Aspectos metodológicos y conceptuales sobre el uso de microdatos censales.

CENSO 2023

Actualización:
Mayo 2026

Índice general

Contenido

Índice general 2

Índice de cuadros 3

1) Introducción 4

2) Características generales del Censo 2023 4

3) Tablas de microdatos disponibles 6

3.1. Tabla de personas ampliada 6

3.2. Tabla de viviendas 7

4) Ponderación de los microdatos censales 8

4.1. Omisión censal y uso de los ponderadores 8

4.2. Conceptos básicos para la estimación con datos ponderados 8

4.2.1. Interpretación de los ponderadores 9

4.2.2. Consecuencias en el procesamiento y redondeo 10

4.3. Procedimiento de redondeo y preservación de totales 10

4.4. Cómputo de los errores estándar en R 11

4.5. Comparabilidad entre censos 14

5) Guía de variables y uso de identificadores en los microdatos censales 15

5.1. Identificadores y estructura de los microdatos 15

5.2. Variables analíticas y de clasificación 15

6) Agregación y niveles de análisis 16

6.1. Nivel hogar 16

6.2. Nivel Vivienda 17

7) Tratamiento de inconsistencias y variables imputadas 17

7.1. Origen del agua para beber y cocinar 18

7.2. Tipo de tenencia de la propiedad de la vivienda 19

7.3. Evacuación del baño 20

7.4. Ocupación 20

7.5. Tratamiento de hogares con ausencia de adultos 22

7.6. Ajuste en la georreferenciación de viviendas colectivas 22

7.7. Asistencia a un centro educativo 23

8) Variables agregadas 24
8.1. Discapacidad 24
8.2. Departamento de residencia cinco años antes 25
8.3. Departamento de nacimiento 25
8.4. Total de personas en el hogar 25
8.5. Condición de actividad económica 26
8.6. MUNICIPIO 26
8.7. Nivel educativo de la población 27
9) Anonimización 27
10) Versiones publicadas. 29
11) Anexos 30

Índice de cuadros

Cuadro 1 - Recuento ponderado y sin ponderar de personas según tipo de vivienda 6
Cuadro 2 - Recuento ponderado y sin ponderar de hogares según tipo de vivienda 6
Cuadro 3 - Recuento sin ponderar de viviendas por tipo, según condición de ocupación 7
Cuadro 4 - Ajuste en la georreferenciación de viviendas colectivas 23
Cuadro 5 - Construcción de la variable POBPCOAC 30
Cuadro 6 - Construcción de la variable NIVELEDU 31

1) Introducción

El presente documento tiene como objetivo especificar las características de los microdatos correspondientes al Censo de Población, Hogares y Viviendas 2023. Esta documentación está orientada a optimizar la explotación estadística de los datos censales por parte de las/os usuarias/os y a asegurar su correcta interpretación, en concordancia con los principios de transparencia y rigor técnico del Instituto Nacional de Estadística (INE).

En particular, se detallan las unidades de análisis (viviendas, hogares y personas), la delimitación del universo de estudio y la fuente de procedencia de los datos. Asimismo, se describe el tratamiento de las variables fundamentales y el proceso de generación de variables agregadas, el tratamiento de inconsistencias y variables imputadas, los protocolos de anonimización aplicados y la gestión de las versiones de los microdatos. Finalmente, se proporcionan las directrices para la correcta aplicación de los ponderadores necesarios para obtener estimaciones representativas de la población total.

2) Características generales del Censo 2023

El Censo 2023 fue un censo de derecho, es decir, las personas fueron censadas en su lugar de residencia habitual, y la fecha de referencia del censo es el 31 de mayo de 2023.

El universo de estudio del Censo 2023 abarca a todos los residentes habituales del país, relevados en viviendas particulares y colectivas, o en situación de calle.

Las unidades de análisis comprenden las viviendas, los hogares y las personas residentes en el territorio nacional.

- Viviendas: se clasifican en particulares o colectivas[1] y según su condición de ocupación (ocupadas con moradores presentes, ocupadas con moradores ausentes o desocupadas).
- Hogares: se identifican tanto en viviendas particulares como colectivas, y se vinculan exclusivamente a viviendas clasificadas como ocupadas con moradores presentes[2].

---

[1] La **vivienda particular** es aquella construcción que está destinada a albergar uno más hogares particulares. la vivienda en la que residen uno o más hogares particulares. Se entiende por hogar particular al conjunto de personas con o sin vínculos de parentesco que habitan bajo un mismo techo y que, al menos para su alimentación, comparten un fondo común (es decir, participan de una "olla común"). Una persona que vive sola también constituye un hogar particular. La **vivienda colectiva**, en cambio, está destinada a hogares colectivos (personas generalmente no ligadas por vínculos de parentesco que comparten la vivienda por razones de trabajo, atención médica, estudio, religiosas, de reclusión, etc.

[2] **Ocupada con residentes presentes:** Vivienda en la que residen en forma habitual uno o más hogares y en el momento de la visita del Censista están presentes uno o más de sus moradores.

Ocupada con residentes ausentes: Vivienda en la que residen en forma habitual uno o más hogares y que, en el momento de la visita del Censista, no se encuentra presente ninguno de sus moradores o persona que pueda

- Personas: Corresponden a los residentes habituales asociados a un hogar (particular o colectivo) o identificados en situación de calle.

Para facilitar el análisis, se dispone de la variable **UNIVERSO**, la cual permite identificar a las personas según su pertenencia a hogares particulares, hogares colectivos o situación de calle.

El operativo de campo del Censo 2023 se llevó a cabo, entre el 29 de abril y el 30 de setiembre, mediante un diseño combinado que integró múltiples modalidades de recolección de datos, con el fin de optimizar la cobertura y eficiencia del operativo.

- **CAPI** (Computer-Assisted Personal Interviewing): cuestionario administrado por un entrevistador mediante una entrevista presencial asistida por computadora.
- **CAWI** (Computer-Assisted Web Interviewing): cuestionario autoadministrado por los informantes a través de internet.
- **CATI** (Computer-Assisted Telephone Interviewing): cuestionario administrado por un entrevistador de forma telefónica.

El conteo del operativo de campo se complementó con la integración de 12 mil personas, aproximadamente, provenientes de registros administrativos. Específicamente, personas de hasta 15 años de edad, que fue posible integrarlas a un hogar a partir de la información del Certificado de Nacido Vivo (CNV): niños, niñas y adolescentes, no censados, asociados a su madre (registrada en el operativo censal) usando el CNV. Para esta población, agregada desde registros administrativos, se cuenta con información sobre sexo y edad, asistencia a un centro educativo y las variables asociadas al hogar y la vivienda de residencia.

La variable **AGREGADO_A_HOGAR_CENSADO** permite identificar a los menores de hasta 15 años agregados desde registros administrativos y la variable **MADRE_PERID** identifica a la madre de dichos menores.

Por otra parte, el diseño del operativo censal implicó la aplicación de dos tipos de cuestionarios: Completo y Básico:

- El cuestionario Completo se aplicó a las viviendas particulares ocupadas con moradores presentes y a ciertas viviendas colectivas (hoteles, pensiones u otras casas de hospedaje, residenciales de estudiantes, casas de peones, internados religiosos y otra colectiva[3]). La variable **CUESTIONARIO_COMPLETO** identifica a las personas a las que se les aplicó este instrumento.
- El cuestionario Básico se utilizó para personas en situación de calle, el remanente de viviendas colectivas[4] y los hogares particulares relevados durante la fase de

---

brindar la información. Se establece que la vivienda está habitada ya sea por diálogo con los vecinos, portero, etc. o porque se encuentra indicios de ello (se observa ropa tendida, una mascota dentro, etc.).

[3] **Otra colectiva**: Comprende las viviendas colectivas no clasificadas según definiciones anteriores, por ejemplo los campamentos de trabajadores (forestales, de carreteras, etc.). Esta opción se utilizó en casos excepcionales.

[4] Hospitales y sanatorios con internación permanente, residenciales de ancianos, establecimientos de reclusión, penitenciarios, hogares y/o centros de atención / protección, cuarteles militares o policiales, refugios y centros de rehabilitación no carcelaria con internación permanente.

recuperación. Este tipo de cuestionario contiene las variables sexo, edad y en casos específicos se relevan ciertas variables de migración, educación y fecundidad según corresponda, y es identificado mediante la variable **CUESTIONARIO_BASICO**.

3) Tablas de microdatos disponibles

3.1. Tabla de personas ampliada

En esta tabla cada fila representa una persona, registradas durante el operativo de campo, más las personas de hasta 15 años provenientes de registros administrativos.

La tabla contiene 3.151.118 registros, de los cuales:

- 3.135.309 corresponden a personas censadas en viviendas ocupadas con moradores presentes,
- 12.305 corresponden a personas de hasta 15 años provenientes de registros administrativos, y vinculados a hogares.
- 3.504 corresponden a personas en situación de calle.

Cada fila de personas presenta, además, las variables de los cuestionario de hogar y vivienda correspondientes, y su ponderador para expandir la población censada a la población estimada al 31 de mayo de 2023 (3.499.451 personas).

A partir de esta tabla de personas ampliada es posible caracterizar las unidades de análisis de hogar y vivienda ocupada, llevando a cabo una agregación que permita cambiar el nivel de la unidad de análisis (de personas a hogares o viviendas); aplicando las reglas que se detallan más adelante, en el capítulo Agregación y niveles de análisis.

Cuadro 1 - Recuento ponderado y sin ponderar de personas según tipo de vivienda

	Personas en viviendas particulares	Personas en viviendas colectivas	Personas en situación de calle	Total
Recuento sin ponderar	3.100.348	47.266	3.504	**3.151.118**
Recuento ponderado	3.446.277	49.670	3.504	**3.499.451**

Cuadro 2 - Recuento ponderado y sin ponderar de hogares según tipo de vivienda

	Hogares en viviendas particulares	Hogares en viviendas colectivas	Total

Recuento sin ponderar	1.252.004	3.058	1.255.062
Recuento ponderado	1.373.546	3.375	1.376.921

3.2. Tabla de viviendas

El objetivo de esta tabla es presentar el recuento de viviendas obtenido en el operativo censal. Cada fila representa una vivienda e incluye viviendas ocupadas (con moradores presentes y ausentes) y desocupadas.

La tabla de microdatos de viviendas no incluye ponderador (por lo tanto, no cuenta con ningún ajuste por omisión censal) y abarca un universo de viviendas más amplio que la tabla de personas ampliada, ya que esta última considera a las viviendas ocupadas con moradores presentes. A los efectos de no exponer a los usuarios a posibles estimaciones o inferencias incorrectas, esta tabla no cuenta con una variable clave que permita su vinculación con la tabla de personas ampliada.

La tabla de datos de viviendas contiene las siguientes variables:

- Geográficas (departamento, localidad, sección, segmento[5], área, región, municipio, barrio y centro comunal zonal de Montevideo)
- Tipo de vivienda (particular o colectiva)
- Tipo de vivienda particular (casa, apartamento en edificio en altura con ascensor, etc.)
- Tipo de vivienda colectiva (Hotel, pensión u otra casa de hospedaje, etc.)
- Condición de ocupación de la vivienda (ocupada o desocupada)
- Estado de ocupación de la vivienda (morador presente o morador ausente).
- Estado de desocupación de la vivienda (uso temporal, alquiler o venta, construcción, etc.).

Cuadro 3 - Recuento sin ponderar de viviendas por tipo, según condición de ocupación

Viviendas censadas	Particulares	Colectivas	Total
Ocupadas con moradores presentes	1.243.794	2.560	1.246.354
Ocupadas con moradores ausentes	87.710	1.866	89.576

[5] Ver capítulo Anonimización

Total ocupadas	1.331.504	4.426	1.335.930
Desocupadas	315.114	8.000	323.114
Recuento total de viviendas	1.646.618	12.426	1.659.044

4) Ponderación de los microdatos censales

4.1. Omisión censal y uso de los ponderadores

El Censo 2023 presentó un nivel de omisión del 10,3%. Dado que la omisión censal se concentra en segmentos específicos de población, el uso directo de los hogares y personas efectivamente censados, sin aplicar correcciones o ajustes adicionales, conduce a estimaciones sesgadas. Por esta razón, luego de análisis e intercambios con actores académicos y organismos internacionales, el INE decidió incorporar ponderadores a los microdatos censales.

El objetivo de la estrategia es enriquecer los microdatos censales, mediante ponderadores que permitan estimaciones con el menor sesgo posible para todas las variables del censo, y alcanzar conteos consistentes con la estimación de personas residentes al 31 de mayo de 2023, (3.499.451 personas), por sexo, edad y departamento. Los ponderadores fueron confeccionados a nivel de hogar (a nivel de personas para los casos en situación de calle), permitiendo así la expansión de personas y de hogares, en la tabla de personas ampliada.

Las viviendas clasificadas como ocupadas con moradores ausentes y desocupadas quedan por fuera del ajuste por ponderación, por ende, no se puede estimar la magnitud de su omisión censal y así realizar inferencias sobre el universo de viviendas en el país. Vale aclarar que, al expandir el recuento de viviendas ocupadas con moradores presentes (a partir del uso de ponderadores en la tabla de personas ampliada), se obtiene una estimación del total de viviendas ocupadas, corrigiendo la omisión de hogares en viviendas que habían quedado asociadas a viviendas ocupadas con moradores ausentes.

4.2. Conceptos básicos para la estimación con datos ponderados

Las siguientes secciones 4.2, 4.3 y 4.4 provienen del documento **Metodología de Ponderación del Censo 2023,** cuya lectura se recomienda.

4.2.1. Interpretación de los ponderadores

A diferencia de un censo tradicional donde cada registro representa exactamente a una persona, en el Censo 2023 cada hogar y persona tiene asignado un ponderador (o peso) único. Es fundamental que se comprenda la naturaleza técnica de estos valores antes de procesar los microdatos.

El ponderador indica a cuántas unidades de la población total representa cada registro de la base de microdatos. Debido a los procesos de ajuste por no respuesta y calibración mediante distancias chi-cuadrado (o estimadores de regresión), estos valores presentan dos propiedades particulares:

1. **Valores con decimales:** Los ponderadores son el resultado de la aplicación de modelos de propensión y calibración, por lo que no son números enteros. Esto implica que las sumas de personas o características (ej. total de personas con nivel educativo terciario) resultarán habitualmente en valores decimales (ej. 1,240.7 personas). Estos decimales no indican "fracciones de personas", sino que son el residuo matemático de distribuir el peso de la omisión entre los hogares efectivamente censados.

2. **Ponderadores menores a uno:** es frecuente encontrar una proporción importante de registros con pesos menores a la unidad (ej. 0.85). Si bien esto puede resultar contraintuitivo, ya que implica que una unidad "no se representa íntegramente a sí misma", ello no afecta las propiedades estadísticas de los estimadores. Esta situación ocurre porque ciertos grupos (por ejemplo, determinados tramos de edad y sexo) quedaron sobrerrepresentados en el operativo de campo. Para corregir este sesgo y reproducir fielmente la estructura poblacional, el proceso de calibración debe asignar a estos grupos ponderadores menores a la unidad. Un fenómeno derivado de lo anterior es la existencia de segmentos censales donde la población ponderada (expandida) resulta inferior a la cantidad de personas efectivamente contadas. Aunque esto puede parecer un error de procesamiento, es un resultado estadísticamente consistente con el método de calibración adoptado.

Es importante destacar que el uso de estos ponderadores es teóricamente preferible y metodológicamente necesario. Las estimaciones expandidas, aun cuando resulten inferiores al conteo directo en ciertos segmentos, poseen un menor sesgo que los datos crudos, ya que han sido ajustadas para reflejar la verdadera estructura de la población. Ignorar estos pesos y optar por el conteo directo no solo mantiene los sesgos de sobrerrepresentación, sino que generaría un problema de benchmarking insalvable: la suma de los datos de estos dominios no coincidiría con los totales poblacionales de referencia. Por lo tanto, el uso del ponderador garantiza que el análisis sea estadísticamente válido y plenamente coherente con el marco demográfico del Censo 2023.

El sistema de ponderación desarrollado provee la base necesaria para la mayoría de los análisis sociodemográficos del país. Sin embargo, en investigaciones que requieran aperturas territoriales muy pequeñas o temáticas de muy baja prevalencia, la ponderación puede no ser suficiente para brindar estimaciones estables. En esos casos, se recomienda a

los usuarios evaluar la precisión de las estimaciones mediante el coeficiente de Variación (CV) y, de ser necesario, recurrir a niveles de agregación geográfica o temática superiores para garantizar la robustez de las conclusiones.

4.2.2. Consecuencias en el procesamiento y redondeo

Estas características exigen un manejo cuidadoso de los datos para garantizar la coherencia interna y la consistencia con los totales oficiales del INE:

● **No redondear antes de tiempo:** Todos los cálculos deben realizarse utilizando el valor completo del ponderador con todos sus decimales. Redondear los pesos individualmente antes de realizar las sumas destruiría la precisión del ajuste y la coherencia demográfica.

● **Gestión de la visibilidad y preservación de totales:** Dado que existen ponderadores pequeños, un redondeo tradicional podría "invisibilizar" celdas con pocas observaciones. Por este motivo, el INE propone un criterio de redondeo específico (detallado en la sección siguiente) que asegura que cualquier categoría observada en la muestra tenga una representación mínima de 1 en las tablas finales, preservando simultáneamente la suma total de la población.

4.3. Procedimiento de redondeo y preservación de totales

Como se detalló anteriormente, el uso de ponderadores con decimales genera totales que no son enteros. Para la publicación de resultados y para asegurar la concordancia con los totales de control (benchmarking), se adopta un procedimiento de redondeo que garantiza que la suma de las categorías coincida exactamente con el redondeo de la suma total, evitando discrepancias en las tablas finales.

Para ello, el INE utiliza un algoritmo basado en el reparto de residuos mayores, al cual se le añade una restricción de “visibilidad estadística”. Este criterio asegura que cualquier grupo observado en los microdatos mantenga su presencia en los resultados finales. La lógica del algoritmo es la siguiente:

1. **Garantía de representación mínima:** El primer paso del algoritmo fuerza que, si una categoría tiene al menos una observación en los microdatos (independientemente de que su peso ponderado sea menor a uno), el valor resultante sea al menos 1. Esto evita que grupos o categorías con muy baja prevalencia desaparezcan de las estimaciones por un efecto puramente matemático del redondeo.

2. **Ajuste por residuos:** Una vez asegurado el mínimo, el algoritmo calcula la diferencia entre la suma de los valores truncados y el total poblacional esperado. Esa diferencia (faltante) se

distribuye sumando una unidad a aquellas categorías que presentaban los mayores residuos decimales antes del truncamiento.

Para aplicar este criterio de forma reproducible, se proporciona la siguiente función en R:

```
redondear_censo <- function(x, digits = 0) {
 up <- 10 ^ digits
 x <- x * up
 y <- ifelse(x > 0 & x < 1, 1, floor(x))

 #Cálculo de la diferencia respecto al total esperado (benchmarking)
 diferencia <- round(sum(x)) - sum(y)

 # Distribución de la diferencia entre los residuos más altos
 if (diferencia > 0) {
  indices <- tail(order(x - y), diferencia)
  y[indices] <- y[indices] + 1
 }
 y / up
}
```

Este procedimiento garantiza que las tablas procesadas por los usuarios sean coherentes con los totales demográficos oficiales y que no se pierda información de celdas con baja frecuencia pero presencia real en el territorio.

4.4. Cómputo de los errores estándar en R

Para facilitar el uso de los microdatos del censo y garantizar la precisión de las estimaciones bajo el método descrito en el capítulo 6 *Medidas de la calidad de las estimaciones* del documento Metodología de Ponderación del Censo 2023, se recomienda el uso del paquete srvyr (o survey) en el entorno R.

Para procesar correctamente el diseño muestral equivalente del Censo 2023, se debe utilizar el siguiente bloque de código:

1) Como primer paso carga los paquetes/librerías necesarias:

```
library(srvyr)

library(survey)
```

2) Luego, se realiza la configuración técnica para grandes bases de datos. Para ello, se activa el método de último conglomerado y se evita la validación exhaustiva de estratos que causa errores de memoria en muestras masivas.

```
options(survey.lonely.psu = "remove",

    survey.ultimate.cluster = TRUE)
```

3) Se realiza el siguiente preprocesamiento:

```
personas_ext = personas_ext |>

 mutate(ID_HOGAR=ifelse(SIT_CALLE==1,

                paste('C',ID_CENSO, sep=''),

                ID_HOGAR),

     ESTRATO=ifelse(SIT_CALLE==1,'CALLE',ESTRATO)) |>

 filter(is.na(ID_HOGAR)==FALSE)
```

4) Se define el diseño muestral equivalente en donde lo SEs son computados utilizando el método del último conglomerado[6], donde **ID_HOGAR** es el Identificador del hogar (i.e. unidad de conglomeración), **ESTRATO** (i.e. el segmento censal), **TR** es la tasa de muestreo para realizar la corrección por población finita a nivel de estrato al momento de computar los SEs y **W** es el ponderador.

[6] Ver documento: Metodología Ponderación censo 2023 - Capítulo 6 - Medidas de la calidad de las estimaciones.

```
censo_diseno <- microdatos |>
 as_survey_design(
  id = ID_HOGAR,
  strata = ESTRATO,
  fpc = TR,
  weights = W,
  nest = TRUE
          )
```

Por otra parte, la opción **survey.ultimate.cluster = TRUE** permite que el cálculo se realice de forma eficiente, capturando la variabilidad entre las unidades primarias de muestreo (hogares) para aproximar la varianza total del diseño sin comprometer la capacidad de procesamiento del sistema.

5) Una vez cargado el diseño (**censo_diseno**) se realizan las estimaciones puntuales y se le pueden agregar medidas de calidad de las estimaciones. Por ejemplo, se puede calcular el total de una variable de interés y computar el error estándar (**se**), el coeficiente de variación (**cv**) y el intervalo de confianza (**ci**), el cual, por defecto es realizado al 95%[7].

```
estimacion <- censo_diseno |>
 summarise(
  total_indicador = survey_total(variable_interes,
                                 vartype = c("se",
                                             "cv",
                                            "ci"
                                             )
                                 )
```

[7] Para obtener detalles adicionales sobre las distintas funciones de srvyr, el manejo de diseños complejos y la visualización de resultados bajo un enfoque reproducible, se recomienda consultar el libro (Zimmer, et. al., 2024), el cual proporciona una guía detallada para el procesamiento de microdatos utilizando la gramática de tidyverse, facilitando la transición desde el análisis estadístico tradicional hacia flujos de trabajo más eficientes en R.

)

4.5. Comparabilidad entre censos

Si bien la comparación de cifras entre censos suele ser una práctica habitual, esta no está exenta de problemas. Los operativos de campo presentan distintos niveles de cobertura, con omisiones diferenciales según nivel socioeconómico y área geográficas, que no fueron sometidos a procesos de ajustes en etapas posteriores. Por lo tanto, esto ya es una complejidad para el análisis de variaciones, independientemente de las rondas censales que se comparen.

Para el Censo 2023 se recomienda trabajar siempre con los microdatos ponderados, a los efectos de mitigar los sesgos de la omisión y lograr conteos consistentes con las estimaciones de población. Por ese motivo, al analizar los resultados ponderados del Censo 2023 con censos anteriores, se debe advertir lo siguiente:

- Considerar el impacto del uso de ponderadores en el Censo 2023. Al comparar, se debe tener claro que las variaciones observadas en el tiempo pueden no responder estrictamente a un cambio real del tópico analizado, sino a una mejora en la calidad de la medición gracias a la aplicación de los ponderadores, dado que los censos anteriores no cuentan con ajustes que mitiguen los sesgos de la omisión.
- Evitar la comparación en números absolutos. Teniendo en cuenta lo mencionado en el párrafo anterior, en caso de aplicar los ponderadores en el Censo 2023, se sugiere evitar las comparaciones entre censos a partir de cifras absolutas. Se recomienda, en cambio, utilizar valores relativos como estrategia para minimizar los efectos del ajuste censal en el análisis de variaciones en el tiempo.
- Considerar que las innovaciones metodológicas del Censo 2023 pueden representar un desafío para la comparabilidad. La introducción de nuevas modalidades de recolección, como el CAWI, implica cambios en cómo los usuarios interactúan con el cuestionario. Estos avances tecnológicos, pueden alterar la distribución de las variables, lo que debe ser advertido a los usuarios.

Por esta razón se disponibiliza la variable **RELEVAMIENTO** que identifica la modalidad del registro.

5) Guía de variables y uso de identificadores en los microdatos censales

A continuación se detalla un conjunto de variables relevantes para una mejor comprensión de los microdatos. Para conocer el nombre y las categorías de todas las variables que conforman la boleta censal se debe acceder al diccionario de variables o al cuestionario.

5.1. Identificadores y estructura de los microdatos

DIRECCION_ID: Es la variable clave, en la tabla de vivienda, que vuelve única cada dirección.

VIVID: Código correlativo de identificación de la vivienda dentro de la dirección. Asume valor 1 excepto cuando en la dirección coexisten una vivienda particular y una vivienda colectiva[8].

HOGID: Código correlativo de identificación del hogar dentro de la vivienda.

ID_HOGAR: Código correlativo de identificación único del hogar.

PERID: Código correlativo de identificación de la persona dentro del hogar.

ID_CENSO: Código correlativo de identificación de la persona censada.

5.2. Variables analíticas y de clasificación

UNIVERSO: Identifica el componente del universo al cual pertenecen las personas. Esta variable asume tres valores:

- Particulares
- Colectivos
- Situación de Calle

[8] Sucede en contadas ocasiones cuando dentro de las viviendas colectivas ocupadas, existen habitaciones separadas donde reside en régimen familiar personal administrativo o de servicio del establecimiento, conformando un hogar particular. Ejemplo: en un residencial de ancianos vive el propietario con su familia y por lo tanto no forma parte del hogar colectivo.

CUESTIONARIO_COMPLETO: Identifica a las personas a las que se les aplicó el cuestionario completo.

CUESTIONARIO_BASICO: Identifica a las personas a las que se les aplicó el cuestionario básico.

RELEVAMIENTO: Identifica la modalidad de relevamiento de las personas censadas.

AGREGADO_A_HOGAR_CENSADO: Identifica a las personas de hasta 15 años censadas por enumeración administrativa que fueron integradas a un hogar censado.

MADRE_PERID: Identifica a la madre del menor, referido a la variable **AGREGADO_A_HOGAR_CENSADO.**

W: Valor del ponderador.

6) Agregación y niveles de análisis

6.1. Nivel hogar

El procesamiento de datos a nivel de hogar desde la tabla de personas ampliada requiere la aplicación de filtros específicos. Para ello se recomienda la selección de un caso único por hogar (típicamente mediante el identificador de la persona de referencia PERID == 1) o el uso de funciones de agrupamiento basadas en la combinación de variables DIRECCION_ID, VIVID y HOGID. Para ello se puede usar alguno de los siguientes comandos (ejemplo con el paquete dplyr en R):

1) Selección de un caso único por hogar (se presentan dos ejemplos):

- filter (UNIVERSO %in% c(1:2) & PERID == 1) *# Se elige a las personas provenientes de hogares particulares y colectivos (no en situación de calle) y nos quedamos con un caso por hogar*

- filter(UNIVERSO %in% c(1:2)) %>% distinct(DIRECCION_ID, VIVID, HOGID, .keep_all = TRUE) *# Se elige a las personas provenientes de hogares particulares y colectivos (no en situación de calle) y luego se elige una observación por cada valor único de dirección, vivienda y hogar*

2) Agrupamiento basado en la combinación de variables:

   filter(UNIVERSO %in% c(1:2)) %>% group_by (DIRECCION_ID, VIVID, HOGID) *# Se elige a las personas provenientes de hogares particulares y colectivos (no en situación de calle) y luego se agrupa por los valores únicos de dirección, vivienda y hogar*

6.2. Nivel Vivienda

Análogamente, para el procesamiento de datos de viviendas ocupadas, los usuarios deben filtrar la tabla de personas ampliada seleccionando una observación única por cada combinación de DIRECCION_ID y VIVID, tomando el primer hogar censado (HOGID == 1). Ello permite pasar de una unidad de análisis basada en individuos a una basada en viviendas ocupadas, utilizando el ponderador correspondiente.

En el caso de viviendas con hogares múltiples, considerando que el ponderador fue realizado a nivel de hogares y que la información de infraestructura de la vivienda fue relevada únicamente a través del primer hogar censado, se recomienda asignar el ponderador del primer hogar.

Para ello, se puede usar alguno de los siguientes comandos

(ejemplo con el paquete dplyr en R):

1) Selección de un caso único por vivienda (se presentan dos ejemplos)::

- filter (UNIVERSO %in% c(1:2) & HOGID == 1 & PERID == 1) *# Se elige a las personas provenientes de hogares particulares y colectivos (no en situación de calle) y luego un caso por hogar (siempre del primer hogar de la vivienda (de tener más de uno)).*

- filter(UNIVERSO %in% c(1:2) & HOGID == 1) %>% distinct(DIRECCION_ID, VIVID, .keep_all = TRUE) *# Se elige a las personas provenientes de hogares particulares y colectivos (no en situación de calle) y que pertenezcan al primer hogar de la vivienda y luego se elige una observación por cada valor único de dirección, vivienda.*

2) Agrupamiento basado en la combinación de variables:

filter(UNIVERSO %in% c(1:2)) %>% group_by (DIRECCION_ID, VIVID) *# Se elige a las personas provenientes de hogares particulares y colectivos (no en situación de calle) y luego se agrupa por los valores únicos de dirección, vivienda (aquí quedan agrupados todos los habitantes de la vivienda independientemente de que habiten en el mismo hogar o no).*

La correcta aplicación de estos identificadores garantiza la coherencia entre los totales poblacionales y las unidades habitacionales registradas en el operativo censal.

7) Tratamiento de inconsistencias y variables imputadas

Con el fin de optimizar la calidad de los datos y asegurar la coherencia interna de la información, se implementaron procesos de imputación y recodificación en variables que presentaron inconsistencias lógicas en las respuestas obtenidas. Bajo el principio de transparencia metodológica, se mantienen las variables originales, siendo opción de los usuarios optar por una de las dos versiones (original o imputada), así como realizar análisis comparativos entre ambas[9].

Se han generado variables auxiliares identificadas con el sufijo _TIPO (por ejemplo, I_VIVVO05_TIPO) para distinguir el origen de cada registro:

- C (cuestionario censal): dato proveniente del relevamiento directo en campo
- I (imputación): dato ajustado mediante procedimientos de corrección

7.1. Origen del agua para beber y cocinar

La variable **VIVDV05** consulta sobre "origen del agua utilizada para beber y cocinar" (Pregunta 12 del cuestionario de vivienda).

La pregunta presenta las siguientes categorías de respuesta:

1.....OSE o red general
2.....Pozo surgente o protegido
3.....Pozo surgente o no protegido
4.....Aljibe
5.....Cachimba
6.....Arroyo, río
7.....Otro

Al examinar la distribución de esta variable se identificó una incidencia atípica de la categoría "Otro" en áreas geográficas donde la cobertura de la red general es universal, particularmente zonas urbanas de alta densidad poblacional,[10]

[9] Durante el proceso de actualización de los microdatos censales se detectó una cantidad de niñas y niños menores de 5 años mucho mayor a la esperada en la localidad de Arachania, departamento de Rocha, representando al 18% del total de residentes de dicha localidad. Esta proporción es mucho mayor a la del promedio departamental (5%) y de las localidades contiguas de La Aguada/Costa Azul (4%), La Paloma (6%) y La Pedrera (6%). Una revisión pormenorizada de los datos censales de Arachania permitió constatar que una buena parte de estos niños y niñas no formaban parte de los hogares particulares censados y fueron incluidos injustificadamente en el operativo de campo. Se recomienda a usuarios tener los recaudos correspondientes al momento de analizar los datos censales de esta localidad. Vale mencionar que su incidencia a nivel municipal y departamental no es significativa.

[10] "En la actualidad el 99% de la población nucleada del país accede al agua potable, las 24 horas, los 365 días del año". http://www.ose.com.uy/agua/agua-potable

Se infiere que este fenómeno responde al contexto de la **emergencia hídrica de 2023**, que desplazó la percepción del informante hacia el consumo de agua embotellada en lugar de la fuente de suministro de la vivienda.[11]

En consecuencia, se procedió a la recodificación de 13.562 registros hacia la categoría "OSE o red general" cuando la vivienda estaba ubicada en áreas urbanas o cuando existía infraestructura de conexión interna (el agua llega por cañería dentro de la vivienda). Como se mencionó anteriormente, se crea una nueva variable (**I_VIVDV05**) manteniéndose la original (**VIVDV05**).

7.2. Tipo de tenencia de la propiedad de la vivienda

La variable HOGTE02 consulta sobre el estatus de propietario de la vivienda, (Pregunta 2 del cuestionario de hogar).

La pregunta se aplicó a los hogares particulares que declararon ser propietarios de la vivienda, y presenta las siguientes categorías de respuesta:

1.....de la vivienda y el terreno, los está pagando
2.....de la vivienda y el terreno, ya los pagó
3.....solamente de la vivienda, la está pagando
4.....solamente de la vivienda, ya la pagó

Al comparar la distribución de esta variable con fuentes externas, específicamente con las estimaciones de la Encuesta Continua de Hogares (ECH), se detectaron dos tipos de inconsistencias que requirieron un proceso de ajuste:

- Elevado porcentaje de propiedad exclusiva de la vivienda: se observó una prevalencia atípica de hogares que declararon ser propietarios únicamente de la vivienda en zonas donde, según la ECH, esa modalidad es muy baja o incluso inexistente. Para mitigar esta inconsistencia, en viviendas clasificadas como "apartamento en altura" se procedió a la reasignación hacia la categoría de propiedad de vivienda y terreno (45.235 hogares).
- Elevado porcentaje de hogares propietarios del terreno en asentamientos irregulares: se detectó una proporción alta de hogares que declararon ese tipo de tenencia en áreas clasificadas como asentamientos por la Dirección Nacional de integración Social y Urbana (DINISU). Una vez identificada esta inconsistencia, se

[11] El espíritu de la pregunta "Origen del agua" fue hallar la cantidad de población que tiene acceso al agua potable proporcionada por el ente estatal OSE. Analizándose una posible causa para la cantidad de hogares con el origen "Otro" se cree que tal vez las personas al responder "El origen del agua utilizada para beber y cocinar", respondieron literalmente la pregunta e indicaron "Otro" cuando por ejemplo usan bidones de agua comerciales para beber y cocinar. En el año 2023, Uruguay se enfrentó a una grave crisis hídrica que provocó que en los hogares el agua potable tuviera cierto grado de salinidad (sobre todo en el área metropolitana) y que el aumento de la venta de bidones de agua aumentara considerablemente (Se promulgó la ley Nº 20159 en la cual se exoneró del IVA a la enajenación de aguas minerales y sodas dada la emergencia hídrica).

procedió a imputar la categoría "propietario solo de la vivienda" (contemplando si la está pagando o ya la pagó según corresponda) a los 24.997 hogares en esta condición.

Se genera una variable imputada (**I_HOGTE02**) manteniéndose la original (**HOGTE02**).

7.3. Evacuación del baño

La variable HOGSH03 consulta sobre la forma de evacuación del baño, (Pregunta 6 del cuestionario de hogar).

La pregunta se aplicó a los hogares particulares que disponen de baño en la vivienda, y presenta las siguientes categorías de respuesta:

1.....Red general o saneamiento
2.....Fosa séptica, pozo negro
3.....Entubado hacia el arroyo
4.....Superficie, hueco en el suelo u otro

Al examinar la distribución de la variable, se detectaron 1.058 hogares en el área rural, que declararon la evacuación del baño hacia "red general o saneamiento". Dado que la infraestructura de saneamiento es una característica exclusiva de algunas zonas urbanas, dicha respuesta se considera una inconsistencia.

Por lo tanto, para los hogares del área rural que reportaron acceso a la red general, se procedió a la recodificación automática hacia la categoría "fosa séptica, pozo negro"[12]. Este criterio se fundamenta en que el 97% de los hogares censados con respuesta válida en zonas rurales utilizan este sistema de evacuación.

Se genera una variable imputada (**I_HOGSH03**) manteniéndose la original (**HOGSH03**).

7.4. Ocupación

La variable ocupación imputada **I_PERAL06** se presenta a un dígito y logra una cobertura de codificación de 98,9% entre las personas censadas que estaban habilitadas a responder la pregunta. El proceso de generación de dicha variable parte de lo obtenido para la variable **PERAL06** que también se presenta a un dígito y que ya estaba presente en la versión de los microdatos anterior.

[12] Del total de hogares en zona rural que se tiene respuesta a la pregunta, 97% respondieron que la evacuación del baño es por fosa séptica, pozo negro.

La variable **PERAL06** sistematiza la ocupación principal de las personas a partir de la pregunta abierta "¿Qué tareas realiza en su trabajo?" (pregunta 38 del cuestionario de personas). Se define la ocupación como el conjunto de tareas desempeñadas por una persona, las cuales se clasifican según su complejidad y nivel de conocimientos requeridos. Para garantizar la comparabilidad estadística, las respuestas recolectadas fueron codificadas *ex post* según el Clasificador Internacional Uniforme de Ocupaciones CIUO-08.

El proceso de codificación automática de la variable **PERAL06** comprendió la implementación de un modelo de aprendizaje supervisado basado en algoritmos de Support Vector Machine (SVM). Este procedimiento permitió asignar códigos a tres dígitos con una precisión global del 79%, cubriendo aproximadamente el 82% de los registros.

No obstante, se identificó una proporción considerable de casos que, si bien no pudieron ser codificados a tres dígitos debido a la exigencia del algoritmo y la ambigüedad de las respuestas, sí podrían ser clasificados de forma confiable a un dígito, nivel de publicación oficial de la variable. Surge así la necesidad de mejorar la cobertura de codificación de la variable **PERAL06** al nivel de un dígito, para los casos no codificados con el modelo anterior, utilizando las descripciones originales provistas por los censados y otros atributos recolectados durante el censo.

Por lo tanto, para esta nueva versión de los microdatos se presenta una nueva variable **I_PERAL06**, generada mediante los siguientes procesos:

1. **Clasificación manual:** Previamente a la clasificación manual se realizó un proceso de normalización y estandarización de texto obteniéndose una lista consolidada de menos de 100 términos únicos. Una vez obtenida la lista definitiva de términos estandarizados, se identificaron varios términos tales que la categoría de un dígito a la que pertenecen es inequívoca, por lo que se llevó a cabo un proceso de clasificación manual, resolviendo el 49.6% de los casos remanentes.

2. **Clasificación automática**: Para los casos restantes, se entrenó un modelo predictivo de Random Forest utilizando como predictores otras variables sociodemográficas y laborales del censo, tales como: departamento de residencia (DEPARTAMENTO), sexo (PERPH02), nivel educativo (NIVELEDU), rama de actividad (PERAL07), categoría de la ocupación (PERAL08), lugar de trabajo (PERAL09). El modelo final alcanza una precisión global del 53.7%, valor que mejora ampliamente la tasa de azar (25.9%).

Los resultados obtenidos fueron contrastados con la distribución del tipo de ocupaciones a un dígito de la ECH 2023, constatándose una mejora sustancial en la cobertura y precisión de la clasificación.

Los usuarios pueden optar por trabajar con la variable original (**PERAL06**) o con la variable imputada (**I_PERAL06**). Además se agrega la variable **I_PERAL06_TIPO** que indica si el valor de la **I_PERAL06** parte de la clasificación inicial (svm), de la clasificación manual (manual) o del modelo predictivo (rf).

7.5. Tratamiento de hogares con ausencia de adultos

En la versión de microdatos publicada en febrero de 2025, se identificaron 142 hogares unipersonales cuyo jefe/a o referente era menor de 12 años (llegando a 558 casos si se considera a menores de 18 años). Esta inconsistencia, detectada a pesar de las reglas de consistencia en los dispositivos de captura (CAPI/CAWI), derivó principalmente del proceso de depuración de registros duplicados, el cual supuso, en algunos casos, la eliminación de los adultos del hogar pero la permanencia de las personas menores[13].

Ante esta situación, y con el objetivo de preservar la estructura por edad de la población sin introducir sesgos por imputación de edad, se optó por una reubicación aleatoria de los menores, bajo los siguientes criterios:

- Proximidad geográfica: asignación de los menores a hogares receptores ubicados en la misma zona geográfica
- Elegibilidad de hogares receptores: se seleccionaron como receptores únicamente hogares particulares con al menos dos integrantes y viviendas con un único hogar
- Ajuste de la variable de parentesco: en los hogares receptores se incrementó la variable de número de personas y se asignó la categoría "Ignorado" en la relación de parentesco con la persona de referencia del hogar
- Condición de ocupación de la vivienda original: las viviendas que figuraban como lugar de residencia de estos hogares unipersonales fueron recodificadas a la condición de “Desocupada – Ignorada”.

7.6. Ajuste en la georreferenciación de viviendas colectivas

Durante el proceso de revisión de los microdatos censales, se identificaron 16 viviendas colectivas (principalmente centros de reclusión) cuya ubicación geográfica original presentaba inconsistencias. Este error afectaba a unas 10.500 personas censadas, impactando en el recuento poblacional a nivel de localidad y, en casos específicos, a nivel departamental. Para subsanar estas discrepancias, se procedió a realizar una reasignación geográfica de estas personas: en el 85% de los casos, las personas fueron reasignadas de una localidad urbana a un paraje rural dentro del mismo departamento; un 11% pasó del área rural a una localidad urbana. En situaciones en las que la vivienda colectiva se encontraba asignada a un departamento incorrecto (correspondiente al 4% de los casos o 464 personas), se realizó la reclasificación al departamento correcto. Dada la baja incidencia relativa, este ajuste no altera significativamente las estimaciones de población departamental. A continuación, se presenta la tabla con el detalle de las correcciones

[13] Ver: Documento Metodológico Censo 2023, Sección 14.1 Sobreenumeración

realizadas, facilitando a los usuarios la identificación de los cambios entre las versiones de los microdatos:

Cuadro 4 - Ajuste en la georreferenciación de viviendas colectivas

		Microdatos 02/2025		Microdatos 04/2026	
Tipo de Vivienda Colectiva	Cantidad de Personas	Código de Departamento	Código de Localidad	Código de Departamento	Código de Localidad
Centro de Reclusión	5	01	01020	01	01900
Centro de Reclusión	670	01	01020	01	01900
Centro de Reclusión	4.675	01	01020	01	01900
Centro de Reclusión	51	01	01900	01	01020
Centro de Reclusión	904	01	01900	01	01020
Centro de Reclusión	239	05	05321	04	04900
Centro de Reclusión	199	01	01900	05	05900
Centro de Reclusión	26	01	01020	07	07320
Centro de Reclusión	167	08	08900	08	8220
Centro de Reclusión	874	10	10021	10	10900
Centro de Reclusión	673	13	13823	13	13900
Centro de Reclusión	83	16	16923	16	16900
Centro de Reclusión	1.138	16	16421	16	16900
Centro de Reclusión	288	17	17220	17	17900
Escuela Militar	10	03	03880	03	03631
Centro de Salud	479	16	16923	16	16900

7.7. Asistencia a un centro educativo

La variable **ASISTENCIA** fue construida a partir de la pregunta del cuestionario censal **PERED01**: ¿Asiste actualmente a un establecimiento de enseñanza preescolar, primaria, secundaria, superior o técnica?, correspondiente al módulo de educación, vinculada a la concurrencia actual a centros educativos. Para su elaboración se realizó un proceso de consistencia y validación de la información declarada, verificando la coherencia entre las respuestas asociadas a asistencia, nivel educativo, grados aprobados reportados en el ciclo y edad de la persona, con el objetivo de minimizar inconsistencias derivadas de errores de respuesta.

En primera instancia, se identificaron los casos con declaración de asistencia a un centro educativo en la pregunta **PERED01** clasificándolos en la categoría “Asiste”, mientras que aquellos que declararon no estar asistiendo a un centro educativo o que nunca asistieron se los clasifica en la categoría “No asiste”. Posteriormente se cotejaron con fuentes externas, como Registros Administrativos, aquellas respuestas inconsistentes o con información faltante. Para el caso de los niños provenientes de RRAA que fueron incorporados a hogares censados (niños de 0 a 15 años), se recuperó la información de asistencia existente en Registros Administrativos. Finalmente, los casos donde no fue posible determinar la

condición de asistencia a partir de las fuentes disponibles, fueron clasificados en la categoría "Sin dato".

Una vez obtenida la variable **ASISTENCIA**, se cotejó su distribución con la Encuesta Continua de Hogares (ECH 2023) a fin de detectar posibles desviaciones en la distribución. La variable resultante constituye una variable agregada elaborada a partir de la integración y depuración de las distintas variables que componen el módulo de educación aplicado en el relevamiento censal.

8) Variables agregadas

8.1. Discapacidad

En cumplimiento con las recomendaciones internacionales para la comparabilidad estadística, el relevamiento de la discapacidad en el Censo 2023 se basó en el set de preguntas propuesto por el Grupo de Washington (GW). Estas variables recogen las limitaciones en seis dominios fundamentales básicos: visión, audición, movilidad, comunicación, cognición y autocuidado.

En cada uno de los dominios se registró el grado de la dificultad para realizar la actividad, siendo las opciones de respuesta:

1..... No tiene dificultad
2..... Tiene alguna dificultad
3..... Tiene mucha dificultad
4..... No puede hacerlo o le resulta imposible

Dada la estructura del módulo (cuatro categorías de respuesta por dominio), se generó la variable agregada **DISC_TIENE** para facilitar el análisis de prevalencia según diferentes umbrales de severidad. La presencia de discapacidad está determinada por las personas que declaran tener "mucha dificultad" o "no puede hacerlo" en al menos uno de los seis dominios mencionados[14].

Si bien en el formulario censal el set de preguntas se realizó sin filtro de edad, se recomienda a los usuarios considerar a la población de 5 años o más para el análisis de los resultados, siguiendo las orientaciones del GW.

Asimismo, la variable **DIFICULTAD** indica para las personas de 5 años o más, el mayor grado de dificultad que presenta en el conjunto de los seis dominios[15].

[14] Por tanto, presentar "alguna dificultad" no se contabiliza como discapacidad.
[15] Ej. Si sólo en un dominio presenta dificultad severa y en los otros cinco no presenta dificultad, queda categorizado con el valor 4 (severa).

8.2. Departamento de residencia cinco años antes

Se generó la variable Departamento de residencia cinco años antes (**DEPTO_RES_2018**) a partir de la pregunta censal "¿En qué departamento vivía hace 5 años (2018)?". Esta pregunta se aplica exclusivamente a personas de 5 años o más que declararon haber migrado alguna vez y reportaron una antigüedad en la residencia actual menor a cinco años.

No obstante, la variable se genera y reporta para toda la población de 5 o más años, aplicando los siguientes criterios de imputación y asignación:

- **Personas menores de 5 años**: se asigna la categoría *No corresponde*

- **Personas que nunca migraron o personas migrantes con antigüedad de cinco años o más en el departamento de residencia actual:** el departamento de residencia cinco años antes se iguala al departamento de residencia actual (2023).

- **Personas migrantes con antigüedad menor a cinco años de duración en la residencia actual:** se asigna el departamento declarado como lugar de residencia cinco años antes, según lo informado en el módulo de migración.

- **Personas que residían en el exterior cinco años antes:** se asigna la categoría *Extranjero*, distinguiendo así la migración internacional reciente de la migración interna.

Los usuarios deben considerar que esta variable permite la construcción de matrices de origen-destino de la migración interna reciente.

8.3. Departamento de nacimiento

La variable Departamento de nacimiento (**DEPTO_NACIM**) identifica el lugar de nacimiento de la población, permitiendo el estudio de la migración interna de "toda la vida". Su construcción sigue una lógica análoga a la utilizada en la variable Departamento de residencia 5 años antes, integrando las respuestas del módulo de migración con procesos de validación y consistencia.

8.4. Total de personas en el hogar

La variable **HOGPR01** indica la cantidad total de personas que residen en un hogar (particular o colectivo) e incluye a niños, niñas y adolescentes provenientes de la

enumeración por registros administrativos que fueron vinculados a un hogar censado (ver sección 2.2).

8.5. Condición de actividad económica

La variable **POBCOAC** clasifica a la población de 12 años o más según su relación con el mercado de trabajo durante el periodo de referencia, y se construye a partir del módulo de preguntas censales sobre actividad laboral. Las categorías de esta variable se definen según los siguientes criterios:

- **Persona ocupada:** comprende a toda persona que, en la semana anterior al registro del formulario, trabajó por lo menos una hora a cambio de una remuneración (en dinero o especie)[16], o que, aunque no lo hizo, tiene algún trabajo o negocio al que seguro volverá.
- **Población desocupada:** comprende a las personas que no estaban ocupadas en la semana de referencia pero que realizaron acciones concretas de búsqueda activa de empleo o intentaron establecer un negocio propio durante las últimas cuatro semanas.
- **Población inactiva (jubilados o pensionistas):** personas que no participan en el mercado laboral y que declararon ser jubiladas o pensionistas.
- **Población inactiva (otras causas):** Incluye al resto de la población no económicamente activa que no se clasifica como jubilada ni pensionista (por ejemplo, estudiantes de tiempo completo, personas dedicadas exclusivamente a las tareas del hogar o rentistas).

8.6. MUNICIPIO

En esta nueva versión de los microdatos censales, la variable **MUNICIPIO** ha sido actualizada, en concordancia con las series electorales de marzo de 2025, aprobadas por la Corte Electoral en la Circular N° 12.208. Esta actualización implicó la creación de nuevos municipios (pasando de 125 a 136)[17] y la revisión de los límites de otros ya existentes en el país.

El proceso de asignación geográfica a nivel de vivienda se realizó mediante un geoprocesamiento riguroso, debido a que los límites municipales (de definición político-

[16] La definición de trabajo incluye además:
- Algún trabajo puntual y concreto a cambio de remuneración en dinero o en especie; por ejemplo changa, venta ambulante, alguna limpieza, etc.
- Colaboración o cuidado de animales, cultivos o huerta que no fueran para su consumo propio.
- Colaboración en un negocio familiar aún sin recibir remuneración.

Excluye: Quehaceres del hogar.

[17] Si bien se amplía el número de municipios, no todo el territorio nacional se encuentra municipalizado (existiendo departamentos parcialmente municipalizados y otros sin municipios).

electoral) no siempre coinciden exactamente con las unidades geoestadísticas utilizadas por el INE para fines censales, especialmente en áreas rurales. Los criterios aplicados fueron:

1. Compatibilización cartográfica: se utilizó la capa de municipios generada en el marco del Grupo de Trabajo de Límites Administrativos (GTLA), liderado por el IDE e integrado por la Corte Electoral, OPP, IGM, DINOT, e INE. Ello permitió asegurar una base cartográfica oficial y unificada.
2. Asignación por ubicación espacial: se contrastó digitalmente la ubicación de cada vivienda censada (utilizando puntos de edificación para áreas rurales y zonas censales para áreas urbanas) con la capa de polígonos de los municipios, asignando el municipio que contenía el punto geográfico de la vivienda.

8.7. Nivel educativo de la población

La variable **NIVELEDU** indica el mayor nivel educativo alcanzado por la persona al momento del relevamiento censal, con independencia de si dicho nivel fue completado o no. Esta variable agregada es el resultado de un algoritmo de consistencia que integra la edad (**PERNA01**), la asistencia actual (**PERED01**), el nivel que cursa actualmente (**PERED03**) y el nivel más alto que cursó anteriormente (**PERED03_1**).

El algoritmo prioriza la información de "nivel más alto cursado" para quienes no asisten, y el "nivel actual" para quienes se encuentran dentro del sistema educativo, asignándose de forma automática la categoría "Menor de 4 años" para la población de esa edad, dado que para este grupo no se relevaron las variables educativas.

En la presente versión de los microdatos se presenta la variable **NIVELEDU25MAS** que indica el mayor nivel alcanzado para la población de 25 años o más.

9) Anonimización

El Secreto Estadístico garantiza que la información recopilada por el INE será difundida preservando el anonimato de la fuente, tal como lo establece la Ley 16.616. Dicha normativa obliga a tratar los datos individuales con absoluta confidencialidad, impidiendo que, a partir de una información o combinación de informaciones, se vulnere el secreto estadístico y logren identificar individuos específicos.

El archivo de microdatos del Censo 2023 ha sido sometido a un proceso de disociación de datos personales (anonimización) para cumplir con este mandato legal.

Los métodos de protección implementados se basan fundamentalmente en la agregación, recodificación de datos y supresión de variables, sin recurrir a la perturbación de los valores originales.

En consideración de lo anterior, se han aplicado las siguientes restricciones a los microdatos de difusión pública:

- **Unidades geoestadísticas**:

  - Se elimina la variable **ZONA.**

  - En el caso de la variable **SEGMENTO**, solamente para los segmentos con población menor a 20 habitantes, se estableció un procedimiento por el cual tales segmentos fueron absorbidos por segmentos receptores (segmentos contiguos o próximos con suficiente población y características geográficas, sociales e infraestructurales comparables). Una vez agrupados, todos los registros de tales segmentos y de los receptores comparten un nuevo identificador común, reduciendo así el riesgo de que la localización geográfica opere como variable de identificación indirecta en la base de microdatos publicada[18]. Lo mismo opera para la variable **SECCION** que para esta nueva versión de los microdatos se levanta la restricción de su publicación, pero se presenta agrupada cuando la población es menor a 20.

  - **LOCALIDAD**: Las localidades con población menor a 20 habitantes se presentan agrupadas junto al área rural del departamento correspondiente[19].

  - Se publican tres variables dummies, **SEGMENTO_AGRUP**, **SECCION_AGRUP** y **LOCALIDAD_AGRUP** que indican si los datos fueron agrupados o no.

- **Variables sensibles:** Se suprimen las variables relativas al documento de identidad[20], a la identidad de género[21] y al día y mes de nacimiento[22].

- **Lugar de nacimiento, de residencia anterior y de residencia 5 años antes (2018)**: Las variables lugar de nacimiento (**PERMI01**), lugar de residencia anterior (**PERMI06**) y lugar de residencia 5 años antes (2018) **PERMI07** se presentan agrupada en tres categorías (en este departamento, en otro departamento, en otro país), no presentándose la información a nivel de localidad.[23]

- **Lugar de trabajo:** Se eliminan las variables **PERAL09_1** y **PERAL09_3** relacionadas a la localidad del lugar de trabajo.

---

[18] Dicho proceso fue conducido de manera conjunta por el Departamento de Metodología, el Departamento de Estadísticas Sociodemográficas y el Departamento de Infraestructura Geoestadística del INE de Uruguay.
[19] Existen dos localidades que, si bien tienen menos de 20 habitantes se decidió no agrupar con un área cercana, dado que implicaba asociarlas a un departamento distinto.
[20] Se suprimen todas las variables del módulo **Tipo de Documento de Identidad**.
[21] Se suprimen las variables **PERGE01** y **PERGE01_1.**
[22] Se suprime la variable **PERNA02**, presentándose únicamente el año de nacimiento **PERNA02_ANO**
[23] En el caso de las variables **PERMI01** y **PERMI07**, la opción 2 “en otra ciudad o localidad del departamento” pasa a estar agrupada con la opción 1 “en ésta localidad” conformando la categoría “en este departamento” y además no se presentan las variables codificadoras de las localidades: **PERMI01_1**, **PERMI01_3**, **PERMI06_1**, **PERMI06_3**, **PERMI07_1**, **PERMI07_3**.

- **Fecha de nacimiento del último hijo:** Día y mes de nacimiento del último hijo se eliminan (**PERFM04_3, PERFM04_2**), sin embargo se levanta la restricción para el año de nacimiento que se presenta en años simples **PERFM04_1.**

- **Fecha de nacimiento del primer hijo:** Mes de nacimiento del primer hijo **PERFM05_2** no se presenta, sin embargo se levanta la restricción para el año de nacimiento que se presenta en años simples **PERFM05.**

Además, en esta nueva versión de los microdatos se decidió publicar las siguientes variables:

- Barrios de Montevideo **BARRIO85**
- Centros Comunales de Montevideo **CCZ**
- Edad se presenta en edades simples para toda la población censada **PERNA01**
- Año de Nacimiento **PERNA02_ANO**

10) Versiones publicadas.

Se disponibiliza la segunda versión de los microdatos anonimizados denominados:

- **personas_ext_05_2026.csv** (versión ampliada con datos de viviendas, hogares, personas y ponderadores).
- **viviendas_ext_05_2026.csv** (versión sin ponderar, proveniente del recuento realizado en operativo de campo).

Asimismo, se publican las tablas análogas etiquetadas en formato R.Data.

- **personas_ext_05_2026_etiquetada**
- **viviendas_ext_05_2026_etiquetada**

Por último, está planificado disponibilizar a corto plazo el nivel educativo alcanzado para toda la población y el resto de las variables del capítulo educación.

La versión anterior a microdatos (febrero del 2025) se encuentra alojada en la sección “Materiales históricos”.

A medida que se continúen enriqueciendo los microdatos del Censo 2023, se divulgarán y anunciarán nuevas versiones.

11) Anexos

Cuadro 5 - Construcción de la variable POBPCOAC

Edad (PERNA01)	34 - Sin considerar los quehaceres del hogar, trabajó la semana pasada (PERAL01)	35 - Durante la semana pasada, al menos 1 hora: (PERAL13)	36 - Aunque no trabajó la semana pasada, tiene algún trabajo o negocio al que seguro volverá (PERAL03)	37 - Durante las últimas 4 semanas, estuvo buscando trabajo o tratando de establecer su propio negocio (PERAL04)	42 - Es jubilado/a o pensionista (PERAL11)	POBPCOAC
Menor de 12 años	-	-	-	-	-	1 - Menor de 12 años
12 años o más	1 - Sí	-	-	-	-	2 - Ocupado
12 años o más	2 - No	1 - Realizó algún trabajo puntual y concreto a cambio de remuneración en dinero o en especie; por ejemplo changa, venta ambulante, alguna limpieza, etc.	-	-	-	2 - Ocupado
12 años o más	2 - No	2 – Colaboró en el cuidado de animales, cultivos o huerta que no fueran para su consumo propio	-	-	-	2 - Ocupado
12 años o más	2 - No	3 - Colaboró en un negocio familiar aún sin recibir remuneración	-	-	-	2 - Ocupado
12 años o más	2 - No	4 - No realizó actividades generadoras de ingresos, ni siquiera durante 1 hora	1 - Sí	-	-	2 - Ocupado
12 años o más	2 - No	4 - No realizó actividades generadoras de ingresos, ni siquiera durante 1 hora	2 - No	1 - Sí	-	3 - Desocupado
12 años o más	2 - No	4 - No realizó actividades generadoras de ingresos, ni siquiera durante 1 hora	2 - No	2 - No	1 - Sí	4 - Inactivo, jubilado o pensionista
12 años o más	2 - No	4 - No realizó actividades generadoras de ingresos, ni siquiera durante 1 hora	2 - No	2 - No	2 - No	5 - Inactivo, otras causas

Cuadro 6 - Construcción de la variable NIVELEDU[24]

Edad (PERNA01)	25 -Asistencia actual a establecimiento de enseñanza (PERED01)	27 - Nivel cursado actualmente (PERED03)	29 - Nivel educativo más alto que cursó (PERED03_1)	Máximo nivel educativo alcanzado (NIVELEDU)
Menor de 4 años	-	-	-	0 - Menor de 4 años
4 años o más	3 - No asiste pero asistió	-	1 - Inicial o Preescolar	12 - Nunca asistió
4 años o más	4 - Nunca asistió	-	-	12 - Nunca asistió
4 años o más	1 o 2 - Sí asiste, público o privado	1 - Inicial o Preescolar	-	1 - Inicial o Preescolar
4 años o más	1 o 2 - Sí asiste, público o privado	2 o 3 - Primaria común o especial	-	2 - Primaria común o especial
4 años o más	3 - No asiste pero asistió	-	2 o 3 - Primaria común o especial	2 - Primaria común o especial
4 años o más	1 o 2 - Sí asiste, público o privado	13 - Educación Media Básica o Ciclo Básico (Liceo o UTU)	-	4 - Educación Media Básica o Ciclo Básico
4 años o más	3 - No asiste pero asistió	-	13 - Educación Media Básica o Ciclo Básico (Liceo o UTU)	4 - Educación Media Básica o Ciclo Básico
4 años o más	1 o 2 - Sí asiste, público o privado	14 - Educación Media Superior o Bachillerato (Liceo o UTU)	-	5 - Educación Media Superior o Bachillerato
4 años o más	3 - No asiste pero asistió	-	14 - Educación Media Superior o Bachillerato (Liceo o UTU)	5 - Educación Media Superior o Bachillerato
4 años o más	1 o 2 - Sí asiste, público o privado	15 - Capacitaciones o cursos de UTU*	-	6 - Capacitaciones o cursos de UTU
4 años o más	3 - No asiste pero asistió	-	15 - Capacitaciones o cursos de UTU*	6 - Capacitaciones o cursos de UTU
4 años o más	1 o 2 - Sí asiste, público o privado	9 - Magisterio o profesorado	-	7 - Magisterio o profesorado
4 años o más	3 - No asiste pero asistió	-	9 - Magisterio o profesorado	7 - Magisterio o profesorado
4 años o más	1 o 2 - Sí asiste, público o privado	10 - Terciario no universitario	-	8 - Terciario no universitario
4 años o más	3 - No asiste pero asistió	-	10 - Terciario no universitario	8 - Terciario no universitario
4 años o más	1 o 2 - Sí asiste, público o privado	11 - Universidad o similar	-	9 - Universidad o similar
4 años o más	3 - No asiste pero asistió	-	11 - Universidad o similar	9 - Universidad o similar
4 años o más	1 o 2 - Sí asiste, público o privado	12 - Posgrado (Diploma/Maestría/Doctorado)	-	10 - Posgrado (Diploma/Maestría/Doctorado)
4 años o más	3 - No asiste pero asistió	-	12 - Posgrado (Diploma/Maestría/Doctorado)	10 - Posgrado (Diploma/Maestría/Doctorado)

* Que no acreditan Ciclo Básico ni Bachillerato

[24] Si bien en la presente edición de microdatos se presenta la variable NIVELEDU para la población de 25 o más, el cuadro indica la construcción de la variable para toda la población.