Red de Transmisión del Conocimiento (RTC)
Conferencia Estadística de las Américas, CEA-CEPAL
Construcción de ponderadores y omisión censal: experiencias metodológicas de Uruguay y Perú

Ponderar un censo como una muestra no probabilística

La experiencia de Uruguay con la omisión diferencial del Censo 2023

Juan Pablo Ferreira

Director, División Metodología e Investigación, Instituto Nacional de Estadística (INE), Uruguay

Trabajo conjunto con Juan José Goyeneche

PUNTO DE PARTIDA

En 2023, uno de cada diez residentes no fue censado

POBLACIÓN OBJETIVO DEL CENSO 2023 · N = 3.444.451



Áreas a escala

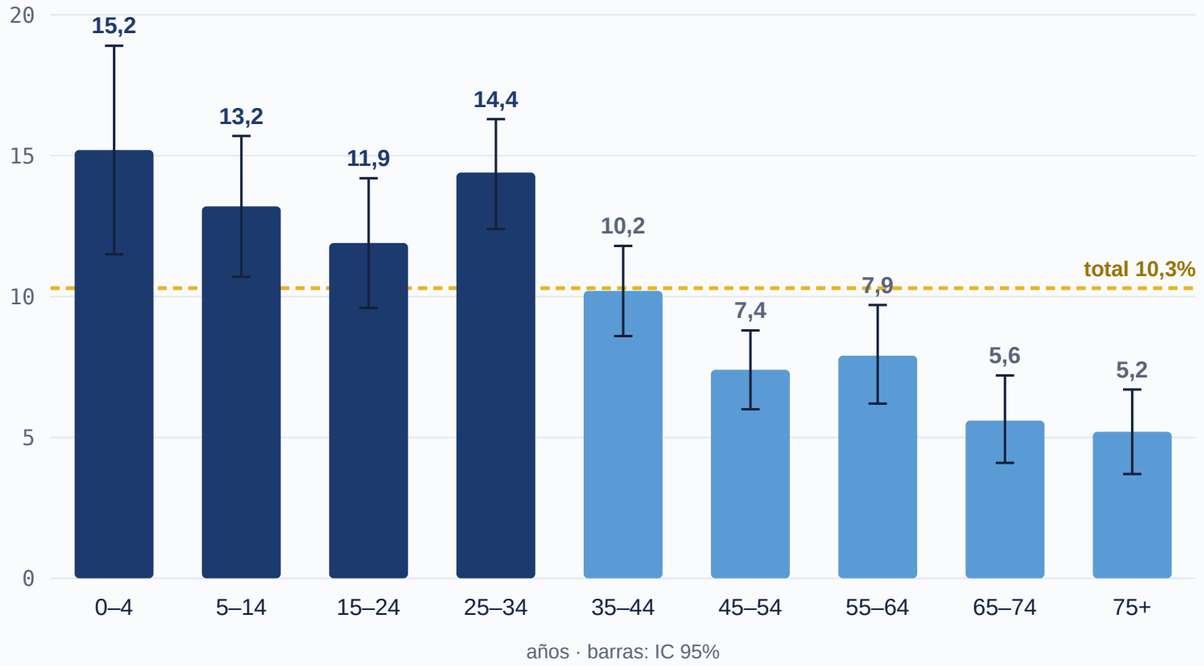
10,3%
omisión estimada por la Encuesta Nacional de Evaluación Censal (ENEC)

4,07%
omisión del censo 2011: un quiebre con la serie histórica

≈ 350.000
personas agregadas con registros administrativos («señales de vida»): el censo combinado

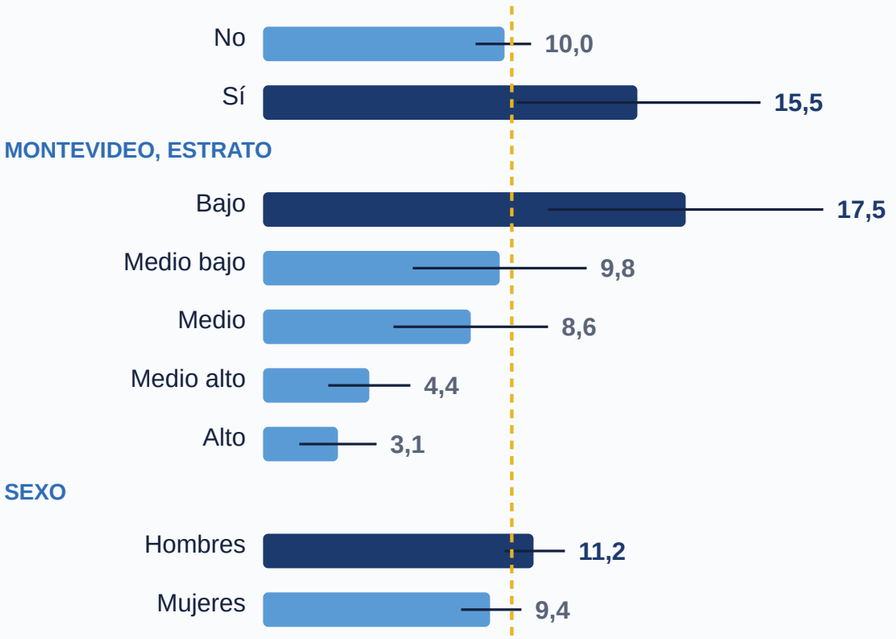
El problema no es el nivel de la omisión, sino dónde se concentra

OMISIÓN POR EDAD, %



años · barras: IC 95%

ASENTAMIENTO



MONTEVIDEO, ESTRATO

SEXO

$$\text{sesgo}\left(\bar{y}_{\text{censados}}\right) = W_M \left(\bar{Y}_{\text{censados}} - \bar{Y}_{\text{omitidos}}\right)$$

W_M: tasa de omisión. Si los omitidos son distintos, el sesgo no desaparece aunque el total sea correcto.

Los registros administrativos recuperan el conteo, pero no corrigen el sesgo

Lo que aportan

- Totales de población por sexo, edad y departamento.
- Un conteo coherente con las proyecciones demográficas.

Lo que no aportan

- Las ~100 variables del cuestionario: educación, vivienda, hogar.
- El hogar: la mayoría de las personas agregadas no queda vinculada a una vivienda.
- Una ubicación fina confiable: se asignan hacia centros urbanos (**sesgo de urbanicidad**).

¿Y la imputación hot deck?

Copia valores de hogares censados parecidos: supone que, dentro de cada clase, los censados representan a los omitidos.

El INE lo evaluó: las estimaciones con y sin los hogares imputados fueron **prácticamente iguales**. No corregía el sesgo.

Conclusión: ponderar a los hogares efectivamente censados, los únicos con información completa.

EL ENFOQUE

Tratamos a los hogares censados como una muestra no probabilística

No sabemos la probabilidad de ser censado: depende del hogar y del operativo. Por eso combinamos dos modelos en un **estimador doblemente robusto**.

1 Propensión de respuesta

¿Qué tan probable era censar a este hogar? Se estima por **segmento censal**.

factor = 1/propensión estimada del segmento

→

2 Calibración

Los pesos se ajustan para reproducir los totales del censo combinado por **sexo, edad simple y departamento**.

modelo implícito: y explicada por sexo, edad y departamento

=

Doble protección

El estimador es consistente si **alguno de los dos** modelos está bien especificado.

No protege si la omisión depende de factores que ninguno de los dos ve.

Con calibración lineal, el estimador doblemente robusto **coincide con un estimador calibrado**: por eso se implementa como un ajuste por propensión seguido de una calibración, con herramientas que cualquier oficina ya usa.

El estimador doblemente robusto en fórmulas

PROPENSIÓN

$$\hat{Y}_{\text{IPW}} = \sum_{i \in s} \frac{y_i}{\hat{\phi}_i}$$

Cada hogar censado representa a $1/\hat{\phi}$ hogares.

MODELO

$$\hat{Y}_{\text{MB}} = \sum_{i \in s} y_i + \sum_{i \in U \setminus s} \hat{m}(x_i)$$

Se predice a los omitidos con $m(x) = \mathrm{E}(y \mid x)$.

DOBLEMENTE ROBUSTO

$$\hat{Y}_{\text{DR}} = \sum_{i \in U} \hat{m}(x_i) + \sum_{i \in s} \frac{y_i - \hat{m}(x_i)}{\hat{\phi}_i}$$

Predicción para todos, más los residuos de los censados ponderados por $1/\hat{\phi}$.

POR QUÉ ES ROBUSTO

$$\mathrm{E}\big[B(\hat{Y}_{\text{DR}})\big] \approx \sum_{i \in U} \mathrm{E}\left[\frac{\phi_i}{\hat{\phi}_i} - 1\right] \mathrm{E}\big[y_i - m(x_i)\big]$$

Es el **producto** de los dos errores de especificación: se anula si cualquiera de los dos modelos es correcto. Si fallan ambos, queda sesgo.

CÓMO SE IMPLEMENTA

$$m(x) = x^{\top}\beta \;\Rightarrow\; \hat{Y}_{\text{DR}} = \sum_{i \in s} w_i\, y_i, \qquad w_i = \frac{g_i}{\hat{\phi}_c}$$

Con $\sum_{i \in s} w_i\, x_i = X$: es un **ajuste por propensión seguido de una calibración**. Aquí, $\hat{\phi}_c$ por segmento y $x = \text{sexo} \times \text{edad} \times \text{departamento}$.

ETAPA 1 · PROPENSIÓN

La tasa de vinculación web de cada segmento mide la dificultad de contacto

$$P(\text{responder}) = P(\text{contacto}) \times P(\text{cooperar} \mid \text{contacto}) \approx P(\text{contacto})$$

Censo obligatorio y pocos rechazos: la cooperación es casi total.

Todo hogar que respondió por internet debía ser visitado para vincular el cuestionario a su vivienda. En cada segmento, la **proporción de cuestionarios web vinculados** funciona como indicador de la probabilidad de contacto.

Segmentos con baja vinculación = más dificultad de contacto = más omisión.

4.507
segmentos censales

4.361
grupos, con mínimo de 20 hogares (se agrupan vecinos)

1,095
factor mediano: la mitad de los grupos se ajusta menos de 10%

2,5
tope del factor; actúa en 46 grupos (1,1% de la población)

Mover el tope entre 2 y 4 afecta a menos de 1,4% de la población.

La propensión de cada segmento sale de un captura–recaptura con los cuestionarios web

1. **Primera captura:** los hogares que respondieron **por internet**. Sabemos que existen: *n* cuestionarios web en el segmento.

2. **Segunda captura:** el censista debía visitarlos y verificar su código. De esos *n*, verificó *m*.

3. **Propensión del segmento:** $\hat{\phi}_c = m/n$, la probabilidad de que el censista llegara a un hogar. La omisión es $1 - \hat{\phi}_c$.

4. Si el grupo tiene menos de 20 casos, se agrupa con segmentos vecinos de la misma sección o departamento.

Segmento	Web *n*	Verificados *m*	$\hat{\phi}_c$	Factor $1/\hat{\phi}$
Costero, Montevideo	50	48	0,96	**1,04**
Periferia norte	50	35	0,70	**1,43**

Ejemplo ilustrativo. Todos los hogares censados del segmento reciben el mismo factor, con tope 2,5.

El supuesto

Dentro de un segmento, el censista tuvo la misma dificultad para llegar a un hogar web que a uno que no respondió por internet.

Por eso se trabaja en áreas chicas: aplicado al país entero, el método da ≈ 8% de omisión; por áreas chicas, ≈ 10–11%, como la ENEC.

El modelo por segmento reproduce lo que mide la encuesta de evaluación

OMISIÓN ESTIMADA POR SEGMENTO

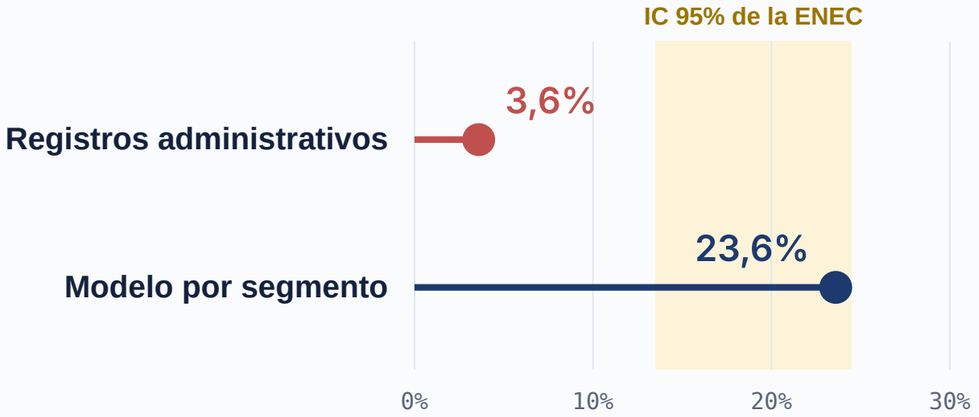


POBREZA POR ÁREAS PEQUEÑAS



r = 0,67 entre omisión y pobreza (Fay-Herriot)

OMISIÓN RURAL, PROMEDIO DE LOS DEPARTAMENTOS, %



Fuente: Ferreira y Goyeneche (2026), figura 6.

Coincide con la encuesta por estrato, edad, sexo y departamento. Los registros administrativos subestiman mucho la omisión rural.

Calibrar solo donde los totales de referencia son confiables

Totales de control

Censo combinado por **sexo × edad simple × departamento**, coherente con las proyecciones oficiales.

No se calibra por localidad: el error de asignación de los registros está en la geografía fina.

Un peso por hogar

Todos los integrantes comparten el peso: **método integrado** (Lemaître–Dufour), que promedia los controles de personas en el hogar.

Mantiene la coherencia hogar–persona.

Corrección por localidad

Antes de calibrar, ajuste de razón donde la localidad se apartaba de un total triangulado: censo 2011, precenso 2023, viviendas ocupadas y **medidores de UTE**.

137 de 677 localidades, todas pequeñas.

CONTROL DE CALIDAD

Pesos estables: la corrección es fuerte solo donde hacía falta

1,07
peso final mediano (rango 0,62 a 4,66)

1,02
efecto de diseño de Kish: 1,23 de 1,26 millones de hogares efectivos

0,99
factor de calibración mediano (rango 0,33 a 2,63)

0
pesos recortados: ninguno lo necesitó

Casos especiales

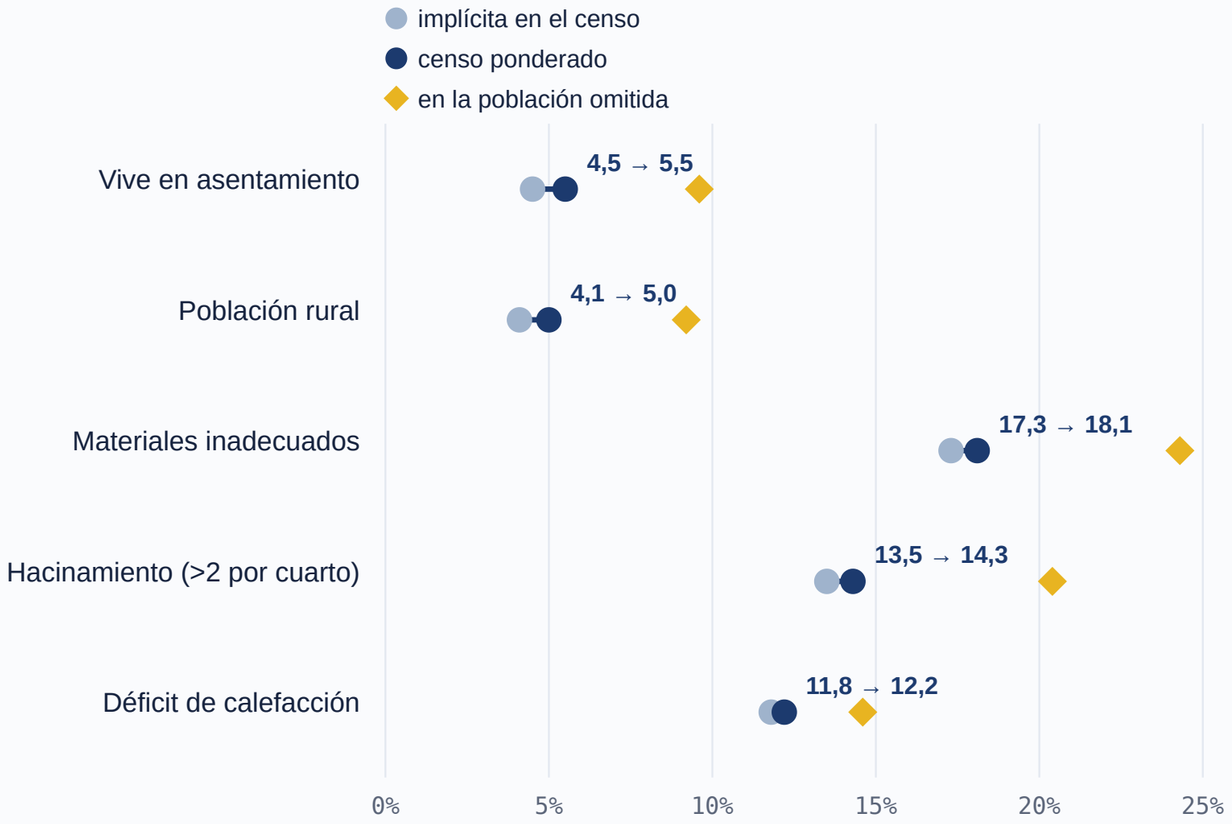
- **Peso 1** para personas en situación de calle y viviendas colectivas de 50 o más residentes: se representan a sí mismas.
- Pesos menores que 1 donde un grupo quedó sobrerrepresentado.

Controles

- Los totales calibrados se reproducen exactamente.
- Los factores más dispersos están en los 25 a 35 años: donde más omisión hubo.

RESULTADOS

Al ponderar, suben los indicadores de privación



La dirección y el tamaño de los cambios coinciden con el diagnóstico independiente de la encuesta de evaluación.

Para los usuarios, una regla: **toda estimación con los microdatos ponderados**; los conteos sin ponderar no se usan.

Microdatos ponderados publicados el 14 de mayo de 2026, con el documento metodológico de los ponderadores.

PRECISIÓN

Un diseño equivalente para que cualquier usuario calcule errores estándar

Diseño equivalente

- **Estratos:** segmentos censales.
- **Conglomerados:** hogares.
- **Corrección por finitud:** 1 − cobertura estimada del segmento.
- Varianza por **último conglomerado**, con los pesos finales.

Publicación por coeficiente de variación: excelente < 5%, … , no publicar ≥ 35%.

Una decisión operativa

Replicar re-estimando las dos etapas sería lo ideal, pero con más de 3 millones de registros es inviable para la producción y para los usuarios.

La aproximación **subestima** un poco la varianza: no es conservadora, y así se informa.

SUPUESTOS Y LÍMITES

Reduce el sesgo de omisión, no lo elimina

Supuesto MAR

Dentro del segmento y de las clases de calibración, ser censado no depende de las variables de interés. Si depende de algo que ningún modelo ve, queda sesgo.

El indicador web

Podría reflejar conectividad. En Uruguay el acceso a internet es casi universal (94% en el quintil más pobre, 98% en el más rico), pero la calidad de conexión y el equipamiento sí varían con el ingreso.

Los totales de referencia

Vienen del censo combinado y se tratan como fijos. Si tienen error, se traslada a los pesos.

¿QUÉ PUEDE HACER CADA OFICINA?

El método que se puede aplicar depende de la infraestructura disponible

	MECANISMO Y REGISTROS	ENCUESTA DE EVALUACIÓN	AJUSTE RECOMENDADO	CONDICIÓN INSTITUCIONAL
A	Sin registros integrados	Débil o ausente	Hot deck **localizado**: donantes del mismo micro-territorio y difíciles de censar	Cartografía y paradatos básicos
B	Omisión diferencial; registros fragmentados	Grande, con potencia para dominios pequeños	Ponderación por el inverso de la propensión, estimada con la encuesta	Capacidad de levantar una encuesta grande
C	Selección explicada por observables; registros integrados de alta cobertura	Disponible para validar	**Doblemente robusto**: propensión + calibración URUGUAY 2023	Registros integrados, acceso legal, capacidad metodológica propia

En los tres escenarios se puede calibrar a proyecciones demográficas por sexo, edad y región.

LO QUE LO HIZO POSIBLE

Más que una técnica, una capacidad institucional construida antes del censo

Acceso legal a registros

Sistema Estadístico Nacional y secreto estadístico; convenios con DGI y BPS.

`Ley 16.616`

Identificador único

Cédula recogida en el censo y disociada en un ID estadístico de acceso restringido.

`Ley 18.331 · URCDP`

Paradatos propios

Modo, contacto y vinculación web guardados en servidores del INE, e interoperabilidad vía AGESIC.

`CAWI + CAPI`

Equipo metodológico

Una división estable de unas diez personas, y proyecciones de población oficiales.

`INE`

Para el próximo censo

Mandato legal e identificador para acceder a registros · interoperabilidad · captura sistemática de paradatos · capacidad metodológica propia · encuesta de evaluación con potencia para los dominios que se van a corregir.

MÁS ALLÁ DE LOS MÉTODOS

La mejor corrección es la que no hace falta: un buen censo.

Ponderar reduce el sesgo, pero depende de supuestos. Ningún estimador sustituye a un operativo que llegue a todos los hogares.

Cartografía y precenso al día

Para que ninguna vivienda quede fuera del marco.

Campo con recursos y supervisión

Censistas suficientes, revisitas y recuperación de casos.

Monitoreo en tiempo real

Paradatos para detectar dónde no se está llegando, mientras todavía hay tiempo.

Gracias

Juan Pablo Ferreira · jferreir@ine.gub.uy
Instituto Nacional de Estadística (INE), Uruguay

INE (2024). Metodología de la Encuesta Nacional de Evaluación Censal 2023.

INE (2026). Censo 2023 ponderado: documento metodológico sobre el diseño de ponderadores.

STATISTICAL JOURNAL OF THE IAOS, 2026

Weighting a census as a non-probability sample: A doubly robust framework for differential undercoverage and the institutional capacity of national statistical offices (Uruguay's 2023 census)

Ferreira y Goyeneche · doi.org/10.1177/18747655261483350