

# Paquete Calidad, para la evaluación de la precisión de estimaciones provenientes de encuestas de hogares.

*Anónimo*

**Abstract** En base al estándar de calidad publicado por el Instituto Nacional de Estadísticas de Chile (INE), se ha desarrollado el paquete “Calidad”, para la implementación de dicho estándar para la evaluación de la precisión de las estimaciones provenientes de encuestas de hogares.

**Palabras clave:** encuestas - hogares - calidad estadística

**Modalidad presentación:** Comunicación Oral

## Origen

El desarrollo del paquete “Calidad”, se da en el marco del estándar de calidad estadística definido por el Instituto Nacional de Estadísticas de Chile (INE), para determinar qué tan precisa y confiable es la información de encuestas de hogares que se publica, respecto a los atributos de la población que se pretende caracterizar.<sup>1</sup>

La relevancia de este estándar y su implementación permitirá a la oficina de estadísticas de Chile estandarizar sus rutinas de control de calidad de los datos publicados y ofrecer un modelo de control de calidad de las estimaciones, que además, puede ser utilizado por otras instituciones tanto a nivel nacional como internacional.

El objetivo inicial del paquete es entregar a los usuarios de las estadísticas oficiales, tanto de las instituciones gubernamentales como investigadores y estudiantes, una herramienta de código abierto y sencilla utilización, que permita implementar de manera práctica los criterios de calidad estadística definidos.

Para lograr este objetivo y alineado a la política institucional del INE Chile, de impulsar el trabajo con R como principal herramienta para el procesamiento y análisis de datos, se han construido en entorno de R los criterios de calidad propuestos por el estándar.

Este paquete permite el trabajo con diversas encuestas de hogares de relevancia nacional como:

- La Encuesta de Caracterización Socioeconómica Nacional (CASEN)
- La Encuesta Nacional de Empleo (ENE)
- La Encuesta Suplementaria de Ingresos (ESI)
- La Encuesta de Presupuestos Familiares (EPF)
- La Encuesta Nacional Urbana de Seguridad Ciudadana (ENUSC)

Entre otras encuestas con diseño estadístico multietápico, no solo de Chile.

Es relevante mencionar que el paquete “Calidad” usa en su base el paquete “Survey” de Thomas Lumley, utilizado para el trabajo con encuestas con diseño complejo.

## Principales características

“Calidad” posee dos familias de funciones:

- `create_`: permiten crear los insumos para la evaluación del estándar
- `evaluate_`: permiten ejecutar la evaluación del estándar

Con las familia “create” Podemos hacer los siguientes cálculos:

1. `calidad::create_mean`: calcular la media (ingreso)
2. `calidad::create_prop`: proporción o razón (ocupación)
3. `calidad::create_tot`: conteo de unidades (ocupación)
4. `calidad::create_tot_con`: suma de variables continuas (ingreso)
5. `calidad::create_median`: mediana (ingreso)

<sup>1</sup><https://www.ine.cl/docs/default-source/institucionalidad/buenas-pr%C3%A1cticas/clasificaciones-y-est%C3%A1ndar-evaluaci%C3%B3n-de-calidad-de-estimaciones-publicaci%C3%B3n-27022020.pdf>

## A continuación un ejemplo, calculamos la media de la edad en la Encuesta Nacional Urbana de Seguridad Ciudadana 2019 (ENUSC):

```
library(survey); library(calidad)

dc <- svydesign(ids = ~Conglomerado, strata = ~VarStrat, weights = ~Fact_Pers, data = enusc)
options(survey.lonely.psu = "certainty")

calidad::create_mean(rph_edad, disenio = dc)

##              mean          se    gl      n    coef_var
## rph_edad 44.27603 0.2804819 486 24465 0.006334848
```

Como se puede observar, adicionalmente a la estimación se nos entrega el Error Estandar (SE), los Grados de Libertad (GL), la cantidad de casos muestrales (n) y el coeficiente de variación (coef\_var). Estos son los insumos que serán evaluados con las funciones de la siguiente familia

### Los argumentos recibidos por estas funciones son:

- var** = Variable de interés, ej: desocupación.
- denominador**<sup>2</sup> = denominador de la división en estimaciones de razón, ej: mujeres ocupadas / hombres ocupados.
- dominios** = variables de desagregación, ej: sexo+region.
- subpop** = Posibles variables de sub poblaciones, fuerza de trabajo.
- disenio** = diseño complejo generado con el paquete survey.
- ci** = para obtener intervalos de confianza.

Como ya se mencionó con la segunda familia de funciones se pueden evaluar los insumos creados. Existe una función de evaluación para cada tipo de estimación:

1. **calidad::evaluate\_mean**
2. **calidad::evaluate\_prop**
3. **calidad::evaluate\_tot**
4. **calidad::evaluate\_tot\_con**
5. **calidad::evaluate\_median**

Estas funciones se aplican sobre los insumos generados por las funciones *create*, obtenemos 3 columnas nuevas de evaluación y una cuarta columna que nos advierte si la estimación es *fiable*, *poco fiable* o *no fiable*.

### A continuación un ejemplo de evaluación, evaluando un estimador de media con “evaluate\_mean”:

```
calidad::evaluate_mean(calidad::create_mean(rph_edad, disenio = dc))

##              mean          se    gl      n    coef_var    eval_n    eval_gl
## rph_edad 44.27603 0.2804819 486 24465 0.006334848 n suficiente gl suficiente
##              eval_cv calidad
## rph_edad cv <= 15   fiable
```

Obtenemos 3 columnas nuevas de evaluación y una cuarta columna que nos advierte si la estimación es fiable.

### Próximos pasos

Actualmente el paquete se encuentra en el desarrollo de nuevas funcionalidades, orientadas a flexibilizar los análisis según las preferencias de los investigadores, así como para lograr ser útil a nivel Latinoamericano, para esto se está colaborando con la Comisión Económica Para Latinoamérica (CEPAL) en los ajustes y funcionalidades necesarias. Esperamos que “calidad” sea útil para diferentes oficinas nacionales de estadísticas y usuarios independientes.

<sup>2</sup>disponible solo para estimadores de razón