

# Package: calidad (via r-universe)

October 17, 2024

**Type** Package

**Title** Assesses the Quality of Estimates Made by Complex Sample Designs

**Version** 0.6.0

**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 (2024, [chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://repositorio.cepal.org/server/api/core/bitstreams/f04569e6-4f38-42e7-a32b-e0b298e0ab9c/content](https://repositorio.cepal.org/server/api/core/bitstreams/f04569e6-4f38-42e7-a32b-e0b298e0ab9c/content)).

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.3.2

**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], Andrea López [ctb], Javiera Preuss [ctb]

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

**Repository** CRAN

**Date/Publication** 2024-10-16 19:40:02 UTC

## Contents

assess . . . . .	2
casen . . . . .	4
create_html . . . . .	5
create_mean . . . . .	5
create_prop . . . . .	7
create_prop_internal . . . . .	8
create_ratio_internal . . . . .	9
create_size . . . . .	10
create_total . . . . .	12
eclac_standard . . . . .	13
ene . . . . .	14
enusc . . . . .	15
epf_personas . . . . .	16
get_cv . . . . .	17
get_df . . . . .	17
get_survey_table . . . . .	18
quadratic . . . . .	19
standardize_columns . . . . .	19
standardize_design_variables . . . . .	20
<b>Index</b>	<b>21</b>

---

assess	<i>Assess the quality of mean estimations</i>
--------	---

---

### Description

assess evaluates the quality of mean estimations 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_2020", "eclac_2023"),
  domain_info = FALSE,
  ...
)
```

### Arguments

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

scheme	character variable indicating the evaluation protocol to use. Options are "chile", "eclac_2020", "eclac_2023".
domain_info	Logical. If TRUE, indicates that the study domain information is available and will be used for assessment. This affects how the evaluation is conducted, leveraging specific domain-level data to refine the assessment results. When FALSE, domain-specific adjustments are omitted, and a generalized assessment is performed.
...	<p>additional parameters for the evaluation. The complete list of parameters is:</p> <ol style="list-style-type: none"> <li>1. General Parameters <ul style="list-style-type: none"> <li>• df degrees of freedom. Default: 9.</li> <li>• n sample size. Default for chile scheme: 60. Default for CEPAL schemes: 100.</li> </ul> </li> <li>2. chile Parameters <ul style="list-style-type: none"> <li>• cv_lower_ine lower limit for CV. Default: 0.15.</li> <li>• cv_upper_ine upper limit for CV. Default: 0.3.</li> </ul> </li> <li>3. CEPAL 2020 Parameters <ul style="list-style-type: none"> <li>• cv_cepal limit for CV. Default: 0.2.</li> <li>• ess effective sample size. Default: 140.</li> <li>• unweighted unweighted count. Default: 50.</li> <li>• log_cv logarithmic coefficient of variation. Default: 0.175.</li> </ul> </li> <li>4. CEPAL 2023 Parameters <ul style="list-style-type: none"> <li>• cv_lower_cepal lower limit for CV. Default: 0.2.</li> <li>• cv_upper_cepal upper limit for CV. Default: 0.3.</li> <li>• ess effective sample size. Default: 60.</li> <li>• cvlog_max maximum logarithmic coefficient of variation. Default: 0.175.</li> <li>• CCNP_b unweighted count before adjustment. Default: 50.</li> <li>• CCNP_a unweighted count after adjustment. Default: 30.</li> </ul> </li> </ol>

**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	<i>Encuesta de Caracterización Socioeconómica Nacional 2020 - CASEN en Pandemia 2020</i>
-------	--

---

### 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

<http://observatorio.ministeriodesarrollosocial.gob.cl/encuesta-casen-en-pandemia-2020>

### Examples

```
data(casen)
```

---

create_html	<i>Create html table with the results of the evaluation</i>
-------------	---

---

**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	<i>Create the inputs to evaluate the quality of mean estimations</i>
-------------	--

---

**Description**

create\_mean generates a 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,
```

```

    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 another function, if TRUE avoid lazy eval errors.
rm.na	boolean remove NA values if required.
deff	boolean design effect.
rel_error	boolean relative error.
unweighted	boolean add non-weighted count if 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	<i>Create the inputs to evaluate the quality of proportion estimations</i>
-------------	--

---

### 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 required.

standard\_eval    boolean indicating if the function is wrapped inside another 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,
```



```

    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,

```

```

    ci = FALSE,
    deff = FALSE,
    ess = FALSE,
    ajuste_ene = FALSE,
    unweighted = FALSE,
    rel_error = FALSE,
    rm.na = 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.
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
unweighted	boolean Add non weighted count if it is required
rel_error	boolean Relative error
rm.na	boolean indicating if NA values must be removed

### Value

dataframe that contains the inputs and all domains to be evaluated

---

create\_size

*Create the inputs to evaluate the quality of total estimations*

---

### 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.

**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("chile", "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 another function, if TRUE avoid lazy eval errors.
rm.na	boolean remove NA values if required.
deff	boolean design effect.
rel_error	boolean relative error.
unweighted	boolean add non-weighted count if required.
df_type	character use degrees of freedom calculation approach from INE Chile or CEPAL. Options are "chile" or "eclac".
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_size("ocupado", "zona+sexo", design = dc)
```

---

create_total	<i>Create the inputs to evaluate the quality of the sum of continuous variables</i>
--------------	---

---

**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 another function, if TRUE avoid lazy eval errors.
rm.na	boolean remove NA values if required.

deff	boolean design effect.
rel_error	boolean relative error.
unweighted	boolean add non-weighted count if 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)
```

---

eclac_standard	<i>Turn on all the indicators needed for the eclac standard</i>
----------------	---

---

**Description**

This function activates the appropriate indicators based on the selected eclac standard and whether proportion indicators are needed.

**Usage**

```
eclac_standard(eclac, env = parent.frame(), proportion = FALSE)
```

**Arguments**

eclac	A logical value indicating the eclac standard.
env	The environment from which to retrieve the existing indicator values. Defaults to the parent frame.
proportion	A logical value indicating whether proportion indicators should be turned on. Defaults to FALSE.

**Value**

A list of logical values indicating which indicators are turned on.

---

ene

*Encuesta Nacional de Empleo - ENE. 2020-efm*

---

### 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)

---

enusc	<i>Encuesta Nacional Urbana de Seguridad ciudadana 2019 - ENUSC 2019</i>
-------	--

---

### Description

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

### Usage

enusc

### Format

dataframe with 24.465 rows y 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\_cent.com** Female perception of insecurity inside malls. Variable elaborated with variables P9 and rph\_sexo. It works combined with Fact\_Pers

**hom\_insg\_cent.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

**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\_barrio** Female perception of insecurity neighborhood. Variable elaborated with variables P9 and rph\_sexo. It works combined with Fact\_Pers

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

### Examples

```
data(enusc)
```

---

epf\_personas

*VIII Encuesta de Presupuestos Familiares*

---

### Description

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

### Usage

```
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/sociales/ingresos-y-gastos/encuesta-de-presupuestos-familiares>



**Examples**

```
data(epf_personas)
```

---

get_cv	<i>Get the coefficient of variation</i>
--------	---

---

**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	dataframe with results
design	design
domains	vector with domains
type_est	type of estimation: all or size.
env	parent environment

**Value**

dataframe with results including including CV

---

get_df	<i>Get degrees of freedom</i>
--------	-------------------------------

---

**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**

data	dataframe
domains	string with domains
df_type	string Use degrees of freedom calculation approach from INE Chile or eclac, by default "chile".

**Value**

dataframe with results including degrees of freedom

---

get_survey_table	<i>Calculates multiple estimations. Internal wrapper for survey package</i>
------------------	---

---

**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**

var	string objective variable
domains	domains
complex_design	design from survey
estimation	string indicating if the mean must be calculated
env	parent environment
fun	function required regarding the estimation
denom	denominator. This parameter works for the ratio estimation
type_est	type of estimation: all or size

**Value**

dataframe containing main results from survey

---

quadratic	<i>Calcula el valor de una función cuadrática</i>
-----------	---

---

**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**

```
quadratic(p)
```

**Arguments**

p numeric vector with the values of the estimations for proportions

**Value**

numeric vector

---

standardize_columns	<i>standardize and sort column names</i>
---------------------	--

---

**Description**

Receive the survey table in raw state and sort it

**Usage**

```
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, 4
- ene, 14
- enusc, 15
- epf\_personas, 16

assess, 2

casen, 4  
create\_html, 5  
create\_mean, 5  
create\_prop, 7  
create\_prop\_internal, 8  
create\_ratio\_internal, 9  
create\_size, 10  
create\_total, 12

eclac\_standard, 13  
ene, 14  
enusc, 15  
epf\_personas, 16

get\_cv, 17  
get\_df, 17  
get\_survey\_table, 18

quadratic, 19

standardize\_columns, 19  
standardize\_design\_variables, 20