

# CEMRepBoot 1.0

Cálculo de Errores Muestrales por Réplicas Bootstrap

## Manual de uso

NOTAS  
TÉCNICAS  
INDEC  
Nº 6



Ministerio de Economía  
Argentina

Instituto Nacional de  
Estadística y Censos  
República Argentina

indec

**Cálculo de Errores Muestrales por Réplicas Bootstrap  
(CEMRepBoot 1.0). Manual de uso  
Notas Técnicas INDEC n°6  
Noviembre de 2020**

Instituto Nacional de Estadística y Censos (INDEC)

**Dirección:** Marco Lavagna

**Dirección Técnica:** Pedro Ignacio Lines

**Dirección Nacional de Difusión y Comunicación:** María Silvina Viazzi

**Coordinación de Producción Gráfica y Editorial:** Marcelo Costanzo

Esta publicación fue realizada por los equipos técnicos de la Dirección Nacional de Metodología Estadística y de la Coordinación de Muestreo.

**Ficha técnica**

**Versión:** 1.0

**Software:** R 4.0.2

**Paquetes requeridos para el uso:** shiny, shinyalert, DT, shinydashboard, shinybusy, openxlsx, survey

**Autor:** Dirección Nacional de Metodología Estadística, Coordinación de Muestreo

**Desarrolladores:** Emanuel Ciardullo y Gregorio García

**Codificación:** ISO-8859-1

**Última actualización:** noviembre de 2020

ISSN 2683-8478

ISBN 978-950-896-663-6

Instituto Nacional de Estadística y Censos

Cálculo de errores muestrales por réplicas bootstrap : CEMRepBoot 1.0 : manual de uso / 1a ed. - Ciudad Autónoma de

Buenos Aires : Instituto Nacional de Estadística y Censos - INDEC, 2024.

Libro digital, PDF - (Notas Técnicas INDEC ; 6)

Archivo Digital: descarga y online

ISBN 978-950-896-663-6

1. Estadísticas. 2. Manual Técnico. 3. Metodología de la Investigación. I. Título.

CDD 310.2



Libro de edición argentina

Buenos Aires, enero de 2024

**Publicaciones del INDEC**

Las publicaciones editadas por el Instituto Nacional de Estadística y Censos están disponibles en [www.indec.gob.ar](http://www.indec.gob.ar) y en el Centro Estadístico de Servicios, ubicado en Av. Presidente Julio A. Roca 609 C1067ABB, Ciudad Autónoma de Buenos Aires, Argentina. También pueden solicitarse al teléfono +54 11 51031-4632 en el horario de atención al público de 9:30 a 16:00. Correo electrónico: [ces@indec.gob.ar](mailto:ces@indec.gob.ar)

Calendario anual anticipado de informes: [www.indec.gob.ar/indec/web/Calendario-Fecha-0](http://www.indec.gob.ar/indec/web/Calendario-Fecha-0)



INDECArgentina

# Índice

1. Introducción .....	4
2. Ejecución de la aplicación .....	5
2.1 Inicio de la aplicación desde RStudio .....	6
2.2 Inicio de la aplicación desde R .....	7
3. Uso de la aplicación y ejemplos .....	13
3.1 Descarga de las bases desde la página web del INDEC .....	14
3.2 Carga de las bases desde un archivo local .....	14
4. Selección del indicador a calcular para las estimaciones .....	16
4.1 Estimación de una proporción .....	16
4.2 Estimación de un total .....	19
4.3 Estimación de una media .....	21
4.4 Estimación de una razón .....	22
5. Estimación en subpoblaciones o dominios .....	23
6. Descarga de resultados .....	26
7. Recomendaciones para interpretar las estimaciones .....	26
8. Tratamientos especiales: ENGHo Gastos .....	28

# 1. Introducción

Con el objetivo de brindar asistencia a los usuarios, el INDEC pone a disposición la herramienta de cálculo CEMRepBoot<sup>1</sup>. Esta recurre al paquete shiny del programa R, y facilita a las personas usuarias de R la posibilidad de realizar estimaciones de diversos indicadores con sus medidas de error e intervalos de confianza a partir de las bases de datos con microdatos de las encuestas. A la fecha de este manual, CEMRepBoot permite operar con: la Encuesta de Actividades de Niños, Niñas y Adolescentes 2016-2017 (EANNA), el Estudio Nacional sobre el Perfil de las Personas con Discapacidad 2018 (ENDI), la 4° Encuesta Nacional de Factores de Riesgo (4° ENFR) y la Encuesta Nacional de Gastos de los Hogares 2017-2018 (ENGHo); todas ellas son publicadas por el Instituto Nacional de Estadística y Censos (INDEC) y sus bases, compartidas en el sitio web.

Estas encuestas tienen la particularidad de utilizar un diseño muestral complejo, es decir, involucran diferentes niveles de estratificación, varias etapas de selección y probabilidades no uniformes en cada una de ellas. Además, los factores de expansión surgidos del diseño son ajustados por elegibilidad, no respuesta y calibración.

Estos diseños y ajustes, si bien resultan en un uso más eficiente de los recursos del Instituto, traen aparejadas complicaciones al momento de estimar indicadores y en particular las medidas del error de muestreo asociadas.

Por tal motivo, el INDEC define para sus encuestas una metodología del cálculo del error muestral a partir de réplicas de los factores de expansión; documentada en las notas técnicas<sup>2</sup> que acompañan los resultados y a las bases de microdatos que comparte el INDEC a los usuarios.

Una característica de la metodología adoptada es que cada base con microdatos de las encuestas es acompañada por una base de réplicas de los factores de expansión. El vínculo de estas bases y la aplicación de la metodología que propone el Instituto permiten a calcular las medidas de calidad o bondad de las estimaciones sin tener que recurrir a publicar tablas de errores u otra alternativa. Las personas usuarias también pueden reproducir la información publicada tanto de las estimaciones como de los indicadores de calidad asociados al error muestral; o bien hacerlo sobre otras estimaciones que el INDEC no publica y la base microdatos habilita.

La herramienta que se describe en este manual facilita el uso de los factores de expansión y las réplicas que acompañan a los datos, ya que automatiza la unión de las bases de datos y permite que la persona usuaria realice estimaciones, visualice y descargue resultados mediante una interfaz gráfica.

Quien desee utilizar otras herramientas para realizar análisis de las bases de datos publicadas puede seguir las instrucciones que se encuentran en las notas técnicas. Estas describen cómo realizar las estimaciones en los programas R, SAS, Stata y Wesvar.

<sup>1</sup> Acrónimo de “Cálculo de Errores Muestrales por Replicaciones Bootstrap”.

<sup>2</sup> Para más detalles consultar los enlaces de las notas técnicas en <https://www.indec.gob.ar/indec/web/Institucional-Indec-Metodologias-2>.

Si bien la herramienta brinda facilidades a la hora de estimar indicadores y sus errores muestrales, posee una serie de limitaciones; entre las más relevantes:

- Solo es posible emplear la herramienta en las encuestas mencionadas y publicadas por el Instituto en los párrafos anteriores.
- Solo es posible estimar parámetros poblacionales del tipo proporciones, totales, medias y razones. Estimadores más complejos como los asociados a estadísticos de orden (medianas, cuantiles, etc.), de medidas de desigualdad (como brechas, coeficiente de Gini, etc.) u otros no están disponibles en la aplicación por el momento.
- No es posible realizar una recodificación o un agrupamiento de las variables de las bases microdatos de las encuestas dentro de la herramienta; si la persona usuaria desea reconvertir o crear una variable a partir de las existentes, deberá hacerlo por afuera del CEMRepBoot.

## 2. Ejecución de la aplicación

**Importante:** para que pueda ejecutarse correctamente, la aplicación requiere conexión a internet, o haber descargado previamente las bases de microdatos y de réplicas de la encuesta.

La aplicación está contenida en una carpeta llamada “CEMRepBoot” en donde se encontrará un archivo .R y varias carpetas (ver sección 2.1). Estas carpetas contienen recursos multimedia que permiten la correcta visualización de la aplicación y el código R que habilita su correcto funcionamiento.

Para ejecutar la aplicación, la persona usuaria deberá tener instalados los programas R y los siguientes paquetes (*package* dentro del entorno R):

- shiny
- shinydashboard
- shinybusy
- shinyalert
- openxlsx
- DT
- survey

En el caso de que utilice RStudio, puede ejecutarse la aplicación desde esta interfaz más fácilmente que mediante la consola de R.

Todos ellos pueden instalarse ejecutando la sentencia *install.packages("xxxxxx")* (ver sección 2.2) escribiendo el nombre del paquete a instalar entre comillas, respetando las mayúsculas y minúsculas como se muestran en la lista anterior.

Si la primera ejecución de la aplicación se realiza desde RStudio, los paquetes se instalarán de forma automática y por única vez (ver sección 2.1). En cambio, si la persona usuaria ejecuta CEMRepBoot por primera vez desde R, deberá instalar de forma manual el paquete shiny para poder abrir la aplicación (ver sección 2.2).

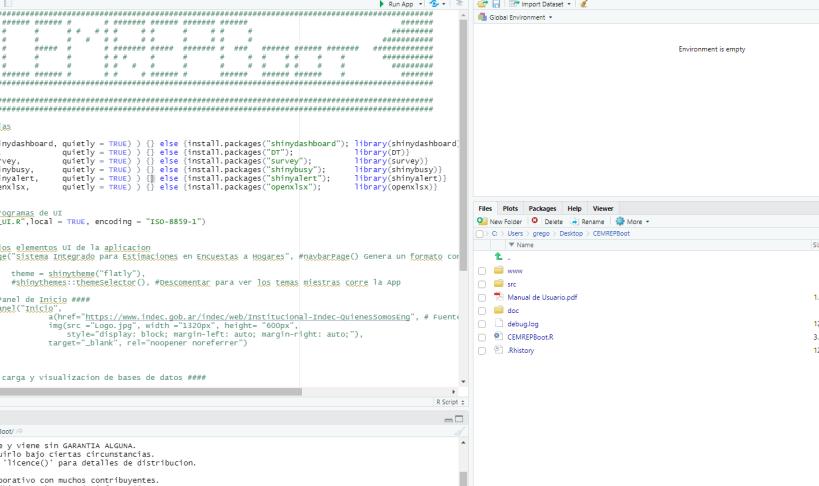
## 2.1 Inicio de la aplicación desde RStudio

Para iniciar la aplicación se debe ejecutar el archivo CEMRepBoot.R incluido en la carpeta que lo contiene desde RStudio empleando el botón Run App.

**Figura 1. Carpeta de la aplicación**

CEMREPBoot				
	Nombre	Fecha de modificación	Tipo	Tamaño
	doc	11/11/2020 19:56	Carpetas de archivos	
	src	11/11/2020 19:56	Carpetas de archivos	
	www	11/11/2020 19:56	Carpetas de archivos	
	CEMREPBoot.R	11/11/2020 19:56	Archivo R	4 KB
	Manual de Usuario.pdf	11/11/2020 19:56	Foxit Reader PDF ...	1.932 KB

Figura 2. Archivo CEMRepBoot.R ejecutado en RStudio

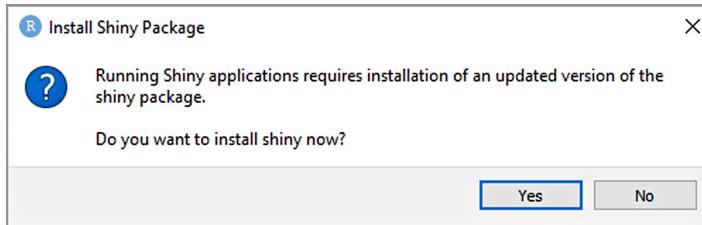


The screenshot shows the RStudio interface with the following details:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Run Tab:** Shows the R script `CEM4RBoot.R` with syntax highlighting and line numbers. The script includes imports for shiny, shinydashboard, DT, shinybusy, shinyalert, and openxlsx, and sets up a UI with a sidebar and a main panel containing a logo and a table.
- Environment Tab:** Shows the global environment is empty.
- Console Tab:** Displays the output of the R script execution, including the creation of the `CEM4RBoot` package and its contents.
- File Explorer:** Shows the project structure with files like `src`, `www`, `src/Manual de Usuario.pdf`, `debug.log`, and `Rhpository`.

Si es la primera vez que la persona usuaria ejecuta la aplicación, se abrirá un cuadro de diálogo que le preguntará si quiere instalar el paquete shiny. Si se presiona “Yes”, RStudio procederá a realizar la instalación (el proceso puede tardar unos minutos dependiendo de la velocidad de conexión a internet).

**Figura 3. Instalación del paquete shiny**



**Figura 4. Instalación de los paquetes necesarios**

```
Install 'shiny'  
Installing 'shiny' ...  
trying URL 'https://cran.rstudio.com/src/contrib/shiny_1.5.0.tar.gz'  
Content type 'application/x-gzip' length 3472148 bytes (3.3 MB)  
=====  
downloaded 3.3 MB  
  
* installing *source* package 'shiny' ...  
** package 'shiny' successfully unpacked and MD5 sums checked  
** using staged installation  
** R  
** inst  
** byte-compile and prepare package for lazy loading  
** help  
*** installing help indices  
*** copying figures  
** building package indices  
** testing if installed package can be loaded from temporary location
```

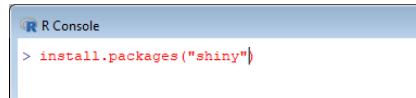
Una vez finalizada la instalación, se ejecutará CEMRepBoot automáticamente. En caso de no ejecutarse al finalizar la instalación, puede ser necesario reiniciar RStudio para que reconozca los paquetes instalados.

## 2.2 Inicio de la aplicación desde R

Es posible iniciar la aplicación desde la consola o línea de comandos de R; esta modalidad no es una dificultad para quienes usan R, pero puede que lo sea para los que no están habituados o familiarizados con ese producto. Por eso, se recomienda iniciar CEMRepBoot desde RStudio siguiendo las instrucciones del apartado anterior.

Sin embargo, para quienes quieran hacerlo sin emplear el RStudio, los pasos son los siguientes:

1. Descargar y descomprimir la aplicación.
2. Ejecutar el programa R.
3. Ejecutar los siguientes comandos en la consola:
  - a. Instalar el paquete shiny manualmente.



```
R Console
> install.packages("shiny")
```

- b. Cargar el paquete shiny (este paso es necesario para tener acceso a las funciones incluidas en el paquete).



```
R Console
> library(shiny)
```

- c. Ejecutar CEMRepBoot con la función runApp del paquete shiny.  
shiny:runApp("Dirección a la carpeta con la aplicación/CEMRepBoot.R")



```
R Console
> runApp ("C:/..../CEMREPBoot/CEMREPBoot.R")
```

Una vez iniciada la aplicación, se abrirá en el navegador una pantalla como la que se observa en la figura 5; las opciones y el control se manejarán desde el tablero ubicado a la izquierda (figura 6).

**Figura 5. Pantalla de inicio de la aplicación**



Figura 6. Panel de control



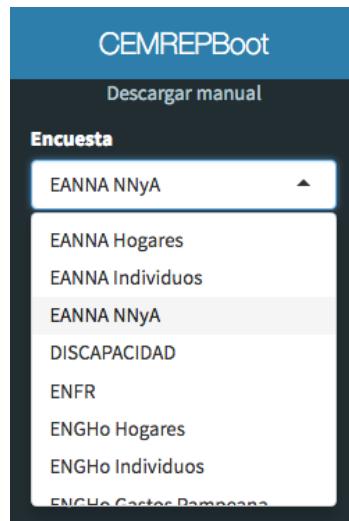
En la figura 6 pueden verse los distintos menús que posibilitan el uso de la aplicación. A continuación, se describe brevemente su funcionalidad:

**[1] Encuesta:** permite la carga de las bases de microdatos para realizar las estimaciones para las distintas encuestas. Por el momento, se encuentran disponibles los operativos:

- “EANNA Hogares”, “EANNA Individuos” y “EANNA NNyA”, que corresponden a los cuestionarios 1, 2 y 3 de la EANNA 2016-2017, respectivamente.
- “ENFR”, que corresponde a la ENFR 2018.
- “DISCAPACIDAD”, que corresponde al ENDI 2018.
- “ENGHo Hogares” y “ENGHo Individuos”, que corresponden a los formularios de hogar e individual de la ENGHo 2017-2018, respectivamente.
- “ENGHo Gastos Pampeana”, “ENGHo Gastos Patagonia”, “ENGHo Gastos Metropolitana”, “ENGHo Gastos Cuyo”, “ENGHo Gastos Noreste” y “ENGHo Gastos Noroeste”, que corresponden al formulario de gasto de la ENGHo 2017-2018 por región.

Las bases de las encuestas pueden ser descargadas a través del sitio web del Instituto, o se las puede incorporar localmente si fueron descargadas previamente. Una vez que se cargaron las bases en la aplicación, este se encarga automáticamente de unir la base de usuario con la base de réplicas.

Figura 7. Menú de selección de encuesta



**[2] Opciones para la carga de datos:** la aplicación permite elegir entre descargar las bases de microdatos desde la página del INDEC seleccionando la opción “Descargar desde INDEC” lo que despliega el menú **[2.1]**, o cargar las bases desde un archivo local seleccionando la opción “Subir desde archivo local”, lo cual despliega el menú **[2.2]**.

- El menú **[2.1]**, que se encuentra desplegado por defecto al iniciar la aplicación, permite al usuario descargar las bases de la encuesta seleccionada en el menú Encuesta [1] presionando el botón “Descargar Bases”. También es posible obtener la nota técnica, la documentación y el diccionario de variables para la misma encuesta.
- El menú **[2.2]** se desplegará al elegir la opción “Subir desde archivo local” y permite al usuario cargar las bases a la aplicación si estas están presentes en un directorio o una carpeta local (figura 8). Las bases de microdatos y de réplicas deben ser cargadas por separado, y se debe indicar un separador de columnas y un separador de decimales; estos últimos son los definidos por la encuesta en la documentación asociada a la base de microdatos. Al presionar el botón de “Carga bases”, los archivos ya deberán estar cargados y los separadores definidos para que la aplicación proceda a unir las bases y habilitar el menú Indicador [3].

Figura 8. Menú de carga de datos\*.

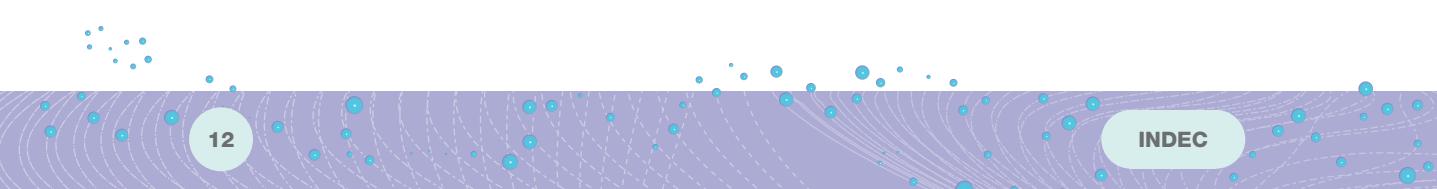


**Importante:** las bases de datos que se carguen en la aplicación deben respetar los nombres de las variables de identificación de registro, ponderadores y pesos replicados originales asociados a la encuesta en cuestión. **Cada encuesta nombra de distinta forma a estas variables.** Por ejemplo, en el caso de la ENGHO 2017-2018, la clave de identificación del registro se denomina *ID*; el ponderador, *pondera*; y los pesos replicados, *whog\_rep*.

[3] **Indicador:** permite seleccionar el indicador que se desea estimar (figura 9). Los disponibles en esta versión son:

- proporción
- total
- media
- razón

Figura 9. Menú de indicadores



Indicador
<b>Proporción:</b> <input type="text"/>
<b>Total:</b> <input type="text"/> <input checked="" type="checkbox"/> Total en las categorías
<b>Media:</b> <input type="text"/>
<b>Numerador razón:</b> <input type="text"/>
<b>Denominador razón:</b> <input type="text"/>

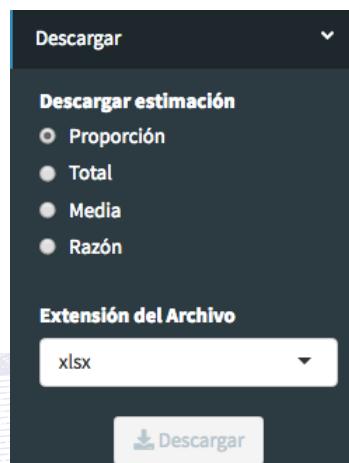
Es importante advertir que para proporciones, totales y medias es posible elegir más de una variable para realizar las estimaciones, y para razones deben especificarse por separado tanto el numerador como el denominador que definen la razón deseada (ver sección 4).

En el caso del total, puede seleccionarse la opción “*Total en las categorías*” para estimar el total de cada categoría (ver sección 4.2).

**[4] Subpoblación:** especifica el dominio de estimación para el que se desean estimaciones. Si se declaran varias variables categóricas, la aplicación tomará la combinación de estas como dominio al presentar los resultados.

**[5] Descargar:** permite descargar las estimaciones calculadas en formato xlsx o RDS. Para la descarga se debe definir el indicador que se quiere descargar y el formato (figura 10). Una vez definidas las opciones, con el botón “Descargar” se inicia la descarga. En el caso de haber elegido varias variables en el menú Indicador [3] y el formato de descarga xlsx, las estimaciones de cada variable se generarán en una hoja separada del archivo xlsx con el nombre **Estimación\_[INDICADOR].xlsx**, donde [INDICADOR] depende del indicador que se desee descargar (proporción, total, media o razón).

Figura 10. Menú de descarga

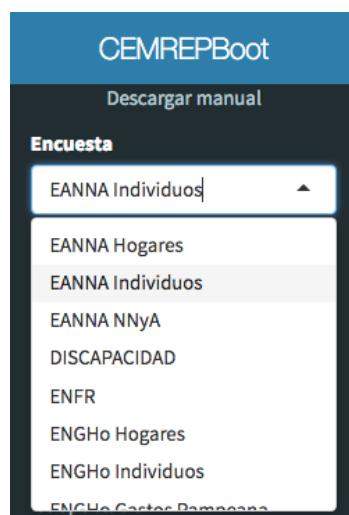


### 3. Uso de la aplicación y ejemplos

Como ya se señaló en la introducción, al momento de la redacción de este manual, la aplicación permite trabajar con todas aquellas encuestas del Instituto que disponen de una base de réplicas para el cálculo de los errores muestrales.

El primer paso es seleccionar en el menú desplegable la encuesta con la que desea trabajar y elegir si descargar las bases o subirlas desde un archivo local. A manera de ejemplo, se utilizará la base correspondiente a la EANNA 2016-2017, en particular, la base de individuos.

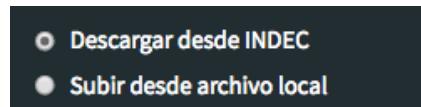
Figura 11. Selección de la base de microdatos EANNA individuos



### 3.1 Descarga de las bases desde la página web del INDEC

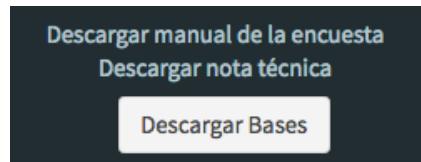
Una vez seleccionada la encuesta con la que se quiere trabajar, elegimos la opción “Descargar desde INDEC”.

Figura 12. Opciones para la carga de datos



Con el botón “Descargar Bases”, la aplicación procederá a descargar las bases con los datos de la encuesta y la base con las réplicas Bootstrap necesaria para las estimaciones de los errores de muestreo desde el sitio del INDEC.<sup>3</sup>

Figura 13. Descarga de las bases desde el sitio web del INDEC



### 3.2 Carga de las bases desde un archivo local

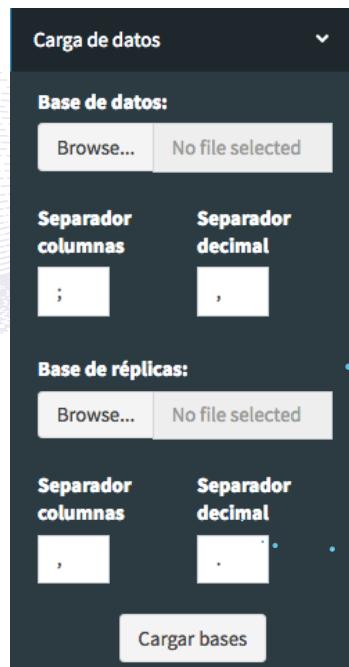
Como se indica en el apartado 2.2, la persona usuaria tiene la opción de cargar las bases de la encuesta y la de réplicas a la aplicación desde un entorno local, pero en este caso la carga es menos inmediata y es necesario completar más información. Una vez seleccionada la encuesta con la que se quiere trabajar, elegimos la opción “Subir desde archivo local”.

Figura 14. Carga de las bases desde un archivo local



<sup>3</sup> Las bases pueden encontrarse en el siguiente enlace: <https://www.indec.gob.ar/indec/web/Institucional-Indec-BasesDeDatos>.

Figura 15. Menú para carga local



En el caso de la EANNA, para la base de microdatos, el separador de columnas es ";" y el de decimales es ",". Para la base de réplicas el separador de columnas es "," y el de decimales, ".". Estos caracteres están señalados en la documentación asociada al manual de uso de las bases que acompañan a cada encuesta y disponible en el apartado correspondiente en el sitio web del Instituto. Se advierte que deben ser ingresados en los cuadros correspondientes sin comillas ni espacios antes o después.

Al elegir los archivos, tal como se muestra en la figura 15, aparecerán unas barras de carga que mostrarán la leyenda *Upload complete* una vez que la carga esté lista. Cuando ambas bases estén cargadas y los separadores, definidos; con el botón "Cargar bases", la aplicación procederá a su lectura y unión, lo que habilitará las opciones en el menú de Indicador [4] y Subpoblación [5].

**Nota:** cuando se cargan las bases de datos desde un archivo local, la aplicación permite solicitar estimaciones de variables construidas con base en variables de la encuesta, con excepción de la ENFR. Por tener esta última distintos ponderadores para cada paso de la encuesta, no es posible relacionar de forma automática el ponderador que le correspondería a la variable construida.

## 4. Selección del indicador a calcular para las estimaciones

Una vez seleccionadas y descargadas las bases, la aplicación habilita el menú Indicador [4]. Para cada indicador se dispondrá de una lista con las variables de la base de datos, donde se podrá seleccionar para cuáles se quieren obtener estimaciones.

Los nombres o las etiquetas de las variables se muestran o identifican como se las definió para la encuesta y como aparecen documentadas en el manual de uso de su base de microdatos. Los diccionarios con las definiciones para cada variable se pueden descargar desde el menú Descargar Bases desde la página del INDEC,<sup>4</sup> como se detalló anteriormente; también existe la posibilidad de hacer clic en el logo del INDEC que aparece en la esquina superior derecha de la aplicación CEMRepBoot, que lo redirigirá automáticamente.

Figura 16. Redirección al sitio web del INDEC

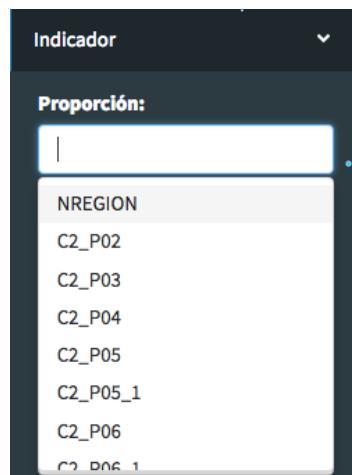


Como se adelantó en varios pasajes de este manual, la aplicación puede resolver estimaciones para proporciones, totales, medias y razones; en un futuro se espera incorporar otros tipos de indicadores o parámetros.

### 4.1 Estimación de una proporción

En el caso de querer estimar una proporción, el usuario deberá posicionarse sobre el cuadro correspondiente. Se desplegará una lista con las variables disponibles en la base, a partir de la cual se podrá seleccionar las que se desee estimar (figura 17).

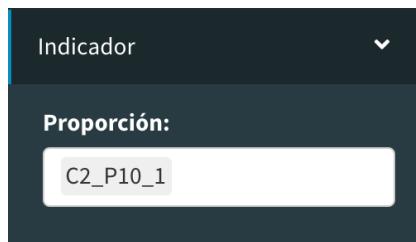
Figura 17. Selector de variables desplegado para la proporción



<sup>4</sup> Ver <https://www.indec.gob.ar/indec/web/Institucional-Indec-BasesDeDatos>.

Siguiendo con el ejemplo de la EANNA, si se quiere calcular la proporción de personas con obra social, se debe seleccionar la variable correspondiente, que en este caso está codificada como *C2\_P10\_1* (figura 18).

**Figura 18. Selección de la variable *C2\_P10\_1***



Una vez seleccionada la variable, el botón “Estimar” que se encuentra debajo del menú Subpoblación habilita a la aplicación a realizar la estimación que se demanda.

Los resultados obtenidos se muestran en la parte central de la pantalla, en las solapas correspondientes a los estimadores elegidos (figura 19). En el cuadro de resultados se muestra la información correspondiente a:

- dominio de estimación
- cantidad de observaciones con la que se calcula el indicador
- estimaciones puntuales para cada categoría de la variable
- límite inferior del intervalo de confianza del 95% para cada categoría
- límite superior del intervalo de confianza del 95% para cada categoría
- coeficiente de variación porcentual de cada estimación puntual
- coeficiente de variación porcentual del denominador.

**Figura 19. Descripción de los resultados arrojados por CEMRepBoot**

The screenshot shows a software interface with a table of results for the 'Proporción' indicator. The table has columns for 'Total\_Pais', 'Var\_Analysis', 'n', '1', '2', '9', 'Li\_1', 'Ls\_1', 'Li\_2', 'Ls\_2', 'Li\_9', 'Ls\_9', 'CV%\_1', 'CV%\_2', 'CV%\_9', and 'CV%\_denominador'. The first row shows 'Total pais' with 'Var\_Analysis' as 'C2\_P10\_1', 'n' as 81729, '1' as 46.99, '2' as 53, '9' as 0.01, 'Li\_1' as 45, 'Ls\_1' as 48.97, 'Li\_2' as 51.02, 'Ls\_2' as 54.98, 'Li\_9' as -0.02, 'Ls\_9' as 0.04, 'CV%\_1' as 2.15, 'CV%\_2' as 1.9, 'CV%\_9' as 119.48, and 'CV%\_denominador' as 0.03.

Callout boxes highlight specific parts of the table:

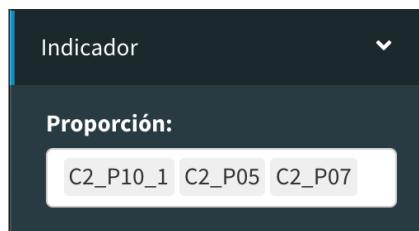
- Domínio de estimación** (top left)
- Cantidad de observaciones** (top middle)
- Límites de los intervalos de confianza** (top right)
- Coeficiente de variación del denominador** (far right)
- Variable de análisis** (bottom left)
- Estimaciones puntuales** (bottom middle)
- Coeficientes de variación porcentuales de las estimaciones** (bottom right)

Para este ejemplo, del cuadro se desprende que la proporción estimada de personas con obra social ( $C2\_P10\_1 = 1$ ) es del 46,99%, con un coeficiente de variación del 2,15% y un intervalo del 95% de confianza de (45,00%; 48,97%).

**Importante:** CEMRepBoot utiliza el “.” como separador decimal al mostrar los resultados.

La aplicación permite a la persona usuaria estimar los indicadores para varias variables en simultáneo. Suponiendo que además de la proporción de personas con obra social se desee estimar la proporción de personas que viven con su padre en el hogar y la proporción de personas por su situación conyugal, etiquetadas en la base de la encuesta con  $C2\_P05$  y  $C2\_P07$ , respectivamente; las variables deben agregarse en el apartado correspondiente a la proporción sumándolas a  $C2\_P10\_1$  como se detalla en la figura 20.

**Figura 20. Selección de las variables  $C2\_P10\_1$ ,  $C2\_P05$  y  $C2\_P07$**



Al solicitar a la aplicación que realice las estimaciones a través del botón “Estimar”, se generarán nuevas solapas en el cuadro de proporción para contener los resultados de las nuevas variables añadidas (figuras 20.1 y 20.2).

**Figura 20.1 Resultados para la variable  $C2\_P05$**

	Proporción	Proporción2	Proporción3												
Total_Pais	Var_Analysis	n	1	2	9	Li_1	Ls_1	Li_2	Ls_2	Li_9	Ls_9	CV%_1	CV%_2	CV%_9	CV%_denominador
Total pais	C2_P05	81730	30.96	69.04	0	28.54	33.37	66.62	71.46	0	0	3.98	1.79	62.62	0.03

**Figura 20.2 Resultados para la variable  $C2\_P07$**

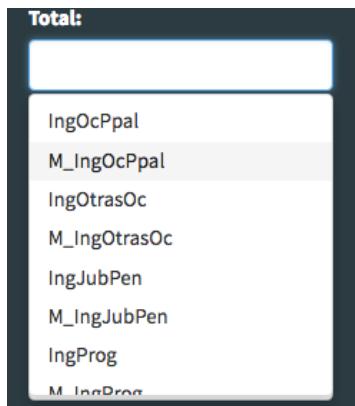
	Proporción	Proporción2	Proporción3												
Total_Pais	Var_Analysis	n	1	2	3	4	5	6	9	Li_1	Ls_1	Li_2	Li_3	Ls_2	Ls_3
Total pais	C2_P07	81730	16.79	23.81	4.58	1.85	4.94	47.92	0.11	16.06	17.51	22.89	24.73	3.99	5.18

El usuario podrá visualizar los distintos cuadros de resultados navegando a través de las solapas generadas para cada variable.

## 4.2 Estimación de un total

En el caso de querer estimar el total de una variable, o el total para las categorías de una variable discreta, la persona usuaria deberá posicionarse sobre el cuadro correspondiente (Total), lo que desplegará una lista con las variables disponibles en la base, a partir de la cual podrá seleccionar la que deseé estimar (figura 21).

Figura 21. Selector de variables desplegado para el total



Para el ejemplo del total de una variable, se procederá a calcular el total de la variable Ingreso de ocupación principal etiquetada en la base de la encuesta de la EANNA con el nombre *IngOcPpal* (figura 22).

Figura 22. Selección de la variable *IngOcPpal*



Una vez seleccionada la variable, el botón “Estimar”, que se encuentra debajo del menú Subpoblación, le pide a la aplicación que calcule el indicador.

Los resultados obtenidos se muestran en la parte central de la pantalla, en las solapas correspondientes a los estimadores elegidos (figura 23). En el cuadro de resultados se muestra la información correspondiente a:

- dominio de estimación
- cantidad de observaciones con la que se calculó el indicador
- estimación puntual del total poblacional de la variable elegida
- límite inferior del intervalo de confianza del 95%
- límite superior del intervalo de confianza del 95%
- coeficiente de variación porcentual de cada estimación puntual

**Figura 23. Resultados para el total de la variable *IngOcPpal***

Total					
Total_Pais	n	Total_IngOcPpal	Li_Total_IngOcPpal	Ls_Total_IngOcPpal	CV%_Total_IngOcPpal
Total país	14255	71368165867	66606453370.517	76129878363.483	3.404

Así, en los resultados que se muestran en la figura 23, puede observarse la estimación del total de ingreso de la ocupación principal en la columna **Total\_IngOcPpal**, con un coeficiente de variación del 3,404% y con su respectivo intervalo de confianza del 95%.

Al igual que en los cálculos de proporciones, pueden indicarse varias variables simultáneamente para el cálculo de totales, para las cuales se procederá de la misma forma indicada y cuyos resultados se dispondrán en distintas pestañas.

Para calcular en una variable categórica la estimación del total de unidades que corresponden a cada una de sus categorías, se procede como en la sección anterior. Por ejemplo, si para la EANNA se desea estimar el total de personas en las distintas categorías ocupacionales, debe seleccionarse en el total la variable **CATOCUP**, marcando la casilla “*Total en las categorías*”, de esta forma la aplicación totaliza por categoría.

**Nota:** al marcar esta casilla, la aplicación calcula el total de cada categoría para todas las variables seleccionadas. No es posible calcular, simultáneamente, totales de variables cuantitativas y de categorías de variables categóricas.

**Figura 24. Selección de la variable *CATOCUP***



Para que la aplicación inicie el cálculo de indicadores siempre se debe presionar el botón “Estimar” debajo del menú Subpoblación. Los elementos del cuadro de resultados obtenidos son los mismos que para el total de una variable (dominio de estimación, cantidad de observaciones, estimación puntual, etc.), calculados para cada categoría de la variable elegida. En lugar de tratar a las variables como numéricas, esta opción trata a la variable como categórica, calculando la suma ponderada para cada una de las categorías de la variable (figuras 25.1, 25.2 y 25.3).

Figura 25.1 Resultados para el total de la variable **CATOCUP** (parte 1)

Total							
Total_Pais	n	Total_CATOCUP_1	Total_CATOCUP_2	Total_CATOCUP_3	Total_CATOCUP_4	Total_CATOCUP_9	Li_Total_CATOCUP_1
Total pais	18562	7053952	1869092	395600	48817	22939	6872419.55

Figura 25.2 Resultados para el total de la variable **CATOCUP** (parte 2)

Total											
Ls_Total_CATOCUP_1	Li_Total_CATOCUP_2	Ls_Total_CATOCUP_2	Li_Total_CATOCUP_3	Ls_Total_CATOCUP_3	Li_Total_CATOCUP_4	Ls_Total_CATOCUP_4	Li_Total_CATOCUP_9	Ls_Total_CATOCUP_9	CV%_Total_CATOCUP_1	CV%_Total_CATOCUP_9	
7235484.45	1753183.999	1984994.001	342060.445	449139.555	24213.101	73420.899	10469.334	35408.666	1.313		

Figura 25.3 Resultados para el total de la variable **CATOCUP** (parte 3)

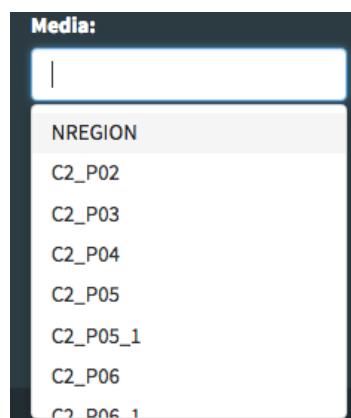
CV%_Total_CATOCUP_1	CV%_Total_CATOCUP_2	CV%_Total_CATOCUP_3	CV%_Total_CATOCUP_4	CV%_Total_CATOCUP_9
1.313	3.164	6.905	25.715	27.735

Así puede verse, por ejemplo, que el total estimado de asalariados ( $CATOCUP = 1$ ) entre personas ocupadas de 18 años o más, que residan en hogares con niños, niñas y adolescentes es de 7.053.952, con un coeficiente de variación del 1,313% y con un intervalo de confianza del 95% definido por (6.872.419,55; 7.235.484,45).

## 4.3 Estimación de una media

En el caso de querer estimar la media de una variable, la persona usuaria deberá posicionarse sobre el cuadro correspondiente (Media), lo que desplegará una lista con las variables disponibles en la base, a partir de la cual deberá seleccionar aquellas que desee estimar (figura 26).

Figura 26. Selector de variables desplegado para la media



Siguiendo el ejemplo con la EANNA, se calculará la media de ingreso de ocupación principal, etiquetada en la base de la encuesta como *IngOcPpal* (figura 27).

**Figura 27. Selección de la variable *IngOcPpal* para la media**



Una vez seleccionada la variable, y presionando el botón “Estimar” debajo del menú Subpoblación, la aplicación procederá a realizar los cálculos de la media. Los resultados obtenidos en este caso son:

**Figura 28. Resultados para la media de la variable *IngOcPpal***

Media					
Total_Pais	n	Media_IngOcPpal	Li_Media_IngOcPpal	Ls_Media_IngOcPpal	CV%_Media_IngOcPpal
Total pais	14255	10723.782	10165.519	11282.046	2.656

Al igual que en los cálculos de proporciones y de totales, pueden seleccionarse varias variables simultáneamente para el cálculo de medias, siguiendo las indicaciones del apartado anterior.

## 4.4 Estimación de una razón

En el caso de querer estimar una razón, la persona usuaria deberá especificar la variable del numerador y la del denominador. Al posicionarse sobre el cuadro correspondiente (Numerador razón o Denominador razón), se desplegará una lista con las variables disponibles en la base, a partir de la cual el usuario seleccionará la que desea utilizar en la estimación del indicador.

Cabe aclarar que, a diferencia del resto de los indicadores, la aplicación en la actualidad no permite estimar más de una razón en forma simultánea.

A modo de ejemplo se va a estimar la razón entre el ingreso de la ocupación principal y el ingreso de otras ocupaciones, variables etiquetadas como *IngOcPpal* y *IngOtrasOc*, respectivamente. El primer paso es seleccionar estas variables en el menú “Indicador” (figura 29).

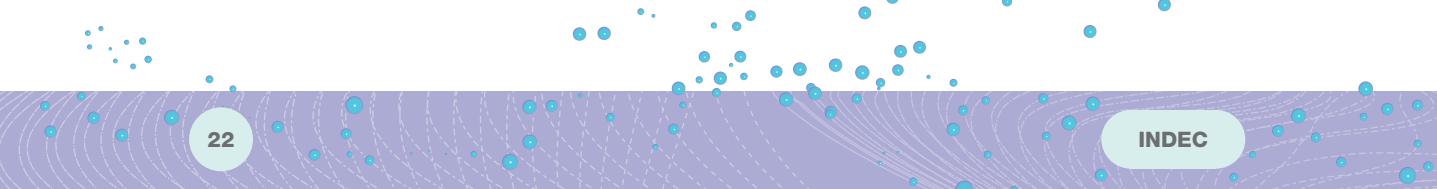


Figura 29. Selección de las variables *IngOcPpal* y *IngOtrasOc*



Una vez seleccionadas las variables con las cuales se desea calcular el indicador, y presionando el botón “Estimar” debajo del menú Subpoblación, la aplicación realiza las estimaciones y presenta los resultados, como puede observarse en la pantalla principal (figura 30).

Figura 30. Resultados para la razón entre las variables *IngOcPpal* y *IngOtrasOc*

Razón						
Total_Pais	n	IngOcPpal/IngOtrasOc	Li_IngOcPpal/IngOtrasOc	Ls_IngOcPpal/IngOtrasOc	CV%_IngOcPpal/IngOtrasOc	CVden%_IngOcPpal/IngOtrasOc
Total país	935	2.086	1.857	2.315	5.601	5.213

## 5. Estimación en subpoblaciones o dominios

Por defecto, todas las estimaciones se realizan a nivel país, la mayor agregación de la encuesta o población objetivo. En el caso de que se quiera utilizar otro dominio de estimación, se debe especificar cuál a través del menú Subpoblación que se encuentra debajo del menú Indicador.

En este menú se desplegará una lista con todas las variables disponibles en la base de datos, a partir de la cual se podrá seleccionar la variable que identifique a las subpoblaciones o los dominios para los cuales se quiera calcular el indicador.

Retomando el ejemplo desarrollado, para el cálculo de una proporción, y recordando que se estimó la proporción de personas con obra social, se supone que esta proporción requiere ser estimada a nivel de región; es decir que se quiere estimar el indicador para las regiones GBA, NOA, NEA, Cuyo, Pampeana y Patagonia, todos dominios de estimación de la EANNA.

A las regiones se las identifica dentro de la base por la variable *NREGION*; por lo que, si la seleccionamos en el menú Subpoblación (figura 31) y solicitamos a la aplicación que realice los cálculos presionando el botón “Estimar”, obtendremos el siguiente resultado de la figura 32.

Figura 31. Selección de la variable *NREGION*

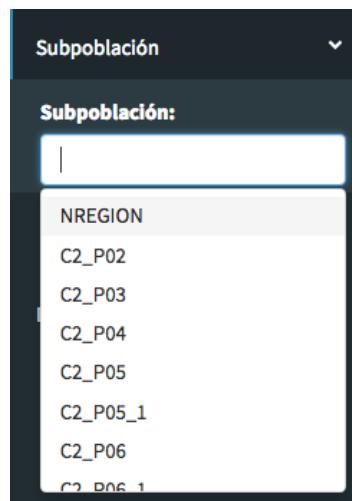


Figura 32. Resultados para la proporción de la variable *C2\_P10\_1* por *NREGION*

Proporción															
<i>NREGION</i>	Var_Analysis	<i>n</i>	1	2	9	<i>Li_1</i>	<i>Ls_1</i>	<i>Li_2</i>	<i>Ls_2</i>	<i>Li_9</i>	<i>Ls_9</i>	<i>CV%_1</i>	<i>CV%_2</i>	<i>CV%_9</i>	<i>CV%_denominador</i>
1	C2_P10_1	13605	42.54	57.43	0.03	39.95	45.12	54.85	60	-0.05	0.12	3.1	2.29	119.49	0.07
40	C2_P10_1	15717	47.03	52.97	0	44.68	49.38	50.62	55.32	0	0	2.55	2.26		0.01
41	C2_P10_1	11289	45.77	54.23	0	41.13	50.41	49.59	58.87	0	0	5.17	4.37		0.01
42	C2_P10_1	10571	48.65	51.35	0	45.5	51.79	48.21	54.5	0	0	3.3	3.13		0.04
43	C2_P10_1	20432	50.02	49.98	0	46.79	53.26	46.74	53.21	0	0	3.3	3.3		0.01
44	C2_P10_1	10115	59.76	40.24	0	57.26	62.26	37.74	42.74	0	0	2.14	3.17		0

Como se puede ver, la columna de **Total\_pais** fue reemplazada por *NREGION* y las estimaciones fueron calculadas para cada una de las regiones.

El menú de Subpoblación permite elegir más de una variable para definir los dominios de estimación. Por ejemplo, si se quiere la proporción de personas por sexo para cada categoría ocupacional dentro de cada región, deben especificarse ambas variables (*NREGION* y *CATOCUP*) en el menú Subpoblación.

Figura 33. Selección de las variables *NREGION* y *CATOCUP*



Como resultado se obtendrá una estimación para cada combinación de los niveles de las variables especificadas, como se observa en la figura 34.

**Figura 34. Resultados para la proporción de la variable *C2\_P10\_1* por *NREGION* y *CATOCUP***

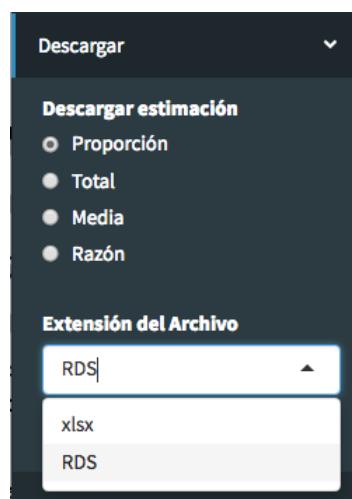
Proporción																
NREGION	CATOCUP	Var_Analysis	n	1	2	9	Li_1	Ls_1	Li_2	Ls_2	Li_9	Ls_9	CV%_1	CV%_2	CV%_9	CV%_denominador
1	1	C2_P10_1	2196	56.5	43.5	0	53.3	59.69	40.31	46.7	0	0	2.88	3.75		2.57
40	1	C2_P10_1	3063	59.51	40.49	0	56.28	62.75	37.25	43.72	0	0	2.77	4.08		2.4
41	1	C2_P10_1	1713	60.1	39.9	0	51.87	68.32	31.68	48.13	0	0	6.99	10.52		2.57
42	1	C2_P10_1	1685	63.14	36.86	0	57.95	68.33	31.67	42.05	0	0	4.19	7.18		3.09
43	1	C2_P10_1	3496	60.43	39.57	0	56.23	64.64	35.36	43.77	0	0	3.55	5.42		2.92
44	1	C2_P10_1	1938	72.58	27.42	0	69.06	76.1	23.9	30.94	0	0	2.47	6.55		2.23
1	2	C2_P10_1	529	24.78	75.22	0	19.89	29.67	70.33	80.11	0	0	10.07	3.32		5.07
40	2	C2_P10_1	756	25.06	74.94	0	21.9	28.21	71.79	78.1	0	0	6.42	2.15		3.85
41	2	C2_P10_1	627	26.46	73.54	0	19.9	33.01	66.99	80.1	0	0	12.64	4.55		5.09
42	2	C2_P10_1	508	30.28	69.72	0	25.08	35.47	64.53	74.92	0	0	8.75	3.8		5.67

**Nota:** es posible que al indicar múltiples variables de forma simultánea en el menú Subpoblación, en alguno de los grupos formados no se hayan registrado observaciones, es decir,  $n=0$ . En estos casos, no es posible realizar estimaciones, por lo que estos grupos serán excluidos de la tabla de resultados.

## 6. Descarga de resultados

El menú de Descarga permite descargar los resultados alcanzados y presentados de las estimaciones requeridas. La persona usuaria deberá seleccionar qué cuadro desea y el formato requerido. Los cuadros deben seleccionarse de a uno por vez y la aplicación permite elegir entre formato xlsx o RDS (figura 35).

Figura 35. Menú Descarga



## 7. Recomendaciones para interpretar las estimaciones

Para ayudar a las personas usuarias a interpretar los resultados de la encuesta, se presentan algunas recomendaciones y sugerencias para identificar estimaciones en las que se debe poner poca o ninguna confianza.

El siguiente cuadro cubre algunas de las situaciones más generales por las que puede atravesar una estimación a la hora de tener que evaluar su precisión o la confianza que se puede poner en ella. Cada encuesta posee sus criterios particulares, si bien las condiciones que se evalúan son las mismas (para más detalle referirse a la nota técnica de la encuesta de interés).

Cuando se leen los resultados oficiales publicados de la encuesta, o se generan estimaciones propias a partir de la base que entrega el Instituto, se deben tener en cuenta estas recomendaciones a la hora de sacar conclusiones del fenómeno que se esté estudiando a partir de la encuesta.

A modo de ejemplo, a continuación se presentan las recomendaciones publicadas en la nota técnica de la Encuesta Nacional de Gastos de los Hogares 2017-2018.

**Cuadro 1. Recomendaciones para interpretar las estimaciones**

Calidad de la estimación	Condición	Recomendaciones
No confiable	<ul style="list-style-type: none"> <li>a) el total de unidades involucradas en el cálculo de la estimación es menor a 100</li> <li>b) la estimación de una razón es menor a 0,03</li> <li>c) la estimación de una proporción es menor al 3%</li> <li>d) el denominador de un cociente, razón, o proporción, tiene un CV &gt; 10% (CV: coeficiente de variación)</li> <li>e) la estimación posee un CV &gt; 33,3%.</li> </ul>	Se recomienda no emplear a la estimación en este caso. Si existe la necesidad de publicarla, se debe advertir que las conclusiones basadas en ella no son confiables o válidas.
Poco confiable	La estimación posee un CV en el rango: $16,6\% < CV \leq 33,3\%$ .	<p>La estimación debe ser considerada con precaución.</p> <p>La inferencia resultante puede presentar, con alta probabilidad, un nivel de error elevado.</p> <p>Se recomienda presentarla con alguna notación advirtiendo de esta situación en la que se encuentra.</p>
Confiable	La estimación posee un CV en el rango: $CV \leq 16,6\%$ .	La estimación puede ser considerada sin restricciones. No se requiere una notación especial para esta situación.

**Fuente:** INDEC, Encuesta Nacional de Gastos de los Hogares 2017-2018.

En la aplicación, cuando alguna de las estimaciones calculadas cumpla una de las condiciones para ser considerada poco confiable o no confiable, se generará automáticamente una nueva columna con advertencias al final de los cuadros de resultados, en la que se identificará con \* las estimaciones poco confiables y con \*\* las estimaciones no confiables.

Por ejemplo, si en la base de datos de la EANNA individuos queremos calcular la proporción de personas por sexo para cada categoría ocupacional, veremos que para alguna de las categorías de esta variable se generarán advertencias, como muestra la figura 36.

**Figura 36. Ejemplo de advertencias en las estimaciones**

Proporción													
CATOCUP	Var_Analysis	n	1	2	L1_1	Ls_1	L1_2	Ls_2	CV%_1	CV%_2	CV%_denominador	Adv_1	Adv_2
1	C2_P03	14092	56.08	43.92	54.83	57.34	42.66	45.17	1.14	1.46	1.31		
2	C2_P03	3630	59.53	40.47	56.75	62.3	37.7	43.25	2.38	3.49	3.16		
3	C2_P03	698	73.54	26.46	68.56	78.51	21.49	31.44	3.45	9.59	6.91		
4	C2_P03	111	46.21	53.79	30.85	61.56	38.44	69.15	16.96	14.57	25.71	**	**
9	C2_P03	31	44.93	55.07	20.66	69.2	30.8	79.34	27.56	22.48	27.74	**	**

Como se puede ver, hay dos categorías del cuadro con advertencias para cada estimación puntual. En particular, para la categoría CATOCUP = 4, las advertencias surgen de que el coeficiente de variación del denominador porcentual es mayor que 10%; en el caso de CATOCUP = 9, el CV% del denominador es mayor que 10% y, además, el número de unidades con las que se calculó el indicador es menor que 100. En ambas situaciones las estimaciones son no confiables.

## 8. Tratamientos especiales: ENGHo Gastos

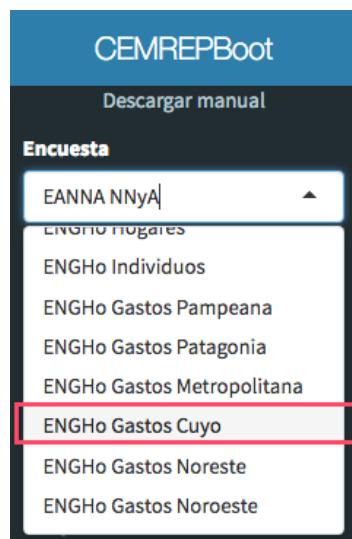
El módulo de gastos de la ENGHo posee indicadores particulares que no se asemejan a los de otras encuestas del Instituto.

Estos son los indicadores de *estructura de gastos*: que pueden pensarse como el *gasto total en artículos de una “división de gasto”* respecto al *gasto total en un “nivel de agregación geográfico”* dado.

- **División de gasto:** grupo en la que se clasifican los artículos adquiridos por un hogar (división, grupo, clase, subclase).
- **Nivel de agregación geográfico:** división geográfica sobre la que se considera el total de gasto de los hogares (región, provincia, total país).

Esta base de datos presenta particularidades con respecto al resto de las encuestas que implicaron desarrollar un módulo específico en CEMRepBoot para estimar las estructuras de gastos. La primera particularidad es que al calcular indicadores para este módulo se considera una gran diversidad de desagregaciones del gasto y de los niveles geográficos. Por otro lado, en esta base hay más de un registro por hogar, lo que dificulta los cálculos.

Figura 37. Selección de las bases correspondientes a la ENGHo Gastos en la región Cuyo



La base de datos correspondientes al formulario de gastos de los hogares contiene una gran cantidad de registros, lo que la vuelve muy difícil de manejar para los equipos estándares de 8 GB de memoria RAM. Por este motivo, en el selector encuesta de CEMRepBoot se presenta la ENGHo Gastos por región (figura 37). No es posible realizar estimaciones a nivel nacional (si el usuario carga de forma manual la base completa CEMRepBoot trabajará únicamente con los registros correspondientes a la región seleccionada).

Al seleccionar alguna de las opciones de ENGHo Gastos, el panel para elegir el indicador a calcular se modificará según se muestra en la figura 34.

Figura 38. Menú Indicador en la ENGHo Gastos



La persona usuaria en este caso deberá identificar la variable de análisis (*monto* o *ingreso*) y solo podrá optar por calcular un total o una razón. Por ejemplo, si se quiere calcular una estructura de gastos por división, debe seleccionarse el monto como variable de análisis y división en la casilla correspondiente al numerador de razón (figura 39).

Figura 39. Selección de la variable *monto*

Indicador

Variable de análisis:

total:

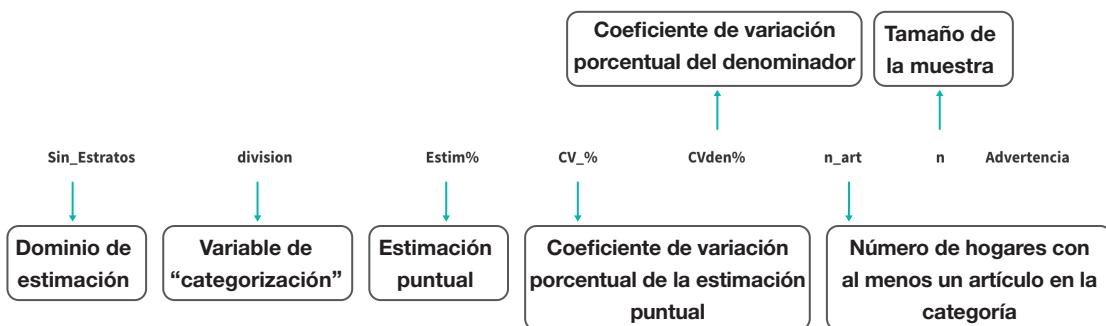
Total en las categorías

Razón:

division

A las columnas ya conocidas, se le agrega la de **n\_art**, que es el número de hogares con al menos un artículo en la categoría, ya que un mismo hogar puede aportar más de un artículo para el cálculo del **n** (figura 40).

Figura 40. Descripción de los resultados arrojados por CEMRepBoot para ENGHo Gastos



Al momento de estimar la estructura de gastos, cada categoría de la variable de análisis (en este caso división) se presentará como una fila en la salida de la aplicación (figura 41) y si se desea replicar el cálculo en subpoblaciones, cada dominio de estimación tendrá su propia pestaña en el indicador escogido (figura 42).

**Figura 41. Resultados para la estructura de gastos por *DIVISION***

Razón	Sin_Estratos	division	Estim%	CV_%	CVden%	n_art	n	Advertencia
1		A01	23.1	2.5	2.7	2015	2033	
1		A02	2.1	5.7	2.7	871	2033	**
1		A03	6.9	3.7	2.7	1356	2033	
1		A04	13.2	2.8	2.7	2022	2033	
1		A05	5.5	4.3	2.7	1547	2033	
1		A06	6.2	5.9	2.7	1172	2033	
1		A07	16.8	3.7	2.7	1778	2033	
1		A08	4.9	2.2	2.7	1746	2033	
1		A09	8.6	5.4	2.7	1772	2033	
1		A10	2.9	6.8	2.7	433	2033	**
1		A11	5.2	4.1	2.7	1128	2033	
1		A12	4.5	4.9	2.7	1655	2033	

**Figura 42. Resultados para la estructura de gastos por *DIVISION* para cada provincia**

Razón	Razón2	Razón3	provincia	division	Estim%	CV_%	CVden%	n_art	n	Advertencia
50				A01	22	3.4	3.8	814	819	
50				A02	2	7.6	3.8	385	819	**
50				A03	6.1	5.6	3.8	561	819	
50				A04	12.6	3.8	3.8	811	819	
50				A05	5.9	5.8	3.8	664	819	
50				A06	6.6	7.7	3.8	538	819	
50				A07	17.3	5	3.8	728	819	
50				A08	4.9	3	3.8	740	819	
50				A09	9	7.1	3.8	729	819	
50				A10	3.3	8	3.8	228	819	
50				A11	5.1	5.7	3.8	463	819	
50				A12	5.1	6.1	3.8	744	819	