

Package: chensus (via r-universe)

May 29, 2026

Type Package

Title Estimate Totals, Means, Proportions and Confidence Intervals of the Swiss Federal Statistic Office's Surveys

Version 2.1.0

Description Estimates population totals, means, proportions and confidence intervals from the Federal Statistic Office's (FSO) structural and mobility surveys. Give it data from ``Strukturerhebung"/``relevé structurel" or from ``Mikrozensus für Mobilität und Verkehr"/``Microrecensement mobilité et transports", and obtain estimates of totals, means, proportions and confidence intervals.

License GPL (>= 3)

URL <https://github.com/afds-bl/chensus>

Imports dplyr, purrr, rlang, stringr, tidyr, forcats, fastDummies

Suggests knitr, rmarkdown, spelling, testthat (>= 3.0.0), tibble

Depends R (>= 4.1.0)

Config/testthat/edition 3

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

VignetteBuilder knitr

Language en-US

Config/pak/sysreqs libicu-dev

Repository <https://swissstatsr.r-universe.dev>

Date/Publication 2025-09-01 13:23:51 UTC

RemoteUrl <https://github.com/afds-bl/chensus>

RemoteRef HEAD

RemoteSha 468ab0e0008276895bdc0ff0d476d1a732c1f50

Contents

fso_flag_mask	2
mzmv_mean	3
mzmv_mean_map	4
nhanes	5
se_mean	7
se_mean_ogd	8
se_ogd_wrapper	9
se_prop	10
se_prop_ogd	12
se_total	13
se_total_map	14
se_total_ogd	16
se_total_prop	17
se_total_prop_ogd	18
Index	19

fso_flag_mask	<i>Classify Estimate Reliability and Apply Confidentiality Masking</i>
---------------	--

Description

fso_flag_mask applies Swiss Federal Statistical Office (FSO) reliability rules for survey estimates, based on the number of observations (occ). It flags low reliability estimates and masks them when sample size is too small ($occ \leq 4$).

Usage

```
fso_flag_mask(data, lang = c("de", "fr", "it", "en"))
```

Arguments

data	A data frame or tibble.
lang	A character string for the language of the estimate reliability description, one of "de", "fr", "it", "en". Defaults to German if omitted.

Details

FSO estimate reliability criteria:

- occ \leq 4: No estimate (confidential).
- occ \leq 49: Estimate of low reliability.
- occ > 49: Reliable estimate.

Value

A tibble containing the original data with masked estimates when `occ <= 4` and one additional column:

obs_status character column classifying reliability of estimates.

Examples

```
df <- data.frame(occ = c(3, 10, 60), mean_income = c(4000, 4200, 4500))
fso_flag_mask(df)
```

mzmv_mean

Estimate Means of Mobility Survey

Description

`mzmv_mean()` estimates the means, proportions and confidence intervals of FSO mobility surveys.

Usage

```
mzmv_mean(data, ..., weight, cf = 1.14, alpha = 0.1)
```

Arguments

<code>data</code>	A data frame or tibble.
<code>...</code>	Names of variables to be estimated. Can be passed unquoted (e.g., <code>household_size</code>) or programmatically using <code>!!!syms(c("annual_household_income", "household_size"))</code> . Variables have integer values, representing a quantity (number of cars per household) or presence/absence (possession of a car). Negative numbers represent NA.
<code>weight</code>	Unquoted or quoted name of the sampling weights column. For programmatic use with a string variable (e.g., <code>wt <- "weights"</code>), use <code>!!sym(wt)</code> in the function call.
<code>cf</code>	Numeric correction factor of the confidence interval, supplied by FSO. Default is 1.14.
<code>alpha</code>	Numeric significance level for confidence intervals. Default is 0.1 (90% CI).

Value

Tibble (number of rows is length of `variable`) with the following columns:

- `id`: estimated item
- `occ`: number of survey responses
- `wmean`: weighted mean estimate
- `ci`: confidence interval estimate

See Also

See [mzmv_mean_map](#) for estimates on a set of conditions.

Examples

```
# Estimate two means
mzmv_mean(
  data = nhanes,
  annual_household_income, annual_family_income,
  weight = weights
)
# Programmatic use with strings
v <- c("annual_household_income", "annual_family_income")
mzmv_mean(nhanes, weight = "weights", !!!rlang::syms(v))
```

mzmv_mean_map

Estimate Means in Parallel for Multiple Grouping Variables in Mobility Survey

Description

`mzmv_mean_map()` estimates weighted means and confidence intervals for a set of features of the mobility survey, optionally grouped by one or more variables.

Usage

```
mzmv_mean_map(data, variable, ..., weight, cf = 1.14, alpha = 0.1)
```

Arguments

<code>data</code>	A data frame or tibble.
<code>variable</code>	Character vector of variable names to be estimated. Must be quoted (e.g., "annual_family_income"). For multiple variables, pass as a vector (e.g., c("annual_family_income", "annual_household_income")). Does not support bare (unquoted) variable names.
<code>...</code>	Grouping variables. Can be passed unquoted (e.g., gender, birth_country) or quoted (e.g., "gender", "birth_country"). If omitted, results are aggregated across the whole dataset.
<code>weight</code>	Unquoted or quoted name of the sampling weights column (must exist in data). For programmatic use with a string variable (e.g., wt <- "weights"), use <code>!!sym(wt)</code> in the function call.
<code>cf</code>	Numeric correction factor for the confidence interval. Default is 1.14.
<code>alpha</code>	Numeric significance level for confidence intervals. Default is 0.1 (90% CI).

Value

A tibble with columns:

variable Name of the estimated variable.

group_vars Name of the grouping variable.

group_vars_value Value of the grouping variable.

occ Number of cases or observations.

wmean Weighted mean.

ci Confidence interval.

Examples

```
# Multiple quoted variables
mzmv_mean_map(
  nhanes,
  variable = c("annual_family_income", "annual_household_income"),
  gender,
  birth_country,
  weight = weights
)
# No grouping variables
mzmv_mean_map(
  nhanes,
  variable = "annual_family_income",
  weight = weights
)
# Programmatic use
wt <- "weights"
mzmv_mean_map(
  nhanes,
  variable = "annual_family_income",
  gender,
  birth_country,
  weight = !!rlang::sym(wt)
)
```

nhanes

National Health and Nutrition Examination Survey (NHANES)

Description

Demographic survey data from NHANES 2015 to 2016, with data on 9971 participants, including sampling weights.

Usage

nhanes

Format

A data frame with 9971 rows and 13 variables:

PSU SDMVPSU - Masked variance pseudo-PSU

weights WTINT2YR - Full sample 2 year interview weight

strata SDMVSTRA - Masked variance pseudo-stratum

gender RIAGENDR - Gender

age RIDAGEYR - Age in years at screening

birth_country DMDBORN4 - Country of birth

marital_status DMDMARTL - Marital status

interview_lang SIALANG - Language of interview

edu_level DMDHREDU - Household reference person's education level

household_size DMDHHSIZ - Total number of people in the Household

family_size DMDFMSIZ - Total number of people in the Family

annual_household_income INDHHIN2 - Annual household income

annual_family_income INDFMIN2 - Annual family income

Note

The data sets provided in this package are derived from the NHANES database and have been adapted for educational purposes. As such, they are NOT suitable for use as a research database. For research purposes, you should download original data files from the NHANES website and follow the analysis instructions given there.

Source

[NHANES 2015-2016](#)

References

[CDC](#)

Examples

```
library(dplyr)
glimpse(nhanes)
nhanes |> dplyr::count(edu_level)
```

 se_mean

Estimate Means of Numeric Variables in Structural Survey

Description

`se_mean()` estimates the means of numeric variables along with variance and confidence intervals for FSO's structural survey.

Usage

```
se_mean(data, variable, ..., strata, weight, alpha = 0.05)
```

Arguments

<code>data</code>	A data frame or tibble.
<code>variable</code>	Unquoted or quoted name of the numeric variable whose mean is to be estimated. Programmatic usage (e.g., using <code>!!sym()</code>) is supported.
<code>...</code>	Optional grouping variables. Can be passed unquoted (e.g., <code>gender, birth_country</code>) or programmatically using <code>!!!syms(c("gender", "birth_country"))</code> .
<code>strata</code>	Unquoted or quoted name of the strata column. Defaults to <code>zone</code> if omitted.
<code>weight</code>	Unquoted or quoted name of the sampling weights column. For programmatic use with a string variable (e.g., <code>wt <- "weights"</code>), use <code>!!sym(wt)</code> in the function call.
<code>alpha</code>	Numeric significance level for confidence intervals. Default is 0.05 (95% CI).

Value

A tibble with columns:

occ Sample size (number of observations) per group.

<variable> Estimated mean of the specified numeric variable, named dynamically.

vhat, stand_dev Estimated variance of the mean (`vhat`) and its standard deviation (`stand_dev`, square root of the variance).

ci, ci_l, ci_u Confidence interval: half-width (`ci`), lower (`ci_l`) and upper (`ci_u`) bounds.

See Also

[se_prop\(\)](#)

Examples

```

# Direct column references (unquoted)
se_mean(
  data = nhanes,
  variable = age,
  strata = strata,
  weight = weights,
  gender, birth_country
)

# Quoted column names
se_mean(
  data = nhanes,
  variable = "age",
  strata = "strata",
  weight = "weights",
  gender, birth_country
)

# Programmatic use with strings
v <- "age"
wt <- "weights"
vars <- c("gender", "birth_country")
se_mean(
  data = nhanes,
  variable = !!rlang::sym(v),
  strata = strata,
  weight = !!rlang::sym(wt),
  !!!rlang::syms(vars)
)

```

se_mean_ogd

Estimate Means for All Combinations of Grouping Variables (OGD Format) in Structural Survey

Description

se_mean_ogd estimates survey means of a continuous variable for every combination of the supplied grouping variables, using se_mean internally and returning results in a format suitable for Open Government Data (OGD). The output includes means for each combination of grouping variables, as well as for the overall population.

Usage

```
se_mean_ogd(data, variable, ..., strata, weight, alpha = 0.05)
```

Arguments

data	A data frame or tibble.
variable	Variable to estimate the mean for (unquoted or programmatic).
...	Grouping variables (unquoted or programmatic).
strata	Stratification variable (unquoted or programmatic). Defaults to "zone" if omitted.
weight	Sampling weights variable (unquoted or programmatic).
alpha	Significance level for confidence intervals. Default is 0.05 (95% CI).

Value

A tibble with survey mean estimates for all combinations of grouping variables. Grouping variables are converted to factors with "Total" representing the overall group.

See Also

[se_total_ogd](#), [se_prop_ogd](#), [se_ogd_wrapper](#), [se_mean](#)

Examples

```
# Unquoted variables
se_mean_ogd(nhanes, variable = household_size, strata = strata, weight = weights, gender)

# Programmatic use
var <- "household_size"
wt <- "weights"
vars <- "gender"
se_mean_ogd(
  nhanes,
  variable = !!rlang::sym(var),
  strata = strata,
  weight = !!rlang::sym(wt),
  !!!rlang::syms(vars)
)
```

Description

OGD Wrapper for Structural Survey Estimation Functions

Usage

```
se_ogd_wrapper(
  data,
  core_fun,
  ...,
  strata,
  weight,
  alpha = 0.05,
  variable = NULL,
  show_internal = FALSE
)
```

Arguments

data	A data frame or tibble.
core_fun	The core estimation function to use, one of <code>se_mean</code> , <code>se_total</code> , <code>se_prop</code> .
...	Grouping variables (unquoted or programmatic).
strata	Stratification variable (unquoted or programmatic).
weight	Sampling weights variable (unquoted or programmatic).
alpha	Significance level for confidence intervals.
variable	(Optional) Variable to estimate mean for (only needed for <code>se_mean</code>).
show_internal	Show internal estimates of variance, standard deviation (and percent confidence interval for <code>se_total()</code>). Hidden by default.

Value

A tibble with estimates for all combinations of grouping variables.

se_prop	<i>Estimate Proportions of Categorical Variables in Structural Survey</i>
---------	---

Description

`se_prop()` estimates the proportions and confidence intervals for each level of one or multiple categorical variables of FSO's structural survey, by first converting columns into dummy variables and then estimating proportions and confidence intervals.

Usage

```
se_prop(data, ..., strata, weight, alpha = 0.05)
```

Arguments

data	A data frame or tibble.
...	Categorical variables. Can be passed unquoted (e.g., gender, birth_country) or programmatically using <code>!!!syms(c("gender", "birth_country"))</code> .
strata	Unquoted or quoted name of the strata column. Defaults to zone if omitted.
weight	Unquoted or quoted name of the sampling weights column. For programmatic use with a string variable (e.g., <code>wt <- "weights"</code>), use <code>!!sym(wt)</code> in the function call.
alpha	Numeric significance level for confidence intervals. Default is 0.05 (95% CI).

Value

A tibble with proportion estimates for all grouping column combinations, including:

occ Sample size (number of observations) per group.

prop Estimated proportion of the specified categorical variable in the corresponding group.

vhat, stand_dev Estimated variance of the mean (vhat) and its standard deviation (stand_dev, square root of the variance).

ci, ci_l, ci_u Confidence interval: half-width (ci), lower (ci_l) and upper (ci_u) bounds.

Examples

```
# Direct column references (unquoted)
se_prop(
  data = nhanes,
  interview_lang,
  birth_country,
  strata = strata,
  weight = weights
)

# Quoted column names
se_prop(
  data = nhanes,
  "interview_lang",
  gender,
  "birth_country",
  strata = "strata",
  weight = weights,
)

# Programmatic use with strings
wt <- "weights"
vars <- c("interview_lang", "gender", "birth_country")
se_prop(
  data = nhanes,
  strata = strata,
  weight = !!rlang::sym(wt),
  !!!rlang::syms(vars)
)
```

)

se_prop_ogd	<i>Estimate Proportions for All Combinations of Grouping Variables (OGD Format) in Structural Survey</i>
-------------	--

Description

se_prop_ogd estimates survey proportions for every combination of the supplied grouping variables, using se_prop internally and returning results in a format suitable for Open Government Data (OGD). The output includes proportions for each combination of grouping variables, as well as for the overall population.

Usage

```
se_prop_ogd(data, ..., strata, weight, alpha = 0.05)
```

Arguments

data	A data frame or tibble.
...	Grouping variables (unquoted or programmatic).
strata	Stratification variable (unquoted or programmatic). Defaults to "zone" if omitted.
weight	Sampling weights variable (unquoted or programmatic).
alpha	Significance level for confidence intervals. Default is 0.05 (95% CI).

Value

A tibble with survey proportion estimates for all combinations of grouping variables. Grouping variables are converted to factors with "Total" representing the overall group.

See Also

[se_total_ogd](#), [se_mean_ogd](#), [se_ogd_wrapper](#), [se_prop](#)

Examples

```
# Unquoted variables
se_prop_ogd(nhanes, strata = strata, weight = weights, gender, birth_country)

# Programmatic use
wt <- "weights"
vars <- c("gender", "birth_country")
se_prop_ogd(nhanes, strata = strata, weight = !!rlang::sym(wt), !!!rlang::syms(vars))
```

se_total	<i>Estimate Totals of Structural Survey</i>
----------	---

Description

se_total() estimates the totals and confidence intervals of FSO structural surveys.

Usage

```
se_total(data, ..., strata, weight, alpha = 0.05)
```

Arguments

data	A data frame or tibble.
...	Optional grouping variables. Can be passed unquoted (e.g., gender, birth_country) or programmatically using !!!syms(c("gender", "birth_country")).
strata	Unquoted or quoted name of the strata column. Defaults to zone if omitted.
weight	Unquoted or quoted name of the sampling weights column. For programmatic use with a string variable (e.g., wt <- "weights"), use !!sym(wt) in the function call.
alpha	Numeric significance level for confidence intervals. Default is 0.05 (95% CI).

Details

The condition argument has been deprecated and is no longer supported. Please use ... to pass grouping variables either unquoted or programmatically using rlang:

* Interactive use:

```
se_total(data, weight = my_weight, group1, group2)
```

* Programmatic use:

```
weight_var <- "my_weight"
```

```
group_vars <- c("group1", "group2")
```

```
se_total(data, weight = !!rlang::sym(weight_var), !!!rlang::syms(group_vars))
```

Value

A tibble with total estimates for all grouping column combinations, including:

<variable> Value of the grouping variables passed in ...

occ number of observations in survey sample.

total population estimate.

vhat, stand_dev Estimated variance of the total (vhat) and its standard deviation (stand_dev, square root of the variance).

ci, ci_per, ci_l, ci_u Confidence interval: half-width (ci), percentage of the total (ci_per), lower (ci_l) and upper (ci_u) bounds.

See Also

[se_total_map\(\)](#), [se_total_ogd\(\)](#).

Examples

```
# One grouping variable
se_total(
  data = nhanes,
  strata = strata,
  weight = weights,
  gender
)
# Multiple grouping variables
se_total(
  data = nhanes,
  strata = strata,
  weight = weights,
  gender, marital_status, birth_country
)
# Programmatic use and quoted variables
v <- c("gender", "marital_status", "birth_country")
se_total(
  nhanes,
  weight = "weights",
  strata = "strata",
  !!!rlang::syms(v)
)
```

se_total_map

Estimate Totals in Parallel for Multiple Grouping Variables in Structural Survey

Description

`se_total_map()` applies `se_total()` to a data frame for each of several grouping variables, returning a combined tibble of results.

Usage

```
se_total_map(data, ..., strata, weight, alpha = 0.05)
```

Arguments

data	A data frame or tibble.
...	One or more grouping variables. Can be passed unquoted (e.g., <code>gender</code> , <code>birth_country</code>) or programmatically using <code>!!!syms(c("gender", "birth_country"))</code> .
strata	Unquoted or quoted name of the strata column. Defaults to <code>zone</code> if omitted.

weight	Unquoted or quoted name of the sampling weights column. For programmatic use with a string variable (e.g., <code>wt <- "weights"</code>), use <code>!!sym(wt)</code> in the function call.
alpha	Numeric significance level for confidence intervals. Default is 0.05 (95% CI).

Details

This wrapper function allows to efficiently compute totals and confidence intervals for each grouping variable in the structural survey data in parallel.

This function iterates over each grouping variable supplied via `...`, applies `se_total()` to the data grouped by that variable, and combines the results into a single tibble. The grouping variable is renamed to `value` and its name is stored in the `variable` column for clarity.

Value

A tibble with results for each grouping variable, including:

variable The name of the grouping variable.

value The value of the grouping variable.

occ Sample size for the group.

total Estimated total for the group.

vhat, stand_dev Estimated variance of the total (`vhat`) and its standard deviation (`stand_dev`, square root of the variance).

ci, ci_per, ci_l, ci_u Confidence interval: half-width (`ci`), percentage of the total (`ci_per`), lower (`ci_l`) and upper (`ci_u`) bounds.

See Also

[se_total\(\)](#), [se_total_ogd\(\)](#).

Examples

```
# Unquoted variables
se_total_map(
  nhanes,
  weight = weights,
  strata = strata,
  gender, marital_status, birth_country
)
# Programmatic use and quoted variables
v <- c("gender", "marital_status", "birth_country")
se_total_map(
  nhanes,
  weight = "weights",
  strata = "strata",
  !!!rlang::syms(v)
)
```

se_total_ogd	<i>Estimate Totals for All Combinations of Grouping Variables (OGD Format) in Structural Survey</i>
--------------	---

Description

se_total_ogd estimates survey totals for every combination of the supplied grouping variables, using se_total internally and returning results in a format suitable for Open Government Data (OGD). The output includes totals for each combination of grouping variables, as well as for the overall population.

Usage

```
se_total_ogd(data, ..., strata, weight, alpha = 0.05)
```

Arguments

data	A data frame or tibble.
...	Grouping variables (unquoted or programmatic).
strata	Stratification variable (unquoted or programmatic). Defaults to "zone" if omitted.
weight	Sampling weights variable (unquoted or programmatic).
alpha	Significance level for confidence intervals. Default is 0.05 (95% CI).

Value

A tibble with survey estimates for all combinations of grouping variables. Grouping variables are converted to factors with "Total" representing the overall group.

See Also

[se_prop_ogd](#), [se_mean_ogd](#), [se_ogd_wrapper](#), [se_total](#)

Examples

```
# Unquoted variables
se_total_ogd(nhanes, strata = strata, weight = weights, gender, birth_country)

# Programmatic use
wt <- "weights"
vars <- c("gender", "birth_country")
se_total_ogd(nhanes, strata = strata, weight = !!rlang::sym(wt), !!!rlang::syms(vars))
```

se_total_prop	<i>Create a Table with Total and Proportion Estimates for Categorical Variables in Structural Survey</i>
---------------	--

Description

se_total_prop is a wrapper function for se_total() and se_prop() which estimates totals and proportions for categorical variables.

Usage

```
se_total_prop(data, ..., strata, weight, alpha = 0.05)
```

Arguments

data	A data frame or tibble.
...	Optional grouping variables (unquoted).
strata	The name of the strata variable (default is "zone").
weight	The name of the weight variable.
alpha	Significance level for confidence intervals. Default is 0.05.

Value

A tibble with joined total and proportion estimates.

See Also

[se_total\(\)](#), [se_prop\(\)](#).

Examples

```
se_total_prop(  
  data = nhanes,  
  interview_lang,  
  gender,  
  birth_country,  
  strata = strata,  
  weight = weights  
)
```

se_total_prop_ogd *Estimate Totals and Proportions for All Combinations of Grouping Variables (OGD Format) in Structural Survey*

Description

se_total_prop_ogd estimates totals and proportions for each combination of grouping variables using se_total_prop, returning results in a format compatible with Open Government Data (OGD) standards. along with stratification and weighting.

Usage

```
se_total_prop_ogd(data, ..., strata, weight, alpha = 0.05)
```

Arguments

data	A data frame or tibble containing the survey data.
...	Grouping variables (unquoted or programmatic) to compute combinations of totals and proportions.
strata	Stratification variable (unquoted or programmatic). Defaults to "zone" if omitted.
weight	Sampling weight variable (unquoted or programmatic).
alpha	Significance level for confidence intervals. Default is 0.05 (for 95% CI).

Value

A tibble with totals and proportions for all combinations of the specified grouping variables. The output includes confidence intervals and handles missing values by representing them as "Total".

See Also

[se_total_prop](#), [se_ogd_wrapper](#), [se_total_ogd](#), [se_prop_ogd](#)

Examples

```
# With unquoted variables
se_total_prop_ogd(nhanes, gender, birth_country, strata = strata, weight = weights)

# Programmatic usage
vars <- c("gender", "birth_country")
wt <- "weights"
se_total_prop_ogd(nhanes, !!!rlang::syms(vars), strata = strata, weight = !!!rlang::sym(wt))
```

Index

* datasets

nhanes, 5

fso_flag_mask, 2

mzmv_mean, 3

mzmv_mean_map, 4, 4

nhanes, 5

se_mean, 7, 9

se_mean_ogd, 8, 12, 16

se_ogd_wrapper, 9, 9, 12, 16, 18

se_prop, 10, 12

se_prop(), 7, 17

se_prop_ogd, 9, 12, 16, 18

se_total, 13, 16

se_total(), 14, 15, 17

se_total_map, 14

se_total_map(), 14

se_total_ogd, 9, 12, 16, 18

se_total_ogd(), 14, 15

se_total_prop, 17, 18

se_total_prop_ogd, 18