



**International release R Package "calidad"**

**Methodologies and Data Science for Statistical Production**

**INE Chile**

**December 2022**

**Estándar de calidad en  
encuestas de hogares  
ONE Chile**

**Quality criteria for  
household surveys ONE  
Chile**

**Antes de empezar...**

**Before starting...**

# ¿Qué significa calidad estadística?

## What is the meaning of statistical quality?



Miguelito: According to the survey, the average salary of women in Arica is 600,000 pesos.

Mafalda: Yes, but did you know that the interval is between 300,000 and 900,000?.



The only thing I know is that I don't know anything

En 2020 el INE publica un estándar de calidad para las estimaciones con **encuestas de hogares**

In 2020 INE Chile published its statistical quality standard for **household surveys** estimations



## Fundamentos del Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares

Documento de trabajo - Metodológico

En este documento se presenta el trabajo metodológico realizado por el INE en relación con la evaluación de las medidas de calidad de las estimaciones provenientes de encuestas a hogares por muestreo, con el propósito de proveer al Sistema Estadístico Nacional (SEN) de un conjunto de directrices que permitan orientar a la población usuaria respecto a la evaluación de la calidad, uso, análisis e interpretación de la información que produce el INE.

Bajo un marco de aseguramiento de la calidad, la importancia de contar con lineamientos para la evaluación de las estimaciones a través de criterios que consideren diferentes dimensiones (como el tamaño muestral, grados de libertad, coeficiente de variación, error estándar, entre otros) radica en que la fiabilidad estadística, entendida como el grado en que las mediciones obtenidas reflejan la realidad, permite definir cuándo una estimación cumple un requisito mínimo de calidad. En este sentido, el documento entrega un panorama respecto al marco conceptual de las medidas de calidad, su uso y criterios utilizados en las encuestas del INE, así como también en diversas Oficinas Nacionales de Estadística (ONE). De acuerdo con los antecedentes expuestos, se desarrolla una propuesta de lineamientos a través de flujogramas que orienta al conjunto de usuarios en la toma de decisiones al momento de analizar y publicar información.

### Información

Fecha:  
Marzo 2020

[Descargar Documento](#)

**Cuadro estadístico:** arreglo ordenado de datos procesados para facilitar la lectura e interpretación

**Statistical table:** ordered arrangement of data to facilitate reading and interpretation"

## *Ejemplo ilustrativo : Personas microemprendedoras por sexo según categoría ocupacional*

Categoría ocupacional	N° personas micro emprendedores			Distribución			Concentración		
	Total	Hombres	Mujeres	Total	Hombres	Mujeres	Total	Hombres	Mujeres
Total	53.317	30.278	23.040	100	56,79	43,21	100	100	100
Empleador	10.729	6.736	3.994	100	62,78	37,23	20,12	22,25	17,34
Cuenta propia	42.588	23.542	19.046	100	55,28	44,72	79,88	77,75	82,66

Fuente: INE, Quinta Encuesta de Microemprendimiento, 2017

- **Tamaño muestral (tm):** unidades de análisis que nutren las estimaciones (viviendas, hogares y/o personas)
- **Grados de libertad (gl)**
  - Tratamiento diferenciado para las **proporciones y razones definidas entre 0 y 1**

$df_d(1)=$

#UPM con observaciones en la subpoblación  
menos  
#Estratos con observaciones en la subpoblación

- **Coeficiente de variación / error estándar**
  - Proporción y razones definidas entre 0 y 1: **SE**
  - Resto: **CV**

- **Sample size:** units of analysis that contribute to estimations (dwellings, households or people)
- **Degrees of freedom**
  - Different procedure for **Proportions and ratios between 0 and 1**

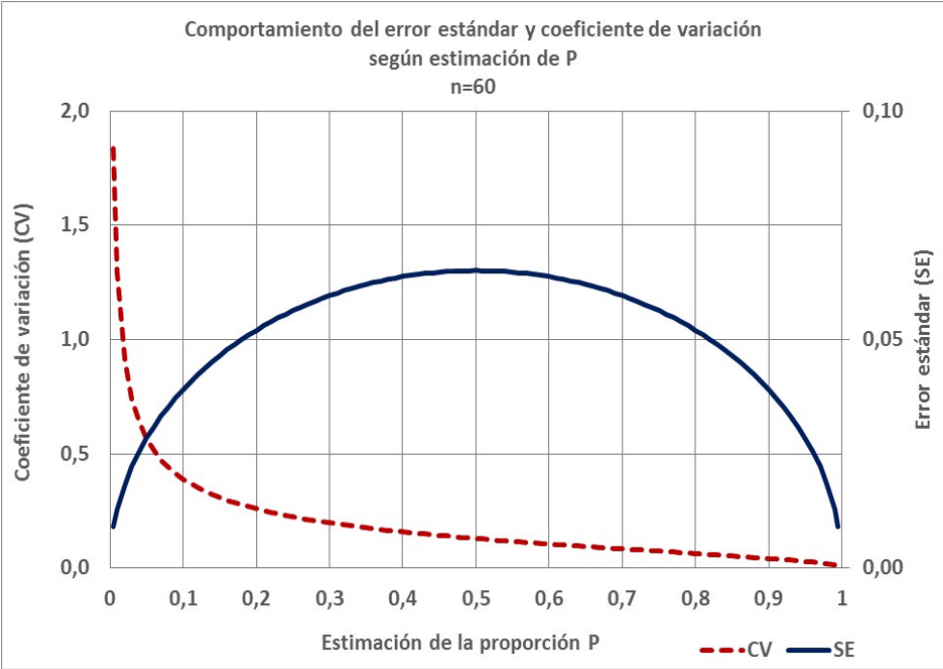
$df_d(1)=$

#UPM con observaciones en la subpoblación  
menos  
#Estratos con observaciones en la subpoblación

- **Coefficient of variation / standard error**
  - Proportions and ratios between 0 and 1: **SE**
  - All others: **CV**

# Dicotomía de las proporciones

## Dicothomy of proportions



P	Error estándar	CV
0,1	0,039	0,391
0,9	0,039	0,043

Es ilógico concluir que la estimación de  $p$  no tiene una calidad aceptable pero la de  $1-p$  sí la tiene.

It makes no sense to conclude that the quality of estimate of  $p$  is not acceptable but that estimate of  $1-p$  is acceptable.



## Indicadores de calidad

- Proporciones y razones definidas entre 0 y 1:  
**Error estándar**
- Otras: Coeficiente de variación

## Quality criteria

- Proportions and ratios between 0 and 1:  
**Standard error**
- All others: Coefficient of variation

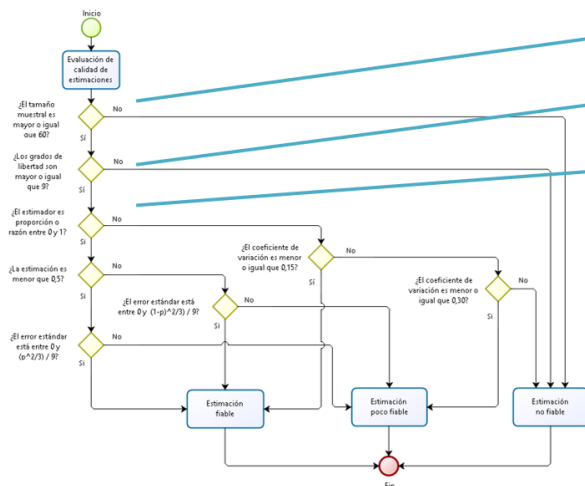
# Estándar INE Chile (INE Chile standard)

## Primera etapa de aplicación de estándar

## First step in quality assessment

Flujograma para evaluación de calidad de las estimaciones

Flowchart for quality assessment



Tamaño muestral  $\in \{(0,60); [60, \infty)\}$   
(Tamaño de muestra efectivo)  $\in \{(0,40); (40; \infty)\}$

Grados de libertad  $\in \{(0,9]; (9; \infty)\}$

Tipo de estimador: Proporción o Razón definida entre 0 y 1

$p < 0,5$ :  $ee \in \left(0, \sqrt[3]{p^2}/9\right]; \left(\sqrt[3]{p^2}/9; \infty\right\}$

$p \geq 0,5$ :  $ee \in \left(0, \sqrt[3]{(1-p)^2}/9\right]; \left(\sqrt[3]{(1-p)^2}/9; \infty\right\}$

CV  $\in \{(0; 0,15]; (0,15; 0,30]; (0,30; \infty)\}$

Fiable

Poco fiable

No fiable

Categories: Reliable, weakly reliable and non-reliable

# Introducción paquete calidad (introduction)

¿Cómo llevar a la práctica los conceptos de calidad?

¿How to put the quality concepts into practice?



- Existen múltiples herramientas (Stata, R, SAS, Python) y todas son válidas
- Una posibilidad es el uso de un **paquete (librería)**
- El paquete **calidad** implementa el estándar mediante **R**

- There are many tools (Stata, R, SAS, Python) and all of them are valid
- One possibility is the use of a **package (library)**
- The **calidad** package implements the criteria using **R**

## Objetivos del paquete

- Facilitar la aplicación del estándar a usuarios externos
- Aumentar la eficiencia de los analistas de datos
- Reducir la probabilidad de errores en la implementación

## Objectives of the package

- Make it easier for external users to apply the standard
- Increase data analysts' efficiency
- Reduce the probability of implementation errors

# Introducción paquete calidad (introduction)

El paquete `calidad` combina conceptos de calidad con el paquete `survey`, desarrollado por Thomas Lumley

`calidad` combines quality concepts with the `survey` package, developed by Thomas Lumley

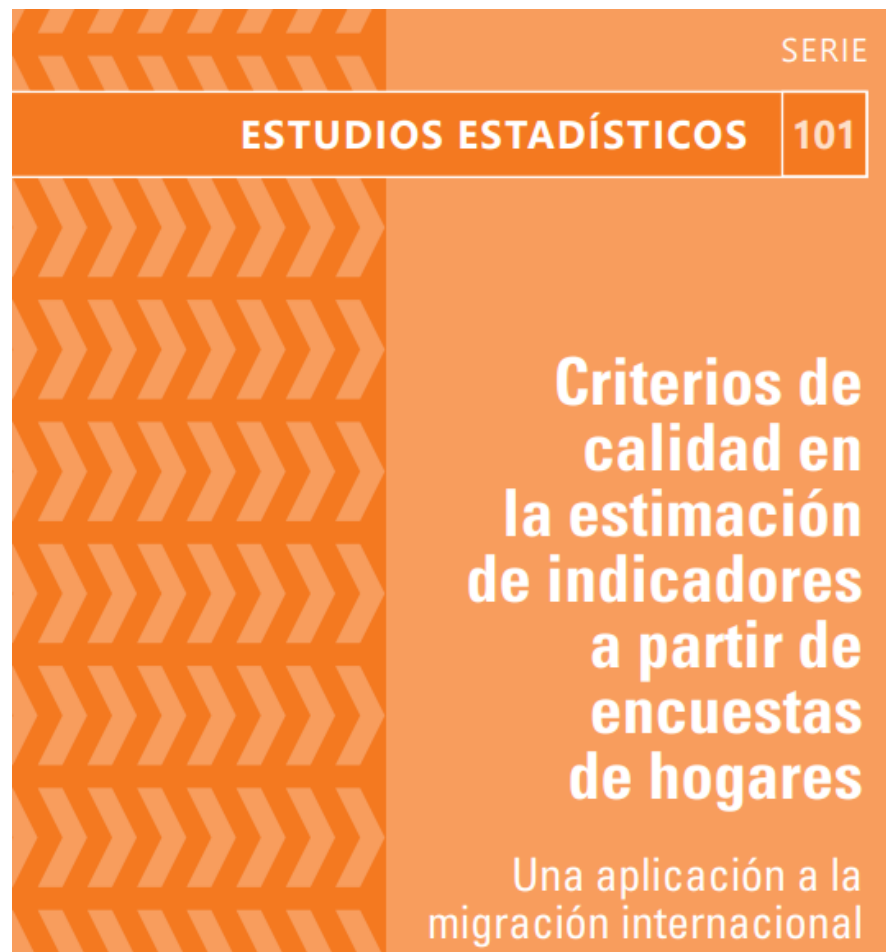


**Desde hace un año el paquete está en github y hace algunas semanas lo subimos a CRAN**

- media
- proporciones
- ~~mediana (con réplicas)~~
- tamaños
- totales

**The package has been available for a year. Two months ago, we uploaded it to CRAN**

- mean
- proportions
- ~~median (replicates)~~
- sizes
- totals



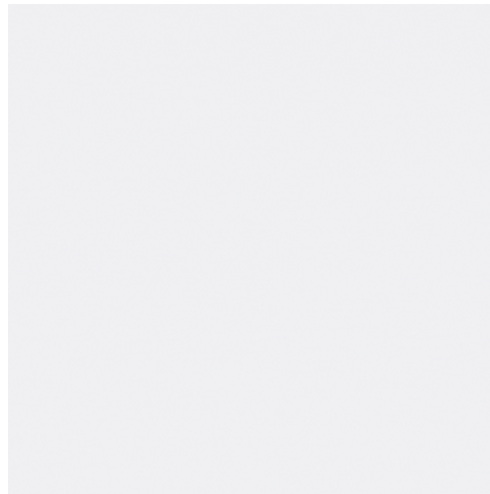
- **El estándar CEPAL considera:**

- coeficiente de variación
- **coeficiente de variación logarítmico**
- tamaño de muestra
- **tamaño de muestra efectivo**
- **conteo de casos no ponderado**
- grados de libertad

- **El estándar CEPAL considera:**

- coefficient of variation
- **logarithmic coefficient of variation**
- sample size
- **effective sample size**
- **non-weighted units**
- degrees of freedom

## New functionalities



## New developments with ECLAC

**Nuevos indicadores de calidad**

**Flexibilización de umbrales**

**Alinear nombres con la teoría**

**New quality indicators**

**Flexible thresholds**

**Names aligned with theory**



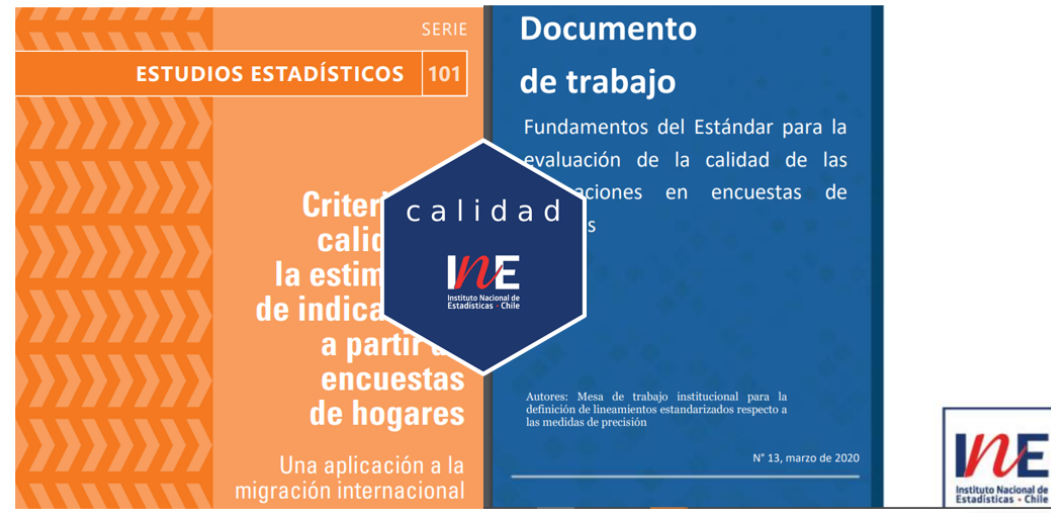
# Estándares combinados (combined standards)

La idea es generar una herramienta que:

- Implemente los dos estándares
- Ofrezca flexibilidad a los usuarios

The idea was to create a tool capable of:

- Implementing both standards
- Offering flexibility to users



- Problemas similares en ONEs y organismos públicos
- Herramientas open source pueden ayudar
- Empoderamiento de la comunidad ligada a las estadísticas oficiales
- En Europa existe una comunidad desde hace algunos años
  - [uRos](#) (ligada a R)
  - [Awesome official statistics software](#) (R, python, Java, C++)
- En América Latina la comunidad es aún incipiente
- Desde hace poco existe un repositorio llamado ["Asombrosos paquetes de R en Latinoamérica"](#)
- Similar problems in ONEs and government organizations
- Open source tools can be very helpful
- Empowerment of the community linked to official statistics
- In Europe there has been a community for some years
  - [uRos](#) (related to R)
  - [Awesome official statistics software](#) (R, python, Java, C++)
- In Latin America the community is still incipient
- Recently there is a repository called ["Amazing R packages in Latin America"](#)

Las dinámicas open source son ideales para la colaboración a distancia

Las oficinas de estadística pueden trabajar unidas para resolver problemas comunes

Open source dynamics are ideal for remote collaboration

Statistical offices can work together to solve common problems



**International release R Package "calidad"**

**Methodologies and Data Science for Statistical Production**

**INE Chile**

**December 2022**