Package ‘calidad’

December 7, 2022

Type Package

Title Assesses the Quality of Estimates Made by Complex Sample Designs

Version 0.4.0

Description Assesses the quality of estimates made by complex sample designs, following the methodology developed by the National Institute of Statistics Chile (2020, <https://www.ine.cl/docs/default-source/institucionalidad/buenas-pr%C3%A1cticas/clasificaciones-y-estandares/est%C3%A1ndar-evaluaci%C3%B3n-de-calidad-de-estimaciones-publicaci%C3%B3n-27022020.pdf>) and by Economic Commission for Latin America and Caribbean (2020, <https://repositorio.cepal.org/bitstream/handle/11362/45681/1/S2000293_es.pdf>).

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.2.2

Suggests knitr, rmarkdown, testthat

VignetteBuilder knitr

Depends R (>= 3.5.0)

Imports rlang, dplyr, purrr, survey, kableExtra, stringr, tidyr, haven

NeedsCompilation no

Author Klaus Lehmann [aut, cre], Ricardo Pizarro [aut], Ignacio Agloni [ctb]

Maintainer Klaus Lehmann <klehmann@fen.uchile.cl>

Repository CRAN

Date/Publication 2022-12-07 20:50:13 UTC

R topics documented:

assess 2
assess

assess the quality of mean estimations

Description

assess assess the quality of mean estimation using the methodology created by INE Chile, which considers sample size, degrees of freedom and coefficient of variation.

Usage

assess(table, publish = FALSE, scheme = c("chile", "eclac"), ...)

Arguments

table dataframe created by crear_insumos_media

publish boolean indicating if the evaluation of the complete table must be added. If it is TRUE, the function adds a new column to the dataframe

scheme string variable, default scheme is "chile" which refers to the evaluation protocol proposed by INE Chile. the alternative is "eclac" to use the eclac protocol

... the list of cepal parameters. The complete list of parameters is

1. General Parameters
   • df degrees of freedom. default: 9
   • n sample size. default ine scheme is 60. default cepal scheme: 100

2. INE parameters
   • cv_lower_ine lower limit for cv. default: 0.15
   • cv_upper_ine upper limit for cv. default: 0.3
3. CEPAL parameters
   • cv_cepal limit for cv. default: 0.2
   • ess effective sample size. default: 140
   • unweighted unweighted count. default: 50

Value

dataframe with all the columns included in the input table, plus a new column containing a label indicating the evaluation of each estimation: reliable, bit reliable or unreliable

Examples

dc <- survey::svydesign(ids = ~varunit, strata = ~varstrat, data = epf_personas, weights = ~fe)
assess(create_mean("gastot_hd", domains = "zona+sexo", design = dc))

| casen | Encuesta de Caracterización Socioeconómica Nacional 2020 - CASEN en Pandemia 2020 |

Description

CASEN data for the year 2020. Contains only a few variables.

Usage

casen

Format

dataframe with 185.437 rows y 6 columns

- **folio** household id
- **sexo** 1 = man; 2 = woman
- **edad** age
- **activ** Economic activity status
- **ing_aut_hog** Household Income
- **pobreza** poverty status: 1 = extreme poverty, 2 = non-extreme poverty, 3 = non-poverty
- **expr** regional sample weights
- **estrato** strata
- **cod_upm** PSU

Source


Examples

data(casen)
create_html

Create html table with the results of the evaluation

Description

Create html table with the results of the evaluation

Usage

create_html(table)

Arguments

table dataframe generated by evaluate function

Value

html table

Examples

library(survey)
library(dplyr)

hogar <- epf_personas %>%
  group_by(folio) %>%
slice(1)
dc <- survey::svydesign(ids = ~varunit, strata = ~varstrat, data = hogar, weights = ~fe)
table <- assess(create_prop("ocupado", domains = "zona+sexo", design = dc))

create_mean

Create the inputs to evaluate the quality of mean estimations

Description

create_mean generates ano dataframe with the following elements: mean, degrees of freedom, sample size and coefficient of variation. The function allows grouping in several domains.

Usage

create_mean(
  var,
  domains = NULL,
  subpop = NULL,
  design,
  ci = FALSE,
create_mean

ess = FALSE,
ajuste_ene = FALSE,
standard_eval = FALSE,
rm.na = FALSE,
deff = FALSE,
rel_error = FALSE,
unweighted = FALSE,
eclac_input = FALSE
)

Arguments

- **var**: numeric variable within the dataframe.
- **domains**: domains to be estimated separated by the + character.
- **subpop**: integer dummy variable to filter the dataframe
- **design**: complex design created by survey package
- **ci**: boolean indicating if the confidence intervals must be calculated
- **ess**: boolean Effective sample size
- **ajuste_ene**: boolean indicating if an adjustment for the sampling-frame transition period must be used
- **standard_eval**: boolean Indicating if the function is wrapped inside a function, if TRUE avoid lazy eval errors
- **rm.na**: boolean Remove NA if it is required
- **deff**: boolean Design effect
- **rel_error**: boolean Relative error
- **unweighted**: boolean Add non weighted count if it is required
- **eclac_input**: boolean return eclac inputs

Value
dataframe that contains the inputs and all domains to be evaluated

Examples
dc <- survey::svydesign(ids = ~varunit, strata = ~varstrat, data = epf_personas, weights = ~fe)
create_mean("gastot_hd", "zona+sexo", design = dc)
create_prop 

Create the inputs to evaluate the quality of proportion estimations

Description

create_prop generates a dataframe with the following elements: sum, degrees of freedom, sample size, standard error and coefficient of variation. The function allows grouping in several domains.

Usage

create_prop(
  var,
  denominator = NULL,
  domains = NULL,
  subpop = NULL,
  design,
  ci = FALSE,
  deff = FALSE,
  ess = FALSE,
  ajuste_ene = FALSE,
  rel_error = FALSE,
  log_cv = FALSE,
  unweighted = FALSE,
  standard_eval = FALSE,
  eclac_input = FALSE
)

Arguments

var numeric variable within the dataframe, is the numerator of the ratio to be calculated.
denominator numeric variable within the dataframe, is the denominator of the ratio to be calculated. If the var parameter is dummy, it can be NULL
domains domains to be estimated separated by the + character.
subpop integer dummy variable to filter the dataframe
design complex design created by survey package
ci boolean indicating if the confidence intervals must be calculated
deff boolean Design effect
ess boolean Effective sample size
ajuste_ene boolean indicating if an adjustment for the sampling-frame transition period must be used
rel_error boolean Relative error
log_cv boolean logarithmic coefficient of variation
unweighted boolean Add non weighted count if it is required
create_prop_internal

standard_eval  boolean  Indicating if the function is wrapped inside a function, if TRUE avoid lazy eval errors

eclac_input  boolean  return eclac inputs

Value
dataframe that contains the inputs and all domains to be evaluated

Examples

library(survey)
library(dplyr)

epf <- mutate(epf_personas, gasto_zona1 = if_else(zona == 1, gastot_hd, 0))
dc <- svydesign(ids = ~varunit, strata = ~varstrat, data = epf, weights = ~fe)
old_options <- options()
options(survey.lonely.psu = "certainty")
create_prop(var = "gasto_zona1", denominator = "gastot_hd", design = dc)
enusc <- filter(enusc, Kish == 1)
dc <- svydesign(ids = ~Conglomerado, strata = ~VarStrat, data = enusc, weights = ~Fact_Pers)
options(survey.lonely.psu = "certainty")
create_prop(var = "VP_DC", denominator = "hom_insg_taxi", design = dc)
options(old_options)

create_prop_internal  internal function to calculate proportion estimations

Description
internal function to calculate proportion estimations

Usage

create_prop_internal(
  var,
  domains = NULL,
  subpop = NULL,
  disenio,
  ci = FALSE,
  deff = FALSE,
  ess = FALSE,
  ajuste_ene = FALSE,
  rel_error = FALSE,
  log_cv = FALSE,
  unweighted = FALSE,
)
create_ratio_internal

standard_eval = TRUE,
rm.na = FALSE,
env = parent.frame()
)

Arguments

var integer dummy variable within the dataframe
domains domains to be estimated separated by the + character.
subpop integer dummy variable to filter the dataframe
disenio complex design created by survey package
ci boolean indicating if the confidence intervals must be calculated
deff boolean Design effect
ess boolean Effective sample size
ajuste_ene boolean indicating if an adjustment for the sampling-frame transition period must be used
rel_error boolean Relative error
log_cv boolean indicating if the log cv must be returned
unweighted boolean Add non weighted count if it is required
standard_eval boolean indicating if the function is inside another function, by default it is TRUE, avoid problems with lazy eval.
rm.na boolean indicating if NA values must be removed
env parent environment to get some variables

Value
dataframe that contains the inputs and all domains to be evaluated

create_ratio_internal  internal function to calculate ratios estimations

Description
internal function to calculate ratios estimations

Usage
create_ratio_internal(
  var,
  denominator,
  domains = NULL,
  subpop = NULL,
  disenio,
create_size

```r
ci = FALSE,
deff = FALSE,
ess = FALSE,
ajuste_ene = FALSE,
unweighted = FALSE,
rel_error = FALSE,
rm.na = FALSE
```

**Arguments**

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>var</td>
<td>numeric variable within the dataframe, is the numerator of the ratio to be calculated.</td>
</tr>
<tr>
<td>denominator</td>
<td>numeric variable within the dataframe, is the denominator of the ratio to be calculated.</td>
</tr>
<tr>
<td>domains</td>
<td>domains to be estimated separated by the + character.</td>
</tr>
<tr>
<td>subpop</td>
<td>integer dummy variable to filter the dataframe</td>
</tr>
<tr>
<td>disenio</td>
<td>complex design created by survey package</td>
</tr>
<tr>
<td>ci</td>
<td>boolean indicating if the confidence intervals must be calculated</td>
</tr>
<tr>
<td>deff</td>
<td>boolean Design effect</td>
</tr>
<tr>
<td>ess</td>
<td>boolean Effective sample size</td>
</tr>
<tr>
<td>ajuste_ene</td>
<td>boolean indicating if an adjustment for the sampling-frame transition period must be used</td>
</tr>
<tr>
<td>unweighted</td>
<td>boolean Add non weighted count if it is required</td>
</tr>
<tr>
<td>rel_error</td>
<td>boolean Relative error</td>
</tr>
<tr>
<td>rm.na</td>
<td>boolean indicating if NA values must be removed</td>
</tr>
</tbody>
</table>

**Value**

dataframe that contains the inputs and all domains to be evaluated

**Description**

create_size generates a dataframe with the following elements: sum, degrees of freedom, sample size and coefficient of variation. The function allows grouping in several domains.
create_size

Usage

create_size(
  var,
  domains = NULL,
  subpop = NULL,
  design,
  ci = FALSE,
  ess = FALSE,
  ajuste_ene = FALSE,
  standard_eval = FALSE,
  rm.na = FALSE,
  deff = FALSE,
  rel_error = FALSE,
  unweighted = FALSE,
  df_type = c("ine", "eclac"),
  eclac_input = FALSE
)

Arguments

  var     numeric variable within the dataframe. When the domain parameter is not used, it is possible to include more than one variable using the + separator. When a value is introduced in the domain parameter, the estimation variable must be a dummy variable.
  domains domains to be estimated separated by the + character.
  subpop  integer dummy variable to filter the dataframe
  design  complex design created by survey package
  ci      boolean indicating if the confidence intervals must be calculated
  ess     boolean Effective sample size
  ajuste_ene boolean indicating if an adjustment for the sampling-frame transition period must be used
  standard_eval boolean Indicating if the function is wrapped inside a function, if TRUE avoid lazy eval errors
  rm.na   boolean Remove NA if it is required
  deff    boolean Design effect
  rel_error boolean Relative error
  unweighted boolean Add non weighted count if it is required
  df_type  string Use degrees of freedom calculation approach from INE Chile or CEPAL, by default "ine".
  eclac_input boolean return eclac inputs

Value

dataframe that contains the inputs and all domains to be evaluated
create_total

Examples

dc <- survey::svydesign(ids = ~varunit, strata = ~varstrat, data = epf_personas, weights = ~fe)
create_size("ocupado", "zona+sexo", design = dc)

create_total

Create the inputs to evaluate the quality of the sum of continuous variables

Description

create_total generates a dataframe with the following elements: sum, degrees of freedom, sample size and coefficient of variation. The function allows grouping in several domains.

Usage

create_total(
  var,
  domains = NULL,
  subpop = NULL,
  design,
  ci = FALSE,
  ess = FALSE,
  ajuste_ene = FALSE,
  standard_eval = FALSE,
  rm.na = FALSE,
  deff = FALSE,
  rel_error = FALSE,
  unweighted = FALSE,
  eclac_input = FALSE
)

Arguments

var numeric variable within the dataframe.
domains domains to be estimated separated by the + character.
subpop integer dummy variable to filter the dataframe
design complex design created by survey package
.ci boolean indicating if the confidence intervals must be calculated
.ess boolean Effective sample size
.ajuste_ene boolean indicating if an adjustment for the sampling-frame transition period must be used
.standard_eval boolean indicating if the function is wrapped inside a function, if TRUE avoid lazy eval errors
.rm.na boolean Remove NA if it is required
deff boolean Design effect
rel_error boolean Relative error
unweighted boolean Add non weighted count if it is required
eclac_input boolean return eclac inputs

Value
dataframe that contains the inputs and all domains to be evaluated

Examples
dc <- survey::svydesign(ids = ~varunit, strata = ~varstrat, data = epf_personas, weights = ~fe)
create_total("gastot_hd", "zona+sexo", subpop = "ocupado", design = dc)

Description
Reduced version of the ENE database. Contains some sociodemographic variables and the necessary information to work with complex design

Usage
ene

Format
dataframe with 87,842 rows y 7 columns
sexo 1 = man; 2 = woman
region region
cae_especifico Economic activity status
fe sample weights
varunit PSU
varstrat strata
fdt It shows if the person belongs to labour force: 1 = yes; 0 = no
ocupado 1 = employed; 0 = non-employed
desocupado 1 = non-employed; 0 = employed

Source
https://www.ine.cl/estadisticas/sociales/mercado-laboral/ocupacion-y-desocupacion

Examples
data(ene)
**Description**

ENUSC data for the year 2019. Contains only a few variables.

**Usage**

enusc

**Format**

dataframe with 24,465 rows and 22 columns

- **rph_sexo** 1 = man; 2 = woman
- **region** 16 regions
- **Fact_Pers** person sample weights
- **Fact_Hog** household sample weights
- **Conglomerado** PSU
- **VarStrat** strata
- **VP_DC** Individual victimization. It works combined with Fact_Pers
- **VA_DC** Household victimization. It works combined with Fact_Hog
- **rph_edad** age
- **P3_1_1** Perception of increased crime in the country. It works combined with Fact_Pers
- **P8_1_1** Cause of increased crime in the neighborhood. It works combined with Fact_Pers
- **muj_insg_taxi** Female perception of insecurity inside taxis. Variable elaborated with variables P9 and rph_sexo. It works combined with Fact_Pers
- **hom_insg_taxi** Male perception of insecurity inside taxis. Variable elaborated with variables P9 and rph_sexo. It works combined with Fact_Pers
- **muj_insg_micro** Female perception of insecurity inside buses. Variable elaborated with variables P9 and rph_sexo. It works combined with Fact_Pers
- **hom_insg_micro** Male perception of insecurity inside buses. Variable elaborated with variables P9 and rph_sexo. It works combined with Fact_Pers
- **muj_insg_centr.com** Female perception of insecurity inside malls. Variable elaborated with variables P9 and rph_sexo. It works combined with Fact_Pers
- **hom_insg_centr.com** Male perception of insecurity inside malls. Variable elaborated with variables P9 and rph_sexo. It works combined with Fact_Pers
- **muj_insg_loc.col** Female perception of insecurity public transport. Variable elaborated with variables P9 and rph_sexo. It works combined with Fact_Pers
epf_personas

**hom_insg_loc.col** Male perception of insecurity public transport. Variable elaborated with variables P9 and rph_sexo. It works combined with Fact_Pers

**muj_insg_bario** Female perception of insecurity neighborhood. Variable elaborated with variables P9 and rph_sexo. It works combined with Fact_Pers

**hom_insg_bario** Male perception of insecurity neighborhood. Variable elaborated with variables P9 and rph_sexo. It works combined with Fact_Pers

**Source**

https://www.ine.cl/docs/default-source/seguridad-ciudadana/bbdd/2019/base-de-datos---xvi-enusc-2019-(csv)?sfvrsn=d3465758_2&download=true

**Examples**

```r
data(enusc)
```

**Description**

Reduced version of the VIII EPF database. Contains some sociodemographic variables and the necessary information to work with complex design.

**Usage**

```r
epf_personas
```

**Format**

dataframe compuesto por 48.308 observaciones y 8 variables

- **sexo** 1 = male; 2 = female
- **zona** 1 = metropolitan area; 2 = rest of the regional capitals
- **ecivil** marital status
- **fe** sample weights
- **varunit** PSU
- **varstrat** strata
- **gastot_hd** household expenditure
- **ocupado** 1 = employed; 0 = non-employed

**Source**

https://www.ine.cl/estadisticas/social/ingresos-y-gastos/encuesta-de-presupuestos-familiares
get_cv

Examples

data(epf_personas)

get_cv

Get the coefficient of variation

Description

Receive a table created with survey and return the coefficient of variation for each cell

Usage

get_cv(table, design, domains, type_est = "all", env = parent.frame())

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>table</td>
<td>dataframe with results</td>
</tr>
<tr>
<td>design</td>
<td>design</td>
</tr>
<tr>
<td>domains</td>
<td>vector with domains</td>
</tr>
<tr>
<td>type_est</td>
<td>type of estimation: all or size.</td>
</tr>
<tr>
<td>env</td>
<td>parent environment</td>
</tr>
</tbody>
</table>

Value

dataframe with results including including CV

get_df

Get degrees of freedom

Description

Receive data and domains. Returns a data frame with the psu, strata and df for each cell

Usage

get_df(data, domains, df_type = "eclac")

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>data</td>
<td>dataframe</td>
</tr>
<tr>
<td>domains</td>
<td>string with domains</td>
</tr>
<tr>
<td>df_type</td>
<td>string Use degrees of freedom calculation approach from INE Chile or eclac, by default &quot;ine&quot;.</td>
</tr>
</tbody>
</table>
get_survey_table

Calculates multiple estimations. Internal wrapper for survey package

Description

Generates a table with estimates for a given aggregation

Usage

get_survey_table(
  var,
  domains,
  complex_design,
  estimation = "mean",
  env = parent.frame(),
  fun,
  denom = NULL,
  type_est = "all"
)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>var</td>
<td>string objective variable</td>
</tr>
<tr>
<td>domains</td>
<td>domains</td>
</tr>
<tr>
<td>complex_design</td>
<td>design from survey</td>
</tr>
<tr>
<td>estimation</td>
<td>string indicating if the mean must be calculated</td>
</tr>
<tr>
<td>env</td>
<td>parent environment</td>
</tr>
<tr>
<td>fun</td>
<td>function required regarding the estimation</td>
</tr>
<tr>
<td>denom</td>
<td>denominator. This parameter works for the ratio estimation</td>
</tr>
<tr>
<td>type_est</td>
<td>type of estimation: all or size</td>
</tr>
</tbody>
</table>

Value

dataframe containing main results from survey
**quadratic**

*Calcula el valor de una función cuadrática*

**Description**

*quadratic* returns the output of a particular function created by INE Chile, which is assessed at the value of the estimated proportion from a sample. If the output of the function is higher than the standard error, it is interpreted as a signal that the estimation is not reliable.

**Usage**

```r
quadratic(p)
```

**Arguments**

- `p` numeric vector with the values of the estimations for proportions

**Value**

numeric vector

---

**standardize_columns** *standardize and sort column names*

**Description**

Receive the survey table in raw state and sort it

**Usage**

```r
standardize_columns(data, var, denom)
```

**Arguments**

- `data` dataframe with results
- `var` string with the objective variable
- `denom` denominator

**Value**

dataframe with standardized data
standardize_design_variables

Standardize the name of design variables

Description

Rename design variables, so we can use the later

Usage

standardize_design_variables(design)

Arguments

design  dataframe

Value

design survey
Index

* datasets
  - casen, 3
  - ene, 12
  - enusc, 13
  - epf_personas, 14

assess, 2

casen, 3
create_html, 4
create_mean, 4
create_prop, 6
create_prop_internal, 7
create_ratio_internal, 8
create_size, 9
create_total, 11

eue, 12
enusc, 13
epf_personas, 14

get_cv, 15
get_df, 15
get_survey_table, 16

quadratic, 17

standardize_columns, 17
standardize_design_variables, 18