INEI INSTITUTO NACIONAL DE ESTADISTICA E INFORMATICA

# Construcción de ponderadores y omisión censal

## Censos Nacionales de Población y Vivienda



Nancy Hidalgo Calle

Dirección Nacional de Censos y Encuestas

Lima, noviembre 2026

# Contenido


1 Introducción

2 Encuesta Postcensal

3 Construcción de ponderadores

4 Innovaciones del Censo 2025 y construcción de ponderadores

5 Conclusiones

# 1. INTRODUCCIÓN: RESULTADOS DEL CENSO NACIONAL 2017

- El Censo Nacional 2017 constituyó la principal operación estadística para conocer las características demográficas, sociales y económicas de la población peruana; y su resultado es la principal referencia para la planificación y formulación de políticas públicas.
- El recuento censal permitió contabilizar un total de 29,381,884 habitantes en el territorio nacional.
- La información permitió caracterizar a la población y conocer su distribución en el territorio nacional.

***Sin embargo, como ocurre en toda operación censal, existió población que no fue empadronada durante el operativo.***

# 2.
# ENCUESTA POSTCENSAL

# 2.1 ENCUESTA POSTCENSAL - OBJETIVO

Evaluar la cobertura y la calidad de la información recolectada en el Censo de Población y Vivienda del 2017.

# 2.2 ENCUESTA POSTCENSAL - FICHA TÉCNICA

## 1 COBERTURA

### COBERTURA GEOGRÁFICA

A nivel nacional, en el área urbana y rural de los 24 departamentos del país y la Provincia Constitucional del Callao.

### COBERTURA TEMPORAL

El trabajo de campo se desarrolló entre noviembre de 2017 y enero de 2018.

## 2 TIPO DE ENCUESTA

La encuesta fue de hecho con las personas entrevistadas en el Censo 2017.

## 3 MÉTODO DE ENTREVISTA

La entrevista fue directa (cara a cara).

## 4 DISEÑO MUESTRAL

### Población objetivo

Todas las viviendas particulares y sus ocupantes, del área urbana y rural del país.

Se excluyen las viviendas colectivas y las personas sin vivienda.

### Marco muestral

El marco muestral básico es de tipo áreas. La fuente de información proviene de la Actualización Cartográfica y Registro de Viviendas y Establecimientos del Pre-censo 2017.

### Unidades de muestreo

**13**
viviendas
**UPM urbana (Área Empadronamiento Urbana - AEU)**

**20 a 60**
viviendas
**UPM rural (Centro Poblado Rural)**

**5**
viviendas
**USM urbano**

**8**
viviendas contiguas
**USM rural**

# 2.2 ENCUESTA POSTCENSAL - FICHA TÉCNICA

## Inferencia de Resultados

- Nacional, Nacional Urbana, Nacional Rural
- Departamentos
- Área Metropolitana de Lima y Callao

## Tamaño de la muestra

**Tamaño global de la muestra**

**5,791** áreas muestrales

**Área urbana**

**4,758** áreas muestrales

**Área rural**

**1,033** áreas muestrales

**Viviendas particulares por entrevistar**

**32,054** viviendas

# 2.3 OMISIÓN CENSAL - RESULTADOS

## Población total, censada y omitida

La evaluación de la cobertura y calidad del censo mediante la Encuesta Post Censal (EPC) se basa en el Sistema Dual de Estimación (DSE: Dual System Estimation).

**Población total**

**31,237,385**

Personas registradas en el territorio peruano

**Población censada**

**29,381,884**

Personas efectivamente censadas

**Población omitida**

**1,856,809**

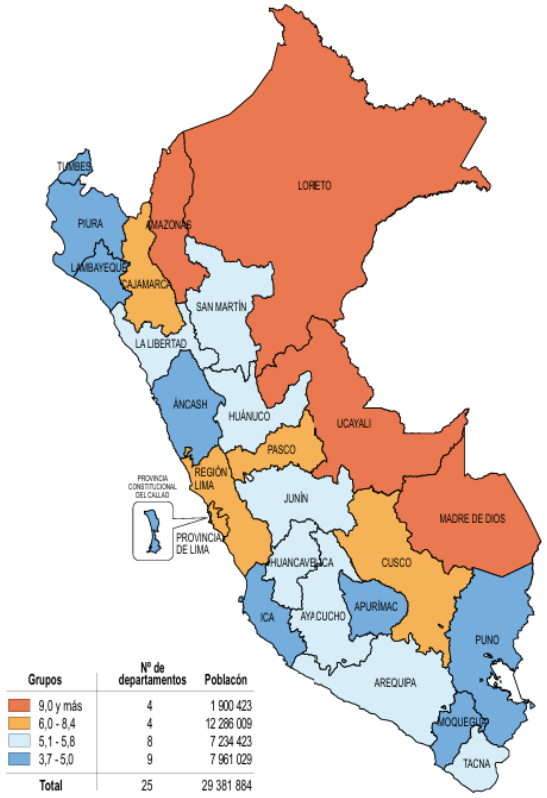
Población no registrada durante el censo

**5.94%** Tasa de omisión

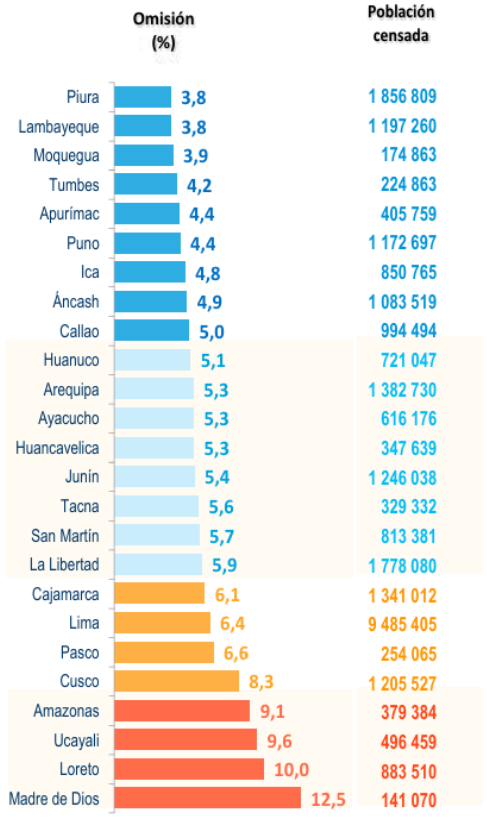
Fuente: Instituto Nacional de Estadística e Informática – Censos Nacionales 2017: XII de Población, VII de Vivienda y III de Comunidades Indígenas – Primeros Resultados.

# 2.3 OMISIÓN CENSAL - RESULTADOS

- La omisión censal presentó diferencias entre departamentos y según las características de las personas y los hogares, evidenciando niveles heterogéneos de cobertura censal.
- Esta heterogeneidad hizo necesario incorporar las particularidades territoriales y poblacionales en la estimación y construcción de los ponderadores.




| Grupos | N° de departamentos | Poblacón |
|---|---|---|
| 9,0 y más | 4 | 1 900 423 |
| 6,0 - 8,4 | 4 | 12 286 009 |
| 5,1 - 5,8 | 8 | 7 234 423 |
| 3,7 - 5,0 | 9 | 7 961 029 |
| Total | 25 | 29 381 884 |




***La omisión censal es un fenómeno diferenciado que requiere procedimientos estadísticos capaces de reconocer las particularidades territoriales y poblacionales.***

# 3.
# CONSTRUCCIÓN DE PONDERADORES

# 3.1 OBJETIVO

Obtener ponderadores que permitan ajustar las estimaciones con desagregación territorial a nivel provincial y distrital.

Ante la presencia de omisión censal y considerando su comportamiento heterogéneo tanto en el territorio como entre los distintos segmentos de la población, fue necesario diseñar e implementar una estrategia específica de ponderación.

**La finalidad** es compensar los efectos de la omisión diferencial y mejorar la representatividad de las estimaciones respecto de la población total; minimizando los posibles sesgos en las principales variables del Censo

# 3.2 ENFOQUE METODOLÓGICO

- Integrar la información censal con los resultados de la Encuesta Postcensal (EPC) para la construcción de los ponderadores.
- Aplicar modelos estadísticos para estimar y ajustar los ponderadores, mejorando su precisión y capacidad de representación de la población.
- Calibrar los ponderadores obtenidos, asegurando la consistencia de las estimaciones con los totales de referencia y los niveles de desagregación territorial establecidos.

# 3.3 FACTORES ASOCIADOS A LA OMISIÓN CENSAL

- **Viviendas**: viviendas ubicadas en edificios.
- **Hogares**: tamaño del hogar, hogares secundarios.
- **Personas**: sexo, diferentes grupos de edades, personas que no forman parte del hogar como los pensionistas o inquilinos.
- **Territorio**: departamento, área urbana o rural, dispersión geográfica y tamaño del área de empadronamiento.

***Estas variables permiten identificar grupos poblacionales con mayor probabilidad de haber sido omitidos.***

# 3.4 INTEGRACIÓN Y ORGANIZACIÓN DE LA BASE DE DATOS

**1 INTEGRACIÓN**

Se fusionaron las bases de la EPC, el Censo 2017 y fuentes externas.

Se emparejaron los registros individuales de la encuesta y el censo.

Se identificaron personas presentes en ambas fuentes y personas registradas únicamente en una de ellas.

**2 ARMONIZACIÓN**

Se armonizaron variables, dominios geográficos y estratos poblacionales.

**3 CONSTRUCCIÓN DE BASE DE DATOS**

Se construyó una base integrada de 471 variables: 46 de la EPC y 425 de otras fuentes.

**FUENTES EXTERNAS**

Registros administrativos:

- Registro Nacional de Municipalidades
- Ministerio de Desarrollo e Inclusión Social
- Ministerio de Educación
- Ministerio de Salud
- Sistema de Focalización de Hogares
- Censo Nacional Agropecuario

# 3.5 ESTRATIFICACIÓN DE PROVINCIAS

- Se identificaron estratos de provincias con características relativamente homogéneas y tasas de omisión significativamente diferentes entre ellos.
- Se utilizó el método de Bosques Aleatorios (Random Forest) para identificar las variables con mayor capacidad para explicar la omisión censal.
- Se partió de un conjunto de 471 variables disponibles, y se seleccionaron inicialmente 50 variables relevantes. Posteriormente, se definieron 21 variables para el procedimiento de estratificación.
- Se utilizó el análisis de conglomerados (Cluster Analysis) para agrupar las provincias según características demográficas, geográficas y territoriales.
- Se evaluaron las diferencias de las tasas de omisión entre estratos, y se ajustó la selección de variables hasta obtener diferencias estadísticamente significativas. El resultado fue una clasificación de 196 provincias en cinco estratos.




| Estratos | Omisión | Intervalo de confianza al 95% | | Coeficiente de variación |
|---|---|---|---|---|
| | | Inferior | Superior | |
| 1 | 5.7 | 5.2 | 6.2 | 4.4 |
| 2 | 7.3 | 6.3 | 8.3 | 7 |
| 3 | 9.9 | 8.6 | 11.1 | 6.4 |
| 4 | 5.2 | 4.7 | 5.7 | 5.3 |
| 5 | 4.3 | 3.5 | 5.1 | 9.6 |

# 3.6 ESTIMACIÓN DE LA  PROBABILIDAD DE OMISIÓN CENSAL

- Se estimó la probabilidad individual de omisión utilizando características personales, familiares, habitacionales y territoriales. La variable dependiente identifica si una persona fue omitida (1) o censada (0).
- Se aplicó un modelo de regresión logística binaria para cada uno de los cinco estratos.
- La muestra de la EPC se dividió en dos partes: 50% para entrenamiento y 50% para validación.
- Se estimaron 100 modelos LASSO con diferentes valores del parámetro lambda, y validación cruzada.
- Se utilizó el área bajo la curva ROC para seleccionar el modelo óptimo, y se evaluó el desempeño mediante exactitud, sensibilidad y especificidad.
- Los resultados de entrenamiento y validación fueron similares, lo que permitió descartar evidencias importantes de sobreajuste.

# 3.7 CÁLCULO DE LOS PONDERADORES AGREGADOS

- ❑ Se aplicaron los coeficientes estimados a La base de datos censal para calcular la probabilidad de omisión de cada persona.
- ❑ Se controlaron los valores extremos de las probabilidades estimadas mediante el método del rango intercuartílico.
- ❑ Se calcularon los ponderadores individuales como el inverso de la probabilidad de inclusión.
- ❑ Los ponderadores permitieron representar a las personas omitidas en el empadronamiento censal.
- ❑ La suma de los ponderadores permitió estimar la población omitida por estrato, provincia y distrito.

Fórmula central:

$$f_i = \frac{1}{1 - P_i}$$

Donde:

- $f_i$: ponderador individual.
- $P_i$: probabilidad estimada de omisión.

# 3.8 CALIBRACIÓN DE LOS PONDERADORES

- Se aplicó el método Raking Ratio, para ajustar los ponderadores individuales.
- El objetivo fue hacer coincidir las estimaciones ponderadas con los totales poblacionales de referencia obtenidos mediante la Encuesta Postcensal (EPC).
- Se utilizaron cuatro variables de control para la calibración: sexo, grupos de edad, área urbana/rural y departamento.
- El procedimiento de calibración se desarrolló en dos fases para controlar y corregir posibles valores extremos de los ponderadores a nivel provincial y distrital.
- Se incorporó posteriormente la población de viviendas colectivas.
- Finalmente, se obtuvieron ponderadores calibrados y consistentes para su aplicación en los niveles departamental, provincial y distrital.

# 4.

# INNOVACIONES DEL CENSO 2025 Y SUS IMPLICANCIAS EN LA CONSTRUCCIÓN DE PONDERADORES

# 4.1 INNOVACIONES DEL CENSO 2025

- ❑ Censo de derecho (de jure), según residencia habitual.
- ❑ Periodo de empadronamiento de tres meses (agosto-octubre), con recuperación censal posterior (noviembre-diciembre).
- ❑ Captura de información mediante dispositivos móviles (tablets).
- ❑ Georreferenciación de viviendas y monitoreo del personal de campo en tiempo real.
- ❑ Mayor incorporación de herramientas tecnológicas en el procesamiento.

# 4.2 IMPLICANCIAS EN LA CONSTRUCCIÓN DE PONDERADORES

- **Mayor disponibilidad de información para el control de cobertura**: la captura digital, la georreferenciación y el monitoreo permiten fortalecer los controles de cobertura y la identificación de posibles omisiones.

- **Incorporación de una etapa de recuperación censal**: permite recuperar información de viviendas y personas que no fueron registradas durante el período de operación de campo, lo que puede reducir la necesidad de ajustes posteriores por omisión.

**La construcción de ponderadores en el Censo 2025, al igual que en el Censo 2017, permite disponer de una estimación de la población total, de gran utilidad para los usuarios, especialmente para las autoridades en la planificación y toma de decisiones.**

**Asimismo, estos resultados contribuyen a mejorar las proyecciones de población, particularmente en los niveles provincial y distrital.**

**Tasa de omisión 2017**

**5,94 %**

**Tasa de omisión 2025**

**4,25 %**

# 5. CONCLUSIONES

1

**La EPC como fuente fundamental**: constituye la principal fuente de información para identificar y estimar la omisión censal, cuyos resultados sirven de base para la construcción y el ajuste de los ponderadores.

2

**Heterogeneidad de la omisión censal**: requiere la construcción de ponderadores diferenciados que permitan compensar las variaciones de la omisión entre los distintos ámbitos territoriales y segmentos de la población.

3

**Estratificación territorial con fines estadísticos**: permite desarrollar modelos diferenciados y obtener ponderadores ajustados a las características particulares de las provincias.

4

**Modelamiento estadístico para la construcción de ponderadores:** Random Forest y LASSO permiten identificar y seleccionar variables relevantes, mientras que la regresión logística estima las probabilidades de inclusión y omisión que sirven de base para el cálculo de los ponderadores individuales.

5

**La calibración como etapa complementaria:** permite ajustar los ponderadores para mejorar la consistencia de las estimaciones ponderadas respecto de los totales poblacionales de referencia, utilizando variables de control y distintos niveles de desagregación territorial.

6

**Utilidad de los ponderadores:** permiten obtener estimaciones de la población total que sirven de referencia para la toma de decisiones, el diseño de políticas públicas, y contribuyen a mejorar las proyecciones de población en los niveles provincial y distrital.

INEI
INSTITUTO NACIONAL DE ESTADISTICA E INFORMATICA

# GRACIAS!!!