50),
  var2 = c(5, NA, 15, NA, 25),
  var3 = c(NA, 30, 20, 50, 10)
)

# Compute row-wise sums (all values must be valid by default)
sum_n(df)

# Require at least 2 valid (non-NA) values per row
sum_n(df, min_valid = 2)

# Require at least 50% valid (non-NA) values per row
sum_n(df, min_valid = 0.5)

# Round the results to 1 decimal
sum_n(df, digits = 1)

# Select specific columns
sum_n(df, select = c(var1, var2))

# Select specific columns using a pipe
df |>
  select(var1, var2) |>
  sum_n()
```

```
# Exclude a column
sum_n(df, exclude = "var3")

# Select columns ending with "1"
sum_n(df, select = ends_with("1"))

# Use with native pipe
df |> sum_n(select = starts_with("var"))

# Use inside dplyr::mutate()
df |> mutate(sum_score = sum_n(min_valid = 2))

# Select columns directly inside mutate()
df |> mutate(sum_score = sum_n(select = c(var1, var2), min_valid = 1))

# Select columns before mutate
df |>
  select(var1, var2) |>
  mutate(sum_score = sum_n(min_valid = 1))

# Show verbose message
df |> mutate(sum_score = sum_n(min_valid = 2, digits = 1, verbose = TRUE))

# Add character and grouping columns
df_mixed <- mutate(df,
  name = letters[1:5],
  group = c("A", "A", "B", "B", "A")
)
df_mixed

# Non-numeric columns are ignored
sum_n(df_mixed)

# Use inside mutate with mixed data
df_mixed |> mutate(sum_score = sum_n(select = starts_with("var")))

# Use everything(), but exclude known non-numeric
sum_n(df_mixed, select = everything(), exclude = "group")

# Select columns using regex
sum_n(df_mixed, select = "^var", regex = TRUE)
sum_n(df_mixed, select = "ar", regex = TRUE)

# Apply to a subset of rows
df_mixed[1:3, ] |> sum_n(select = starts_with("var"))

# Store the result in a new column
df_mixed$sum_score <- sum_n(df_mixed, select = starts_with("var"))
df_mixed

# With a numeric matrix
mat <- matrix(c(1, 2, NA, 4, 5, NA, 7, 8, 9), nrow = 3, byrow = TRUE)
```

```
mat
mat |> sum_n(min_valid = 2)
```

table_categorical	<i>Categorical summary table</i>
-------------------	----------------------------------

Description

Builds a publication-ready frequency or cross-tabulation table for one or many categorical variables selected with `tidyselect` syntax.

With `by`, produces grouped cross-tabulation summaries (using `cross_tab()` internally) with Chi-squared p -values and optional association measures. Without `by`, produces one-way frequency-style summaries.

Multiple output formats are available via `output`: a printed ASCII table ("default"), a wide or long numeric `data.frame` ("data.frame", "long"), or publication-ready tables ("tinytable", "gt", "flextable", "excel", "clipboard", "word").

Usage

```
table_categorical(
  data,
  select,
  by = NULL,
  labels = NULL,
  levels_keep = NULL,
  include_total = TRUE,
  drop_na = TRUE,
  weights = NULL,
  rescale = FALSE,
  correct = FALSE,
  simulate_p = FALSE,
  simulate_B = 2000,
  percent_digits = 1,
  p_digits = 3,
  v_digits = 2,
  assoc_measure = "auto",
  assoc_ci = FALSE,
  decimal_mark = ".",
  output = c("default", "data.frame", "long", "tinytable", "gt", "flextable", "excel",
    "clipboard", "word"),
  indent_text = " ",
  indent_text_excel_clipboard = strrep(" ", 6),
  add_multilevel_header = TRUE,
  blank_na_wide = FALSE,
  excel_path = NULL,
  excel_sheet = "Categorical",
```

```

  clipboard_delim = "\t",
  word_path = NULL
)

```

Arguments

data	A data frame.
select	Columns to include as row variables. Supports tidyselect syntax and character vectors of column names.
by	Optional grouping column used for columns/groups. Accepts an unquoted column name or a single character column name.
labels	An optional character vector of display labels for the variables named in select (must be the same length and in the same order). When NULL (the default), column names are used as-is.
levels_keep	Optional character vector of levels to keep/order for row modalities. If NULL, all observed levels are kept.
include_total	Logical. If TRUE (the default), includes a Total group when available.
drop_na	Logical. If TRUE (the default), removes rows with NA in the row/group variable before each cross-tabulation. If FALSE, missing values are displayed as a dedicated "(Missing)" level.
weights	Optional weights. Either NULL (the default), a numeric vector of length nrow(data), or a single column in data supplied as an unquoted name or a character string.
rescale	Logical. If FALSE (the default), weights are used as-is. If TRUE, rescales weights so total weighted N matches raw N. Passed to <code>spicy::cross_tab()</code> .
correct	Logical. If FALSE (the default), no continuity correction is applied. If TRUE, applies Yates correction in 2x2 chi-squared contexts. Passed to <code>spicy::cross_tab()</code> .
simulate_p	Logical. If FALSE (the default), uses asymptotic p-values. If TRUE, uses Monte Carlo simulation. Passed to <code>spicy::cross_tab()</code> .
simulate_B	Integer. Number of Monte Carlo replicates when <code>simulate_p = TRUE</code> . Defaults to 2000.
percent_digits	Number of digits for percentages in report outputs. Defaults to 1.
p_digits	Number of digits for p-values (except < .001). Defaults to 3.
v_digits	Number of digits for the association measure. Defaults to 2.
assoc_measure	Passed to <code>cross_tab()</code> . Which association measure to report ("auto", "cramer_v", "phi", "gamma", "tau_b", "tau_c", "somers_d", "lambda", "none"). Defaults to "auto".
assoc_ci	Passed to <code>cross_tab()</code> . If TRUE, includes the confidence interval of the association measure. In data formats ("data.frame", "long", "excel", "clipboard"), two extra columns CI_lower and CI_upper are added. In rendered formats ("gt", "tinytable", "flexible", "word"), the CI is shown inline (e.g., .14 [.08, .19]). Defaults to FALSE.
decimal_mark	Decimal separator (". " or ", "). Defaults to ". ".
output	Output format. One of:

- "default" (a printed ASCII table, returned invisibly)
- "data.frame" (a wide numeric data.frame)
- "long" (a long numeric data.frame)
- "tinytable" (requires tinytable)
- "gt" (requires gt)
- "flextable" (requires flextable)
- "excel" (requires openxlsx2)
- "clipboard" (requires clipr)
- "word" (requires flextable and officer)

indent_text	Prefix used for modality labels in report table building. Defaults to " " (two spaces).
indent_text_excel_clipboard	Stronger indentation used in Excel and clipboard exports. Defaults to six non-breaking spaces.
add_multilevel_header	Logical. If TRUE (the default), merges top headers in Excel export.
blank_na_wide	Logical. If FALSE (the default), NA values are kept as-is in wide raw output. If TRUE, replaces them with empty strings.
excel_path	Path for output = "excel". Defaults to NULL.
excel_sheet	Sheet name for Excel export. Defaults to "Categorical".
clipboard_delim	Delimiter for clipboard text export. Defaults to "\t".
word_path	Path for output = "word" or optional save path when output = "flextable". Defaults to NULL.

Details

When `by` is used, each selected variable is cross-tabulated against the grouping variable with `cross_tab()`. Chi-squared statistics, p -values, and the chosen association measure are reported for each variable.

Optional output engines require suggested packages:

- **tinytable** for output = "tinytable"
- **gt** for output = "gt"
- **flextable** for output = "flextable"
- **flextable + officer** for output = "word"
- **openxlsx2** for output = "excel"
- **clipr** for output = "clipboard"

Value

Depends on output:

- "default": prints a styled ASCII table and returns the underlying data.frame invisibly (S3 class "spicy_categorical_table").

- "data.frame": a wide data.frame with one row per variable–level combination. When by is used, the columns are Variable, Level, and one pair of n / \% columns per group level (plus Total when include_total = TRUE), followed by Chi2, df, p, and the association measure column. When by = NULL, the columns are Variable, Level, n, \%.
- "long": a long data.frame with columns variable, level, group, n, percent (and chi2, df, p, association measure columns when by is used).
- "tinytable": a tinytable object.
- "gt": a gt_tbl object.
- "flextable": a flextable object.
- "excel" / "word": writes to disk and returns the file path invisibly.
- "clipboard": copies the table and returns the display data.frame invisibly.

See Also

[table_continuous\(\)](#) for continuous variables; [cross_tab\(\)](#) for two-way cross-tabulations; [freq\(\)](#) for one-way frequency tables.

Examples

```
# Long numeric output
table_categorical(
  data = sochealth,
  select = c(smoking, physical_activity),
  by = education,
  labels = c("Current smoker", "Physical activity"),
  output = "long"
)

# ASCII console output (default)
table_categorical(
  data = sochealth,
  select = c(smoking, physical_activity),
  by = sex
)

# One-way frequency-style table
table_categorical(
  data = sochealth,
  select = c(smoking, physical_activity)
)

# Wide numeric data.frame
table_categorical(
  data = sochealth,
  select = c(smoking, physical_activity),
  by = education,
  labels = c("Current smoker", "Physical activity"),
  output = "data.frame"
)
```

```

# Weighted example
table_categorical(
  data = sochealth,
  select = c(smoking, physical_activity),
  by = education,
  labels = c("Current smoker", "Physical activity"),
  weights = "weight",
  rescale = TRUE,
  simulate_p = FALSE,
  output = "long"
)

# Optional output: tinytable
if (requireNamespace("tinytable", quietly = TRUE)) {
  table_categorical(
    data = sochealth,
    select = c(smoking, physical_activity),
    by = sex,
    labels = c("Current smoker", "Physical activity"),
    output = "tinytable"
  )
}

# Optional output: Excel
if (requireNamespace("openxlsx2", quietly = TRUE)) {
  table_categorical(
    data = sochealth,
    select = c(smoking, physical_activity),
    by = education,
    labels = c("Current smoker", "Physical activity"),
    output = "excel",
    excel_path = tempfile(fileext = ".xlsx")
  )
}

```

table_continuous

Continuous summary table

Description

Computes descriptive statistics (mean, SD, min, max, confidence interval of the mean, n) for one or many continuous variables selected with `tidyselect` syntax.

With `by`, produces grouped summaries with optional group-comparison tests (`test`), p -values (`p_value`), test statistics (`statistic`), and effect sizes (`effect_size / effect_size_ci`). Without `by`, produces one-way descriptive summaries.

Multiple output formats are available via `output`: a printed ASCII table ("default"), a plain numeric `data.frame` ("data.frame"), or publication-ready tables ("tinytable", "gt", "flextable", "excel", "clipboard", "word").

Usage

```
table_continuous(
  data,
  select = dplyr::everything(),
  by = NULL,
  exclude = NULL,
  regex = FALSE,
  test = c("welch", "student", "nonparametric"),
  p_value = FALSE,
  statistic = FALSE,
  effect_size = FALSE,
  effect_size_ci = FALSE,
  labels = NULL,
  ci_level = 0.95,
  digits = 2,
  effect_size_digits = 2,
  decimal_mark = ".",
  output = c("default", "data.frame", "tinytable", "gt", "flextable", "excel",
    "clipboard", "word"),
  excel_path = NULL,
  excel_sheet = "Descriptives",
  clipboard_delim = "\t",
  word_path = NULL,
  verbose = FALSE
)
```

Arguments

<code>data</code>	A <code>data.frame</code> .
<code>select</code>	Columns to include. If <code>regex = FALSE</code> , use <code>tidyselect</code> syntax or a character vector of column names (default: <code>dplyr::everything()</code>). If <code>regex = TRUE</code> , provide a regular expression pattern (character string).
<code>by</code>	Optional grouping column. Accepts an unquoted column name or a single character column name. The column does not need to be numeric.
<code>exclude</code>	Columns to exclude. Supports <code>tidyselect</code> syntax and character vectors of column names.
<code>regex</code>	Logical. If <code>FALSE</code> (the default), uses <code>tidyselect</code> helpers. If <code>TRUE</code> , the <code>select</code> argument is treated as a regular expression.
<code>test</code>	Character. Statistical test to use when comparing groups. One of "welch" (default), "student", or "nonparametric". <ul style="list-style-type: none"> "welch": Welch <i>t</i>-test (2 groups) or Welch one-way ANOVA (3+ groups). Does not assume equal variances.

	<ul style="list-style-type: none"> • "student": Student <i>t</i>-test (2 groups) or classic one-way ANOVA (3+ groups). Assumes equal variances. • "nonparametric": Wilcoxon rank-sum / Mann–Whitney <i>U</i> (2 groups) or Kruskal–Wallis <i>H</i> (3+ groups).
	Used when by is supplied together with <code>p_value = TRUE</code> , <code>statistic = TRUE</code> , or <code>effect_size = TRUE</code> . Ignored otherwise.
<code>p_value</code>	Logical. If TRUE and by is used, adds a <i>p</i> -value column from the test specified by test. Defaults to FALSE. Ignored when by is not used.
<code>statistic</code>	Logical. If TRUE and by is used, the test statistic is shown in an additional column (e.g., $t(df) = \dots$, $F(df1, df2) = \dots$, $W = \dots$, or $H(df) = \dots$). Both <code>p_value</code> and <code>statistic</code> are independent; either or both can be enabled. Defaults to FALSE. Ignored when by is not used.
<code>effect_size</code>	Logical. If TRUE and by is used, adds an effect-size column ("ES"). The measure is chosen automatically: <ul style="list-style-type: none"> • Hedges' <i>g</i> (bias-corrected) - 2 groups, parametric (CI via Hedges & Olkin approximation). • Eta-squared (η^2) - 3+ groups, parametric (CI via noncentral <i>F</i> distribution). • Rank-biserial <i>r</i> (r_{rb}) - 2 groups, nonparametric (CI via Fisher <i>z</i>-transform). • Epsilon-squared (ε^2) - 3+ groups, nonparametric (CI via percentile bootstrap, 2 000 replicates). Defaults to FALSE. Ignored when by is not used.
<code>effect_size_ci</code>	Logical. If TRUE, appends the confidence interval of the effect size in brackets (e.g., $g = 0.45 [0.22, 0.68]$). Implies <code>effect_size = TRUE</code> . Defaults to FALSE.
<code>labels</code>	An optional named character vector of variable labels. Names must match column names in data. When NULL (the default), labels are auto-detected from variable attributes (e.g., haven labels); if none are found, the column name is used.
<code>ci_level</code>	Confidence level for the mean confidence interval (default: 0.95). Must be between 0 and 1 exclusive.
<code>digits</code>	Number of decimal places for descriptive values and test statistics (default: 2).
<code>effect_size_digits</code>	Number of decimal places for effect-size values in formatted displays (default: 2).
<code>decimal_mark</code>	Character used as decimal separator. Either "." (default) or ",", ".,".
<code>output</code>	Output format. One of: <ul style="list-style-type: none"> • "default" (a printed ASCII table, returned invisibly) • "data.frame" (a plain numeric data.frame) • "tinytable" (requires tinytable) • "gt" (requires gt) • "flextable" (requires flextable) • "excel" (requires openxlsx2) • "clipboard" (requires clipr)

	<ul style="list-style-type: none"> • "word" (requires flextable and officer)
excel_path	File path for output = "excel".
excel_sheet	Sheet name for output = "excel" (default: "Descriptives").
clipboard_delim	Delimiter for output = "clipboard" (default: "\t").
word_path	File path for output = "word".
verbose	Logical. If TRUE, prints messages about excluded non-numeric columns (default: FALSE).

Details

Non-numeric columns are silently dropped (set verbose = TRUE to see which columns were excluded). When a single constant column is passed, SD and CI are shown as "---" in the ASCII table.

Optional output engines require suggested packages:

- **tinytable** for output = "tinytable"
- **gt** for output = "gt"
- **flextable** for output = "flextable"
- **flextable + officer** for output = "word"
- **openxlsx2** for output = "excel"
- **clipr** for output = "clipboard"

Value

Depends on output:

- "default": prints a styled ASCII table and returns the underlying data.frame invisibly (S3 class "spicy_continuous_table").
- "data.frame": a plain data.frame with columns variable, label, group (when by is used), mean, sd, min, max, ci_lower, ci_upper, n. When by is used together with p_value = TRUE, statistic = TRUE, or effect_size = TRUE, additional columns are appended (populated on the first row of each variable block only):
 - test_type – test identifier (e.g., "welch_t", "welch_anova", "student_t", "anova", "wilcoxon", "kruskal").
 - statistic, df1, df2, p.value – test results.
 - es_type – effect-size identifier ("hedges_g", "eta_sq", "r_rb", or "epsilon_sq"), when effect_size = TRUE.
 - es_value, es_ci_lower, es_ci_upper – effect-size estimate and confidence interval bounds.
- "tinytable": a tinytable object.
- "gt": a gt_tbl object.
- "flextable": a flextable object.
- "excel" / "word": writes to disk and returns the file path invisibly.
- "clipboard": copies the table and returns the display data.frame invisibly.

See Also

[table_categorical\(\)](#) for categorical variables; [freq\(\)](#) for one-way frequency tables; [cross_tab\(\)](#) for two-way cross-tabulations.

Examples

```
table_continuous(  
  sohealth,  
  select = c(bmi, wellbeing_score)  
)  
  
table_continuous(  
  sohealth,  
  select = c(bmi, wellbeing_score),  
  output = "data.frame"  
)  
  
table_continuous(  
  sohealth,  
  select = c(bmi, wellbeing_score),  
  by = education  
)  
  
table_continuous(  
  sohealth,  
  select = c(bmi, wellbeing_score),  
  by = education,  
  p_value = TRUE,  
  statistic = TRUE  
)  
  
table_continuous(  
  sohealth,  
  select = wellbeing_score,  
  by = education,  
  effect_size_ci = TRUE,  
  effect_size_digits = 3  
)  
  
table_continuous(  
  sohealth,  
  select = wellbeing_score,  
  by = sex,  
  effect_size_ci = TRUE  
)  
  
table_continuous(  
  sohealth,  
  select = "^life_sat",  
  regex = TRUE  
)
```

```

table_continuous(
  sochealth,
  select = c(bmi, life_sat_health),
  labels = c(
    bmi = "Body mass index",
    life_sat_health = "Satisfaction with health"
  )
)

if (requireNamespace("tinytable", quietly = TRUE)) {
  table_continuous(
    sochealth,
    select = starts_with("life_sat"),
    by = education,
    output = "tinytable"
  )
}

if (requireNamespace("gt", quietly = TRUE)) {
  table_continuous(
    sochealth,
    select = starts_with("life_sat"),
    by = education,
    output = "gt"
  )
}

if (requireNamespace("flextable", quietly = TRUE)) {
  table_continuous(
    sochealth,
    select = starts_with("life_sat"),
    by = education,
    output = "flextable"
  )
}

```

table_continuous_lm *Continuous-outcome linear-model table*

Description

Builds APA-style summary tables from a series of simple linear models for one or many continuous outcomes selected with tidysselect syntax.

A single predictor is supplied with `by`, and each selected numeric outcome is fit as `lm(outcome ~ by, ...)`. When `by` is categorical, the function returns a model-based mean-comparison table with fitted means by level derived from the linear model, plus an optional single difference for dichotomous predictors. When `by` is numeric, the table reports the slope and its confidence interval.

Multiple output formats are available via `output`: a printed ASCII table ("default"), a plain wide data.frame ("data.frame"), a raw long data.frame ("long"), or rendered outputs ("tinytable", "gt", "flextable", "excel", "clipboard", "word").

Usage

```
table_continuous_lm(
  data,
  select = dplyr::everything(),
  by,
  exclude = NULL,
  regex = FALSE,
  weights = NULL,
  vcov = c("classical", "HC0", "HC1", "HC2", "HC3", "HC4", "HC4m", "HC5"),
  contrast = c("auto", "none"),
  statistic = FALSE,
  p_value = TRUE,
  show_n = TRUE,
  show_weighted_n = FALSE,
  effect_size = c("none", "f2"),
  r2 = c("r2", "adj_r2", "none"),
  ci = TRUE,
  labels = NULL,
  ci_level = 0.95,
  digits = 2,
  fit_digits = 2,
  effect_size_digits = 2,
  decimal_mark = ".",
  output = c("default", "data.frame", "long", "tinytable", "gt", "flextable", "excel",
    "clipboard", "word"),
  excel_path = NULL,
  excel_sheet = "Linear models",
  clipboard_delim = "\t",
  word_path = NULL,
  verbose = FALSE
)
```

Arguments

<code>data</code>	A data.frame.
<code>select</code>	Outcome columns to include. If <code>regex = FALSE</code> , use <code>tidyselect</code> syntax or a character vector of column names (default: <code>dplyr::everything()</code>). If <code>regex = TRUE</code> , provide a regular expression pattern (character string).
<code>by</code>	A single predictor column. Accepts an unquoted column name or a single character column name. The predictor can be numeric, logical, character, or factor.
<code>exclude</code>	Columns to exclude from <code>select</code> . Supports <code>tidyselect</code> syntax and character vectors of column names.

regex	Logical. If FALSE (the default), uses tidyselect helpers. If TRUE, the select argument is treated as a regular expression.
weights	Optional case weights. Accepts NULL, an unquoted numeric column name, a single character column name, or a numeric vector of length nrow(data).
vcov	Variance estimator used for uncertainty estimates. One of "classical" (default), "HC0", "HC1", "HC2", "HC3", "HC4", "HC4m", or "HC5".
contrast	Contrast display for categorical predictors. One of: <ul style="list-style-type: none"> "auto": show a single reference contrast only when by has exactly two levels. "none": suppress difference columns for categorical predictors.
statistic	Logical. If TRUE, includes a test-statistic column in the wide and rendered outputs. Defaults to FALSE.
p_value	Logical. If TRUE, includes a p column in the wide and rendered outputs. Defaults to TRUE.
show_n	Logical. If TRUE, includes an unweighted n column in the wide and rendered outputs. Defaults to TRUE.
show_weighted_n	Logical. If TRUE and weights is supplied, includes a Weighted n column equal to the sum of case weights in the analytic sample. Defaults to FALSE.
effect_size	Character. Effect-size column to include in the wide and rendered outputs. One of "none" (the default) or "f2".
r2	Character. Fit statistic to include in the wide and rendered outputs. One of "r2" (the default), "adj_r2", or "none".
ci	Logical. If TRUE, includes contrast confidence-interval columns in the wide and rendered outputs when a single contrast is shown. Defaults to TRUE.
labels	An optional named character vector of outcome labels. Names must match column names in data. When NULL (the default), labels are auto-detected from variable attributes; if none are found, the column name is used.
ci_level	Confidence level for coefficient and model-based mean intervals (default: 0.95). Must be between 0 and 1 exclusive.
digits	Number of decimal places for descriptive values, regression coefficients, and test statistics (default: 2).
fit_digits	Number of decimal places for model-fit columns (R2 or adjusted R2) in wide and rendered outputs (default: 2).
effect_size_digits	Number of decimal places for effect-size columns (f2) in wide and rendered outputs (default: 2).
decimal_mark	Character used as decimal separator. Either "." (default) or ",".
output	Output format. One of: <ul style="list-style-type: none"> "default": a printed ASCII table, returned invisibly "data.frame": a plain wide data.frame "long": a raw long data.frame

- "tinytable" (requires tinytable)
- "gt" (requires gt)
- "flextable" (requires flextable)
- "excel" (requires openxlsx2)
- "clipboard" (requires clipr)
- "word" (requires flextable and officer)

excel_path	File path for output = "excel".
excel_sheet	Sheet name for output = "excel" (default: "Linear models").
clipboard_delim	Delimiter for output = "clipboard" (default: "\t").
word_path	File path for output = "word".
verbose	Logical. If TRUE, prints messages about ignored non-numeric selected outcomes (default: FALSE).

Details

table_continuous_lm() is designed for article-style bivariate reporting: a single predictor supplied with by, and one simple model per selected continuous outcome. The model fit is always `lm(outcome ~ by, ...)`.

For categorical predictors, the reported means are model-based fitted means for each level of by, and the reported contrasts are derived from the same fitted linear model.

Compared with `table_continuous()`, this function is the model-based companion for users who want to report bivariate mean comparisons in a linear-model framework. In practice, it is the better choice when you want heteroskedasticity-consistent standard errors (`vcov = "HC*"`), model fit statistics, or case weights via `lm(..., weights = ...)`.

Effect size is reported as Cohen's f^2 , computed from the model R^2 as $R^2 / (1 - R^2)$. When `vcov != "classical"`, standard errors, confidence intervals, and test statistics use the requested heteroskedasticity-consistent estimator, while R^2 and adjusted R^2 remain the ordinary least-squares fit statistics.

When `weights` is supplied, `table_continuous_lm()` fits weighted linear models using `lm(..., weights = ...)` and reports weighted model-based means or slopes accordingly. This is appropriate for case-weighted analyses and weighted article tables, but it is not a substitute for a full complex-survey design workflow.

In wide and rendered outputs, `n` always reports the unweighted analytic sample size used for each outcome. When `show_weighted_n = TRUE`, an additional `Weighted n` column reports the sum of case weights for the same analytic sample.

For dichotomous categorical predictors, the wide outputs report means in reference-level order and labels the contrast column explicitly as `Delta (level2 - level1)`. For categorical predictors with more than two levels, no single contrast or contrast confidence interval is shown in the wide outputs; instead, the table reports level-specific means plus the overall F test when `statistic = TRUE` (or `F(df1, df2)` when the degrees of freedom are constant across outcomes).

Optional output engines require suggested packages:

- **tinytable** for output = "tinytable"

- **gt** for output = "gt"
- **flextable** for output = "flextable"
- **flextable + officer** for output = "word"
- **openxlsx2** for output = "excel"
- **clipr** for output = "clipboard"

Value

Depends on output:

- "default": prints a styled ASCII table and returns the underlying long data.frame invisibly, with class "spicy_continuous_lm_table".
- "data.frame": a plain wide data.frame with one row per outcome and numeric columns for means, optional contrasts, optional test statistics, model fit, and sample size.
- "long": a raw data.frame with one block per outcome and columns describing estimated means, contrasts, test statistics, model fit, and sample size.
- "tinytable": a tinytable object.
- "gt": a gt_tbl object.
- "flextable": a flextable object.
- "excel" / "word": writes to disk and returns the file path.
- "clipboard": copies the wide table and returns it invisibly.

See Also

[table_continuous\(\)](#), [table_categorical\(\)](#)

Examples

```
table_continuous_lm(
  sochealth,
  select = c(wellbeing_score, bmi),
  by = sex
)
```

```
table_continuous_lm(
  sochealth,
  select = c(wellbeing_score, bmi),
  by = sex,
  output = "data.frame"
)
```

```
table_continuous_lm(
  sochealth,
  select = c(wellbeing_score, bmi),
  by = age,
  vcov = "HC3",
  ci = FALSE
)
```

```
table_continuous_lm(  
  sochealth,  
  select = c(wellbeing_score, bmi),  
  by = sex,  
  weights = weight,  
  statistic = TRUE,  
  effect_size = "f2",  
  show_weighted_n = TRUE  
)  
  
if (requireNamespace("tinytable", quietly = TRUE)) {  
  table_continuous_lm(  
    sochealth,  
    select = c(wellbeing_score, bmi),  
    by = sex,  
    output = "tinytable"  
  )  
}  
  
if (requireNamespace("gt", quietly = TRUE)) {  
  table_continuous_lm(  
    sochealth,  
    select = c(wellbeing_score, bmi),  
    by = sex,  
    output = "gt"  
  )  
}  
  
if (requireNamespace("flextable", quietly = TRUE)) {  
  table_continuous_lm(  
    sochealth,  
    select = c(wellbeing_score, bmi),  
    by = sex,  
    output = "flextable"  
  )  
}
```

uncertainty_coef

Uncertainty Coefficient

Description

`uncertainty_coef()` computes the Uncertainty Coefficient (Theil's U) for a two-way contingency table, based on information entropy.

Usage

```
uncertainty_coef(
  x,
  direction = c("symmetric", "row", "column"),
  detail = FALSE,
  conf_level = 0.95,
  digits = 3L,
  .include_se = FALSE
)
```

Arguments

<code>x</code>	A contingency table (of class table).
<code>direction</code>	Direction of prediction: "symmetric" (default), "row" (column predicts row), or "column" (row predicts column).
<code>detail</code>	Logical. If FALSE (default), return the estimate as a numeric scalar. If TRUE, return a named numeric vector including confidence interval and p-value.
<code>conf_level</code>	A number between 0 and 1 giving the confidence level (default 0.95). Only used when <code>detail = TRUE</code> . Set to NULL to omit the confidence interval.
<code>digits</code>	Number of decimal places used when printing the result (default 3). Only affects the <code>detail = TRUE</code> output.
<code>.include_se</code>	Internal parameter; do not use.

Details

The uncertainty coefficient measures association using Shannon entropy. For `direction = "row"`: $U = (H_X + H_Y - H_{XY})/H_X$, where H_X , H_Y are the marginal entropies and H_{XY} is the joint entropy. The symmetric version is $U = 2(H_X + H_Y - H_{XY})/(H_X + H_Y)$. Standard error formulas follow the DescTools implementations (Signorell et al., 2024); see [cramer_v\(\)](#) for full references.

Value

Same structure as [cramer_v\(\)](#): a scalar when `detail = FALSE`, a named vector when `detail = TRUE`. The p-value tests $H_0: U = 0$ (Wald z-test).

See Also

[lambda_gk\(\)](#), [goodman_kruskal_tau\(\)](#), [assoc_measures\(\)](#)

Examples

```
tab <- table(sochealth$smoking, sochealth$education)
uncertainty_coef(tab)
uncertainty_coef(tab, direction = "row", detail = TRUE)
```

varlist	<i>Generate a comprehensive summary of the variables</i>
---------	----------------------------------------------------------

Description

`varlist()` lists the variables of a data frame and extracts essential metadata, including variable names, labels, summary values, classes, number of distinct values, number of valid (non-missing) observations, and number of missing values.

`vl()` is a convenient shorthand for `varlist()` that offers identical functionality with a shorter name.

Usage

```
varlist(
  x,
  ...,
  values = FALSE,
  tbl = FALSE,
  include_na = FALSE,
  .raw_expr = substitute(x)
)
```

```
vl(x, ..., values = FALSE, tbl = FALSE, include_na = FALSE)
```

Arguments

<code>x</code>	A data frame or a transformation of one. Must be named and identifiable.
<code>...</code>	Optional tidyselect-style column selectors (e.g. <code>starts_with("var")</code> , <code>where(is.numeric)</code> , etc.).
<code>values</code>	Logical. If FALSE (the default), only min/max or representative values are displayed. If TRUE, all unique values are listed.
<code>tbl</code>	Logical. If FALSE (the default), the summary is opened in the Viewer (if interactive). If TRUE, a tibble is returned instead.
<code>include_na</code>	Logical. If TRUE, missing values (NA) are included in the Values column. Default is FALSE.
<code>.raw_expr</code>	Internal. Do not use. Captures the original expression from <code>vl()</code> to generate an informative title. Used only for internal purposes.

Details

The function can also apply tidyselect-style variable selectors to filter columns dynamically.

If used interactively (e.g. in RStudio), the summary is displayed in the Viewer pane with a contextual title like `vl: sochealth`. If the data frame has been transformed or subsetted, the title will display an asterisk (*), e.g. `vl: sochealth*`.

For full documentation, see [varlist\(\)](#).

Value

A tibble with one row per (selected) variable, containing the following columns:

- Variable: variable names
- Label: variable labels (if available via the label attribute)
- Values: a summary of the variable's values, depending on the values and include_na arguments. If values = FALSE, a compact summary (max 4 values: 3 + ... + last) is shown. If values = TRUE, all unique non-missing values are displayed. For labelled variables, **prefixed labels** are displayed using labelled::to_factor(levels = "prefixed"). For factors, levels are used as-is. Missing values (NA, NaN) are optionally appended at the end (controlled via include_na).
- Class: the class of each variable (possibly multiple, e.g. "labelled", "numeric")
- N_distinct: number of distinct non-missing values
- N_valid: number of non-missing observations
- NAs: number of missing observations If tbl = FALSE and used interactively, the summary is displayed in the Viewer pane. If the data frame is a transformation (e.g. head(df) or df[, 1:3]), an asterisk (*) is appended to the name in the title (e.g. v1: df*).

Examples

```
varlist(sochealth)
sochealth |> varlist()
varlist(sochealth, where(is.numeric), values = TRUE, tbl = TRUE)
varlist(sochealth, tbl = TRUE)
varlist(sochealth, starts_with("bmi"), tbl = TRUE)

v1(sochealth)
sochealth |> v1()
v1(sochealth, starts_with("bmi"))
v1(sochealth, where(is.numeric), values = TRUE, tbl = TRUE)
```

yule_q

Yule's Q

Description

yule_q() computes Yule's Q coefficient of association for a 2x2 contingency table.

Usage

```
yule_q(x, detail = FALSE, conf_level = 0.95, digits = 3L, .include_se = FALSE)
```

Arguments

x	A contingency table (of class table).
detail	Logical. If FALSE (default), return the estimate as a numeric scalar. If TRUE, return a named numeric vector including confidence interval and p-value.
conf_level	A number between 0 and 1 giving the confidence level (default 0.95). Only used when detail = TRUE. Set to NULL to omit the confidence interval.
digits	Number of decimal places used when printing the result (default 3). Only affects the detail = TRUE output.
.include_se	Internal parameter; do not use.

Details

For a 2x2 table with cells a, b, c, d , Yule's Q is $Q = (ad - bc)/(ad + bc)$. It is equivalent to the Goodman-Kruskal Gamma for 2x2 tables. The asymptotic standard error is $SE = 0.5(1 - Q^2)\sqrt{1/a + 1/b + 1/c + 1/d}$. Standard error formulas follow the DescTools implementations (Signorell et al., 2024); see [cramer_v\(\)](#) for full references.

Value

Same structure as [cramer_v\(\)](#): a scalar when detail = FALSE, a named vector when detail = TRUE. The p-value tests $H_0: Q = 0$ (Wald z-test).

See Also

[phi\(\)](#), [gamma_gk\(\)](#), [assoc_measures\(\)](#)

Examples

```
tab <- table(sochealth$smoking, sochealth$sex)
yule_q(tab)
```

Index

* datasets

- sochealth, 29
- assoc_measures, 3
- assoc_measures(), 6, 13, 20–23, 26, 29, 32, 52, 55
- build_ascii_table(), 32–34
- clipr::write_clip(), 7
- code_book, 4
- contingency_coef, 5
- contingency_coef(), 3, 13
- copy_clipboard, 6
- count_n, 8
- cramer_v, 12
- cramer_v(), 3, 4, 6, 20–23, 25, 26, 29, 32, 52, 55
- crayon::has_color(), 33
- cross_tab, 13
- cross_tab(), 31, 32, 37–40, 45
- datawizard::row_count(), 9, 11
- freq, 16
- freq(), 13, 14, 32, 34, 40, 45
- gamma_gk, 19
- gamma_gk(), 3, 4, 22, 23, 32, 55
- goodman_kruskal_tau, 20
- goodman_kruskal_tau(), 3, 26, 52
- kendall_tau_b, 21
- kendall_tau_b(), 3, 4, 20, 23, 32
- kendall_tau_c, 22
- kendall_tau_c(), 3, 20, 22
- label_from_names, 24
- labelled::var_label(), 24, 31
- lambda_gk, 25
- lambda_gk(), 3, 21, 52
- mean_n, 26
- phi, 28
- phi(), 3, 13, 55
- print.spicy_freq_table(), 16–18, 34
- sochealth, 29
- somers_d, 31
- somers_d(), 3, 20, 22, 23
- spicy_print_table, 32
- spicy_print_table(), 16, 18
- sum_n, 34
- table_categorical, 37
- table_categorical(), 45, 50
- table_continuous, 41
- table_continuous(), 40, 49, 50
- table_continuous_lm, 46
- tidyselect::starts_with(), 9
- uncertainty_coef, 51
- uncertainty_coef(), 3, 21, 26
- varlist, 53
- varlist(), 4, 5, 31, 53
- vl(varlist), 53
- yule_q, 54
- yule_q(), 3, 29