Package ‘pollster’

May 12, 2023

Type Package

Title Calculate Crosstab and Topline Tables of Weighted Survey Data

Version 0.1.6

Author John D. Johnson [aut, cre]

Maintainer John D. Johnson <john.d.johnson@marquette.edu>

Description Calculate common types of tables for weighted survey data. Options include topline and (2-way and 3-way) crosstab tables of categorical or ordinal data as well as summary tables of weighted numeric variables. Optionally, include the margin of error at selected confidence intervals including the design effect. The design effect is calculated as described by Kish (1965) <doi:10.1002/bimj.19680100122> beginning on page 257. Output takes the form of tibbles (simple data frames). This package conveniently handles labelled data, such as that commonly used by ‘Stata’ and ‘SPSS.’ Complex survey design is not supported at this time.

Depends R (>= 2.10)

Imports dplyr (>= 0.8.0), stringr (>= 1.0.0), tidyr (>= 1.1.0), labelled (>= 2.0.0), forcats (>= 1.0.0), rlang (>= 0.4.5)

Suggests ggplot2 (>= 3.3.0), knitr, rmarkdown

License CC0

Encoding UTF-8

LazyData true

RoxygenNote 7.2.0

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2023-05-12 19:00:05 UTC
R topics documented:

- crosstab
- crosstab_3way
- deff_calc
- illinois
- moedef_calc
- moe_crosstab
- moe_crosstab_3way
- moe_topline
- moe_wave_crosstab
- moe_wave_crosstab_3way
- summary_table
- topline
- wtd_mean

Index

16

crosstab

**Description**

crosstab returns a tibble containing a weighted crosstab of two variables

**Usage**

```r
crosstab(
  df,
  x,
  y,
  weight,
  remove = "",
  n = TRUE,
  pct_type = "row",
  format = "wide",
  unwt_n = FALSE
)
```

**Arguments**

- `df` The data source
- `x` The independent variable
- `y` The dependent variable
- `weight` The weighting variable
crosstab_3way

remove
An optional character vector of values to remove from final table (e.g. "refused"). This will not affect any calculations made. The vector is not case-sensitive.

n
logical, if TRUE numeric totals are included. They are included in a separate column for row and cell percentages, but in a separate row for wide format column percentages.

pct_type
Controls the kind of percentage values returned. One of "row," "cell," or "column."

format
one of "long" or "wide"

unwt_n
logical, if TRUE a column "unweighted_n" is included containing the unweighted frequency count. It is not available when pct_type is "column"

Details
Options include row, column, or cell percentages. The tibble can be in long or wide format.

Value
a tibble

Examples

crosstab(df = illinois, x = voter, y = raceethnic, weight = weight)
crosstab(df = illinois, x = voter, y = raceethnic, weight = weight, n = FALSE)

crosstab_3way
weighted 3-way crosstabs

description

crosstab_3way returns a tibble containing a weighted crosstab of two variables by a third variable

Usage

crosstab_3way(
  df,
  x,
  y,
  z,
  weight,
  remove = c(""),
  n = TRUE,
  pct_type = "row",
  format = "wide",
  unwt_n = FALSE
)
Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
<td>The data source</td>
</tr>
<tr>
<td>x</td>
<td>The independent variable</td>
</tr>
<tr>
<td>y</td>
<td>The dependent variable</td>
</tr>
<tr>
<td>z</td>
<td>The second control variable</td>
</tr>
<tr>
<td>weight</td>
<td>The weighting variable</td>
</tr>
<tr>
<td>remove</td>
<td>An optional character vector of values to remove from final table (e.g. &quot;refused&quot;). This will not affect any calculations made. The vector is not case-sensitive.</td>
</tr>
<tr>
<td>n</td>
<td>logical, if TRUE numeric totals are included.</td>
</tr>
<tr>
<td>pct_type</td>
<td>Controls the kind of percentage values returned. One of &quot;row&quot; or &quot;cell.&quot;</td>
</tr>
<tr>
<td>format</td>
<td>one of &quot;long&quot; or &quot;wide&quot;</td>
</tr>
<tr>
<td>unwt_n</td>
<td>logical, if TRUE a column is added containing unweighted frequency counts</td>
</tr>
</tbody>
</table>

Details

Options include row or cell percentages. The tibble can be in long or wide format. These tables are ideal for use with small multiples created with ggplot2::facet_wrap.

Value

a tibble

Examples

crosstab_3way(df = illinois, x = sex, y = educ6, z = maritalstatus, weight = weight)
crosstab_3way(df = illinois, x = sex, y = educ6, z = maritalstatus, weight = weight, format = "wide")

deff_calc

Description

deff_calc returns a single number

Usage

deff_calc(w)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>w</td>
<td>a vector of weights</td>
</tr>
</tbody>
</table>
Details

This function returns the design effect of a given sample using the formula length(w)*sum(w^2)/(sum(w)^2). It is designed for use in the moe family of functions. If any weights are equal to 0, they are removed prior to calculation.

Value

A number

Examples

def_eff_calc(illinois$weight)

Description


Usage

illinois

Format

A data frame with 36207 rows and 9 variables:

- **year**: year of survey
- **fips**: the state fips code
- **sex**: sex of the respondent, labelled value
- **educ6**: highest level of education for respondent, labelled values
- **raceethnic**: one of white, black, Hispanic, or other, labelled values
- **maritalstatus**: one of Married, Widowed/divorced/Sep, or Never Married, labelled values
- **rv**: indicates if the respondent is registered to vote, labelled values
- **voter**: indicates if the respondent voted, labelled values
- **age**: the age of the respondent, numeric values
- **weight**: the number of people each respondent is calculated to represent

Source

https://www.census.gov/topics/public-sector/voting.html
moedeff_calc  
*Calculate the margin of error (including design effect) of a sample*

Description

`moedeff_calc` returns a single number. It is designed for use in the moe family of functions.

Usage

`moedeff_calc(pct, deff, n, zscore = 1.96)`

Arguments

- `pct`: a proportion
- `deff`: a design effect
- `n`: the sample size
- `zscore`: defaults to 1.96, consistent with a 95% confidence interval.

Details

This function returns the margin of error including design effect of a given sample of weighted data using the formula `sqrt(deff)*zscore*sqrt((pct*(1-pct))/(n-1))*100`

Value

A percentage

Examples

`moedeff_calc(pct = 0.515, deff = 1.6, n = 214)`

moe_crosstab  
*weighted crosstabs with margin of error*

Description

`moe_crosstab` returns a tibble containing a weighted crosstab of two variables with margin of error
Usage

```r
moe_crosstab(
    df,
    x,
    y,
    weight,
    remove = c(""),
    n = TRUE,
    pct_type = "row",
    format = "long",
    zscore = 1.96,
    unwt_n = FALSE
)
```

Arguments

- `df` The data source
- `x` The independent variable
- `y` The dependent variable
- `weight` The weighting variable, defaults to `zwave_weight`
- `remove` An optional character vector of values to remove from final table (e.g. "refused"). This will not affect any calculations made. The vector is not case-sensitive.
- `n` logical, if TRUE numeric totals are included.
- `pct_type` Controls the kind of percentage values returned. One of "row" or "cell." Column percents are not supported.
- `format` one of "long" or "wide"
- `zscore` defaults to 1.96, consistent with a 95% confidence interval
- `unwt_n` logical, if TRUE it adds a column with unweighted frequency values

Details

Options include row or cell percentages. The tibble can be in long or wide format. The margin of error includes the design effect of the weights.

Value

a tibble

Examples

```r
moe_crosstab(df = illinois, x = voter, y = raceethnic, weight = weight)
moe_crosstab(df = illinois, x = voter, y = raceethnic, weight = weight, n = FALSE)
```
**moe_crosstab_3way**

**weighted 3-way crosstabs with margin of error**

**Description**

moe_crosstab_3way returns a tibble containing a weighted crosstab of two variables by a third variable with margin of error.

**Usage**

```r
moe_crosstab_3way(
  df,
  x,
  y,
  z,
  weight,
  remove = c(""),
  n = TRUE,
  pct_type = "row",
  format = "long",
  zscore = 1.96,
  unwt_n = FALSE
)
```

**Arguments**

- `df`: The data source
- `x`: The independent variable
- `y`: The dependent variable
- `z`: The second control variable
- `weight`: The weighting variable
- `remove`: An optional character vector of values to remove from final table (e.g. "refused"). This will not affect any calculations made. The vector is not case-sensitive.
- `n`: logical, if TRUE numeric totals are included.
- `pct_type`: Controls the kind of percentage values returned. One of "row" or "cell."
- `format`: one of "long" or "wide"
- `zscore`: defaults to 1.96, consistent with a 95% confidence interval
- `unwt_n`: logical, if TRUE it adds a column with unweighted frequency values

**Details**

Options include row or cell percentages. The tibble can be in long or wide format. These tables are ideal for use with small multiples created with ggplot2::facet_wrap.
moe_topline

Value
a tibble

Examples
moe_crosstab_3way(df = illinois, x = sex, y = educ6, z = maritalstatus, weight = weight)
moe_crosstab_3way(df = illinois, x = sex, y = educ6, z = maritalstatus, weight = weight, format = "wide")

moe_topline weighted topline with margin of error

Description
moe_topline returns a tibble containing a weighted topline of one variable with margin of error

Usage
moe_topline(
  df,
  variable,
  weight,
  remove = c(""),
  n = TRUE,
  pct = TRUE,
  valid_pct = TRUE,
  cum_pct = TRUE,
  zscore = 1.96
)

Arguments
df The data source
variable the variable name
weight The weighting variable, defaults to zwave_weight
remove An optional character vector of values to remove from final table (e.g. "refused"). This will not affect any calculations made. The vector is not case-sensitive.
n logical, if TRUE a frequency column is included percentages, but in a separate row for column percentages.
pct logical, if TRUE a column of percents is included
valid_pct logical, if TRUE a column of valid percents is included
cum_pct logical, if TRUE a column of cumulative percents is included
zscore defaults to 1.96, consistent with a 95% confidence interval
moe_wave_crosstab

Details

By default the table includes a column for frequency count, percent, valid percent, and cumulative percent.

Value

a tibble

Examples

moe_topline(df = illinois, variable = educ6, weight = weight)
moe_topline(df = illinois, variable = educ6, weight = weight, remove = c("LT HS"))

moe_wave_crosstab weighted crosstabs with margin of error, where the x-variable identifies different survey waves

Description

moe_wave_crosstab returns a tibble containing a weighted crosstab of two variables with margin of error. Use this function when the x-variable indicates different survey waves for which weights were calculated independently.

Usage

moe_wave_crosstab(
  df,
  x,
  y,
  weight,
  remove = c(""),
  n = TRUE,
  pct_type = "row",
  format = "long",
  zscore = 1.96,
  unwt_n = FALSE
)

Arguments

df The data source
x The independent variable, which uniquely identifies survey waves
y The dependent variable
weight The weighting variable, defaults to zwave_weight
remove An optional character vector of values to remove from final table (e.g. "refused"). This will not affect any calculations made. The vector is not case-sensitive.
moe_wave_crosstab_3way

n logical, if TRUE numeric totals are included.
pct_type Controls the kind of percentage values returned. One of "row" or "cell." Column percents are not supported.
format one of "long" or "wide"
zscore defaults to 1.96, consistent with a 95% confidence interval
unwt_n logical, if TRUE it adds a column with unweighted frequency values

Details
Options include row or cell percentages. The tibble can be in long or wide format. The margin of error includes the design effect of the weights, calculated separately for each survey wave.

Value
a tibble

Examples
moe_wave_crosstab(df = illinois, x = year, y = maritalstatus, weight = weight)
moe_wave_crosstab(df = illinois, x = year, y = maritalstatus, weight = weight, format = "wide")

Description
moe_wave_crosstab_3way returns a tibble containing a weighted crosstab of two variables by a third variable with margin of error. Use this function when the z-variable indicates different survey waves for which weights were calculated independently.

Usage
moe_wave_crosstab_3way(
  df,
  x,
  y,
  z,
  weight,
  remove = c(""),
  n = TRUE,
  pct_type = "row",
  format = "long",
  zscore = 1.96,
  unwt_n = FALSE
)
Arguments

- **df**: The data source
- **x**: The independent variable
- **y**: The dependent variable
- **z**: The second control variable, uniquely identifies survey waves
- **weight**: The weighting variable
- **remove**: An optional character vector of values to remove from final table (e.g., "refused"). This will not affect any calculations made. The vector is not case-sensitive.
- **n**: logical, if TRUE numeric totals are included.
- **pct_type**: Controls the kind of percentage values returned. One of "row" or "cell."
- **format**: one of "long" or "wide"
- **zscore**: defaults to 1.96, consistent with a 95% confidence interval
- **unwt_n**: logical, if TRUE it adds a column with unweighted frequency values

Details

Options include row or cell percentages. The tibble can be in long or wide format. These tables are ideal for use with small multiples created with ggplot2::facet_wrap.

Value

a tibble

Examples

```r
moe_crosstab_3way(df = illinois, x = sex, y = educ6, z = year, weight = weight)
moe_crosstab_3way(df = illinois, x = sex, y = educ6, z = year, weight = weight, format = "wide")
```

summary_table weighted summary table

Description

**summary_table** returns a tibble containing a weighted summary table of a single variable.

Usage

```r
summary_table(df, variable, weight, name_style = "clean")
```
Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
<td>The data source</td>
</tr>
<tr>
<td>variable</td>
<td>the variable to summarize, it should be numeric</td>
</tr>
<tr>
<td>weight</td>
<td>The weighting variable</td>
</tr>
<tr>
<td>name_style</td>
<td>the style of the column names--one of &quot;clean&quot; or &quot;pretty.&quot; Clean names are all lower case and words are separated by an underscore. Pretty names begin with a capital letter are words a separated by a space.</td>
</tr>
</tbody>
</table>

Details

The resulting tibble includes columns for the variable name, unweighted observations, weighted observations, weighted mean, minimum value, maximum value, unweighted missing values, and weighted missing values.

Value

a tibble

Examples

summary_table(illinois, age, weight)
summary_table(illinois, age, weight, name_style = "pretty")

Description

topline returns a tibble containing a weighted topline of one variable

Usage

topline(
    df,
    variable,
    weight,
    remove = c(""),
    n = TRUE,
    pct = TRUE,
    valid_pct = TRUE,
    cum_pct = TRUE
)
**Arguments**

- **df**
  - The data source

- **variable**
  - The variable name

- **weight**
  - The weighting variable, defaults to `zwave_weight`

- **remove**
  - An optional character vector of values to remove from final table (e.g. "refused"). This will not affect any calculations made. The vector is not case-sensitive.

- **n**
  - logical, if TRUE a frequency column is included percentages, but in a separate row for column percentages.

- **pct**
  - logical, if TRUE a column of percents is included

- **valid_pct**
  - logical, if TRUE a column of valid percents is included

- **cum_pct**
  - logical, if TRUE a column of cumulative percents is included

**Details**

By default the table includes a column for frequency count, percent, valid percent, and cumulative percent.

**Value**

a tibble

**Examples**

```r
topline(illinois, sex, weight)
topline(illinois, sex, weight, pct = FALSE)
```

**Description**

`wtd_mean` returns the weighted mean of a variable. It’s a tidy-compatible wrapper around `stats::weighted.mean()`.

**Usage**

```r
wtd_mean(df, variable, weight)
```

**Arguments**

- **df**
  - The data source

- **variable**
  - The variable, it should be numeric

- **weight**
  - The weighting variable
**wtd_mean**

**Value**

a numeric value

**Examples**

```r
wtd_mean(illinois, age, weight)
```

```r
library(dplyr)
illinois %>% wtd_mean(age, weight)
```
Index

* datasets
  illinois, 5
  crosstab, 2
  crosstab_3way, 3
  deff_calc, 4
  illinois, 5
  moe_crosstab, 6
  moe_crosstab_3way, 8
  moe_topline, 9
  moe_wave_crosstab, 10
  moe_wave_crosstab_3way, 11
  moedef_calc, 6
  summary_table, 12
  topline, 13
  wtd_mean, 14