

Construcción del Mapa de Pobreza Monetaria Provincial y Distrital en el Perú

Utilizando la Metodología de Estimación en Áreas Pequeñas

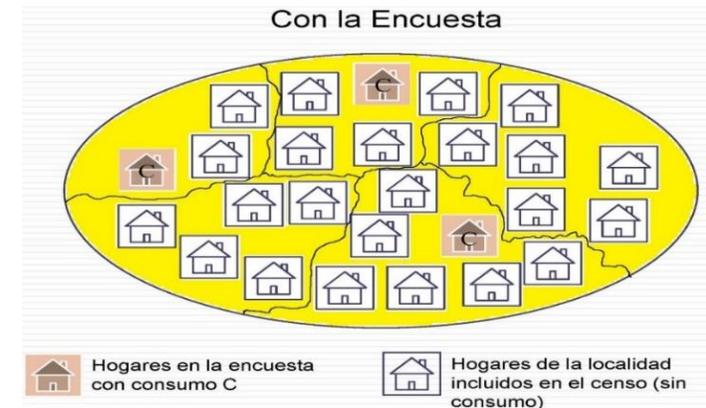
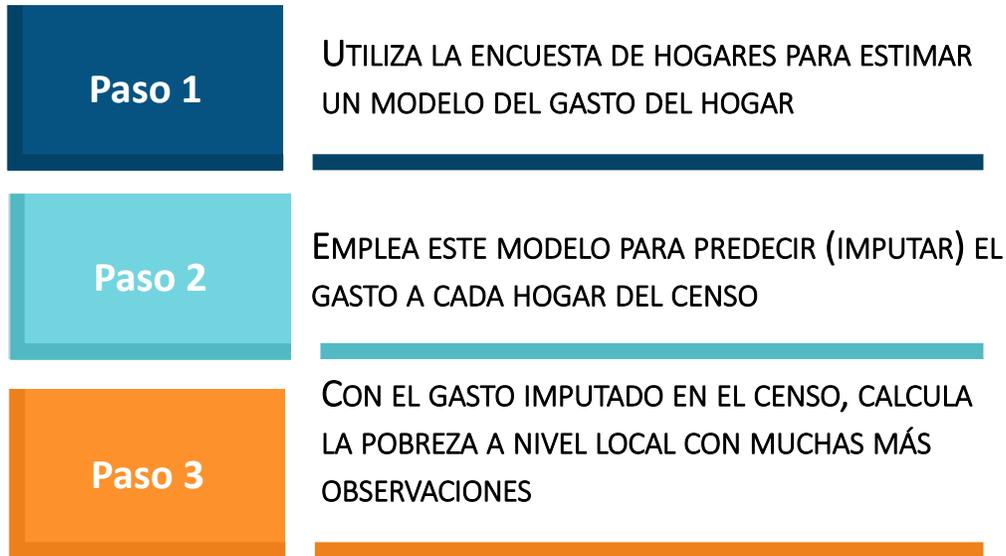
José Antonio Llanos Solórzano
Dirección Nacional de Censos y Encuestas
INEI - PERÚ
Abril de 2021

1. Metodología de Estimación en Áreas Pequeñas
2. Aplicación de la Metodología de Estimación en Áreas Pequeñas al Mapa de Pobreza Monetaria
3. Etapas de la Elaboración del Mapa de Pobreza Monetaria

Metodología de Estimación en Áreas Pequeñas

Metodología ELL

- Desarrollada por Elbers, Lanjouw y Lanjouw ⁽¹⁾, y utilizada por el Banco Mundial.
- Método indirecto basado en modelo, a nivel de unidad.
- Emplea la encuesta de hogares y el censo de población.



(1) Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2003). Micro-level estimation of poverty and inequality. *Econometrica*

Modelo

$$\ln y_{ch} = E[\ln y_{ch} | x_{ch}^T] + u_{ch} = x_{ch}^T \beta + u_{ch}$$

$\ln Y_{ch}$: logaritmo del gasto per cápita del hogar h en el cluster muestral c

X_{ch} : características del hogar h en el cluster c

β : vector que captura la correlación entre $\ln Y_{ch}$ y las X_{ch}

μ_{ch} : error de la regresión (lo que no se puede explicar con las X_{ch})

- No es suficiente con estimar los β (a diferencia de gran parte de la econometría)
- Si sólo se utiliza las estimaciones de los β , sólo se obtendrá la parte del consumo que se puede explicar con las X_{ch}
- Falta agregar la parte que no se puede explicar; es decir, una estimación de los errores

$$\mu_{ch} = \eta_c + \varepsilon_{ch}$$

Errores del Modelo

$$u_{ch} = \eta_c + \varepsilon_{ch}$$

η_c : error a nivel del área (o cluster) c (parte no explicada común para los hogares dentro del área c)

ε_{ch} : error a nivel del hogar (parte no explicada específica del hogar)

Supuestos

- η y ε son independientes entre sí, y no correlacionados con X_{ch}
- η_c es i.i.d. y homocedástico
- ε_{ch} se asume heterocedástico (depende de X_{ch})

Varianza de los Errores

La varianza del error idiosincrático (del hogar) se estima imponiendo la siguiente forma paramétrica de heterocedasticidad:

$$\text{Var}(\varepsilon_{ch}) = \frac{Ae^{H'_{ch}\alpha} + B}{1 + e^{H'_{ch}\alpha}}$$

Donde:

- A, B y α son parámetros a estimar.
- H_{ch} es una función de las variables que se incluyeron en el modelo de consumo

Estimación de A, B y α :

$$A^* = 1.05 * \max(\hat{\varepsilon}_{ch}^2)$$

$$\hat{B} = e^{H'_{ch}\hat{\alpha}}$$

La estimación de α , se obtiene de la siguiente regresión (Modelo Alfa):

$$\ln \left[\frac{\hat{\varepsilon}_{ch}^2}{(A^* - \hat{\varepsilon}_{ch}^2)} \right] = H'_{ch}\alpha + r_{ch}$$

Aplicación de la Metodología de Estimación en Áreas Pequeñas al Mapa de Pobreza Monetaria

- La pobreza monetaria se mide anualmente a partir de la información de la Encuesta Nacional de Hogares (ENAHOG), que utiliza el gasto como indicador de bienestar.
- Mediante Resolución Suprema Nº 097-2010-PCM se constituye la Comisión Consultiva para la Estimación de la Pobreza y otros indicadores relacionados, conformada por organismos nacionales e internacionales; cuyo objetivo es garantizar la calidad, transparencia y confianza de la información en el campo de la medición de la pobreza.

Descripción detallada de la distribución espacial de la pobreza monetaria en el país. Implica un ordenamiento de las divisiones geográficas según alguna medida que indique el nivel de bienestar de la población.

Objetivo

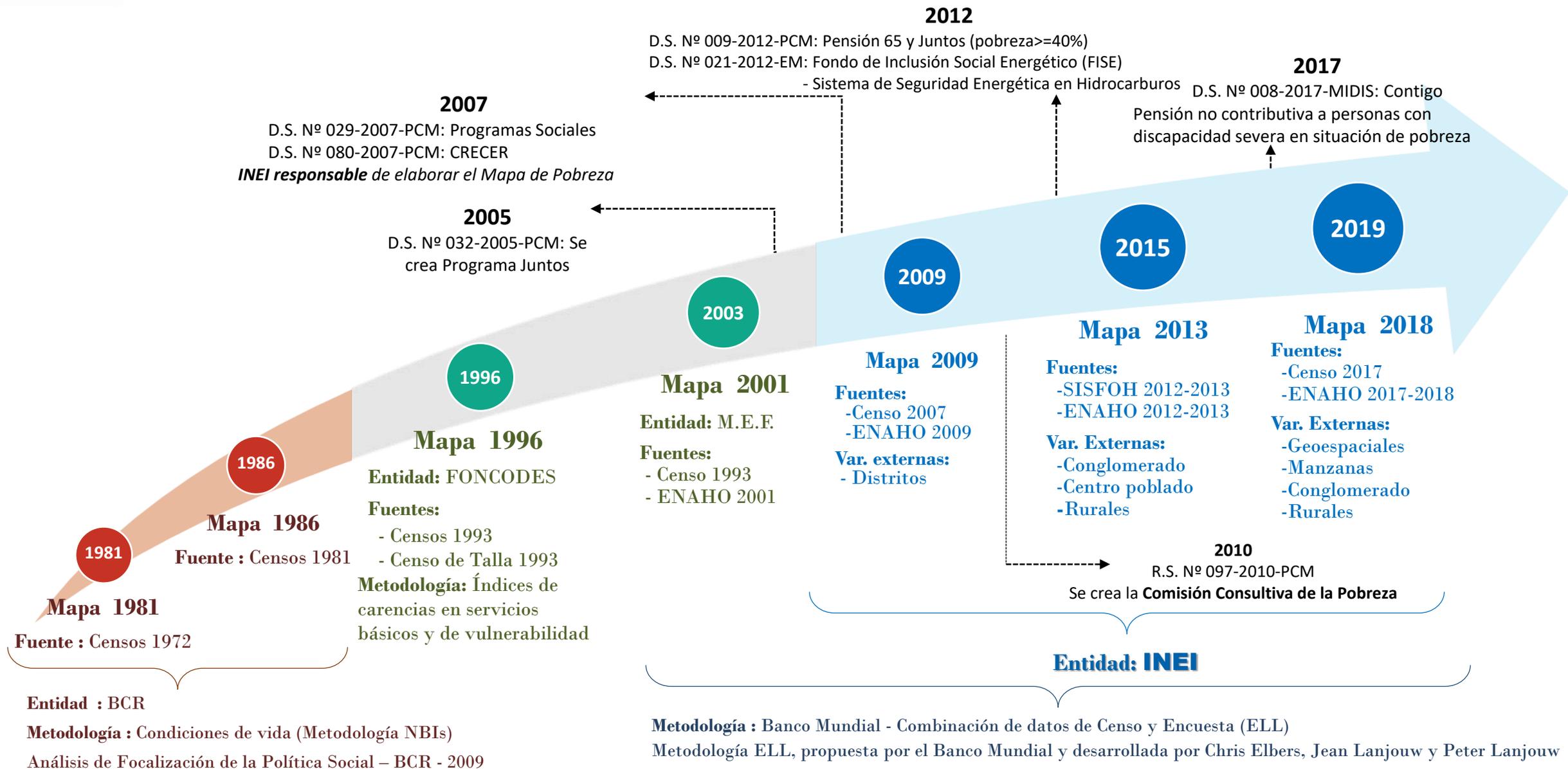
- El objetivo del Mapa de Pobreza es mostrar la distribución geográfica de la pobreza monetaria a nivel de áreas menores (provincia, distrito e infradistrito) para identificar y priorizar las zonas más pobres del país, debido a que la información a nivel de departamento no refleja la heterogeneidad dentro de las provincias ni distritos.

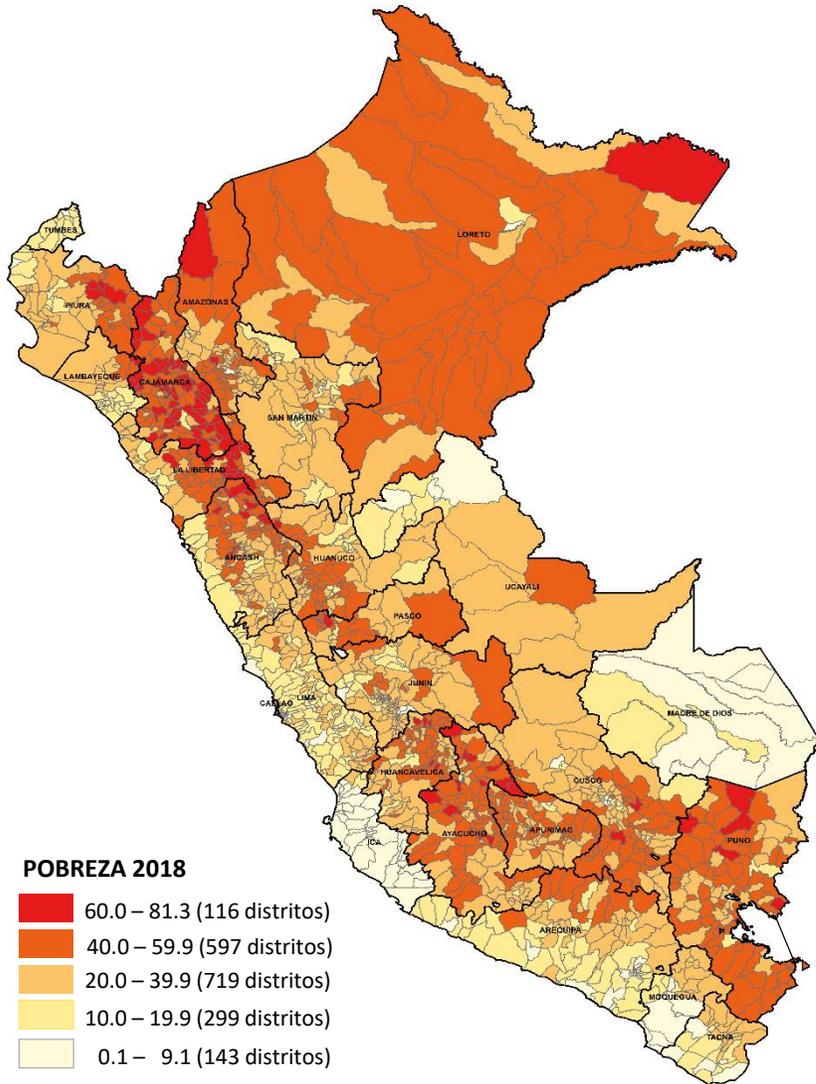
Utilidad

- Es un instrumento para la orientación e implementación de políticas sociales y priorización de los ámbitos de intervención, ya que permite contar con información desagregada geográficamente.

- Facilita la definición de prioridades para la asignación de recursos públicos y es de mucha utilidad en la focalización del gasto público en zonas geográficas que presentan una mayor incidencia de la pobreza.
- Es un insumo y guía para los lineamientos en los diferentes Programas Sociales, porque permite priorizar a la población menos favorecida.
- Identifica las diferencias en las condiciones de los hogares y la población a nivel de áreas pequeñas (provincias, distritos e infra distritos) donde una encuesta no tiene ese nivel de inferencia.

ANTECEDENTES DEL MAPA DE POBREZA



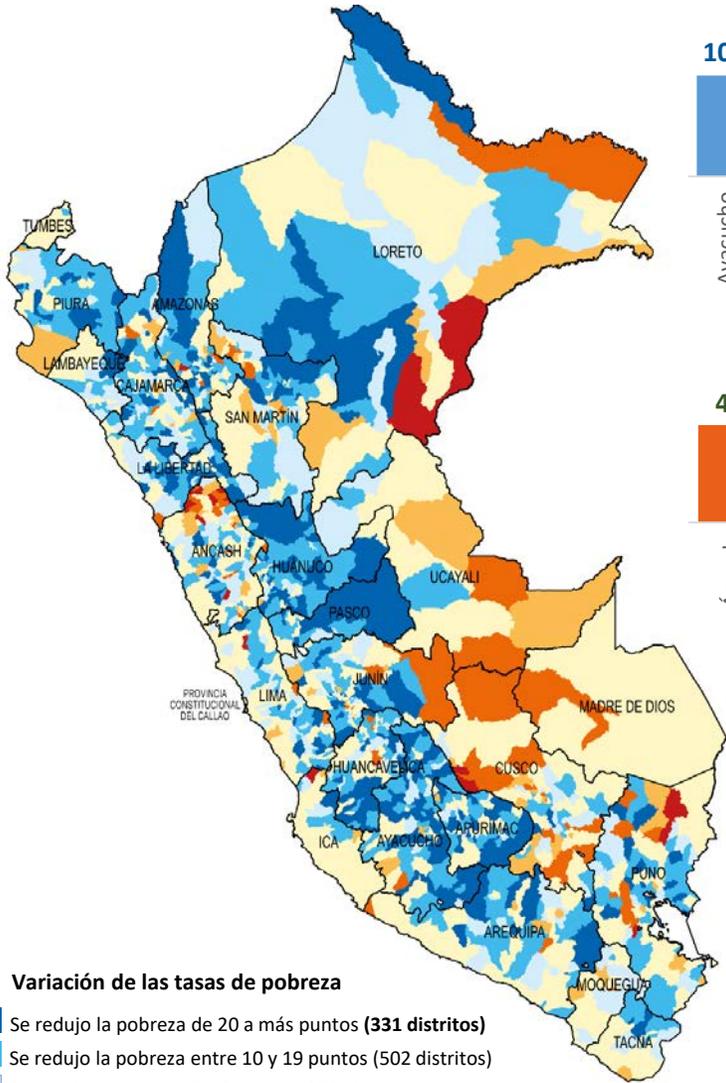


- De los 1 874 distritos, 713 distritos (38,1%) tienen una pobreza monetaria de 40% a más y afecta al 13,4% de la población total del país (4 millones 166 mil 641).
- En 719 (38,4%) distritos la pobreza monetaria se encuentra entre los rangos de 20% a menos de 40%, y afecta al 24,4% (7 millones 607 mil 537) de la población.
- En 442 (23,6%) distritos la pobreza monetaria es menor a 20% y agrupa al 62,3% (19 millones 463 mil 207) de la población del país.

Rango de Pobreza Monetaria	Número de distritos		Población en los distritos	
	Abs.	%	Abs.	%
Menos a 20%	442	23,6	19 463 207	62,3
De 20% a 39,9%	719	38,4	7 607 537	24,4
De 40% a 59,9%	597	31,9	3 424 520	11,0
De 60% y más	116	6,2	742 121	2,4
TOTAL	1 874	100,0	31 237 385	100,0

Distritos	Rango de habitantes
29	Más de 20 mil
237	5 mil a menos 20 mil
244	2 mil a menos 5 mil
203	Menos de 2 mil

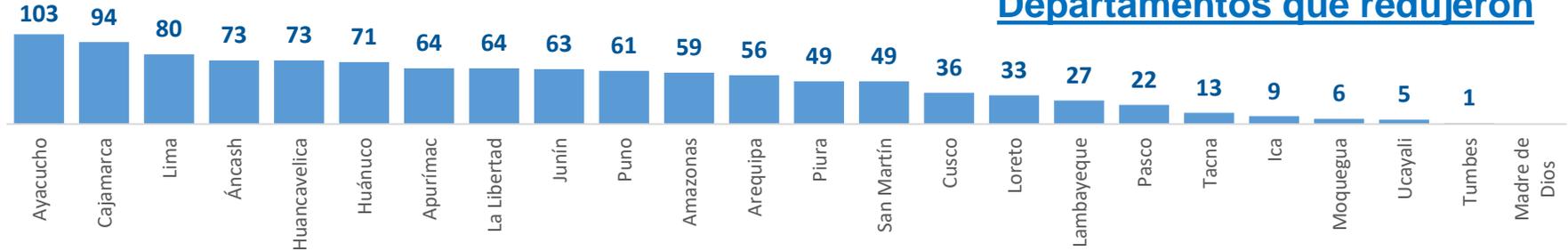
VARIACIONES EN LA POBREZA MONETARIA DISTRICTAL 2013 Y 2018



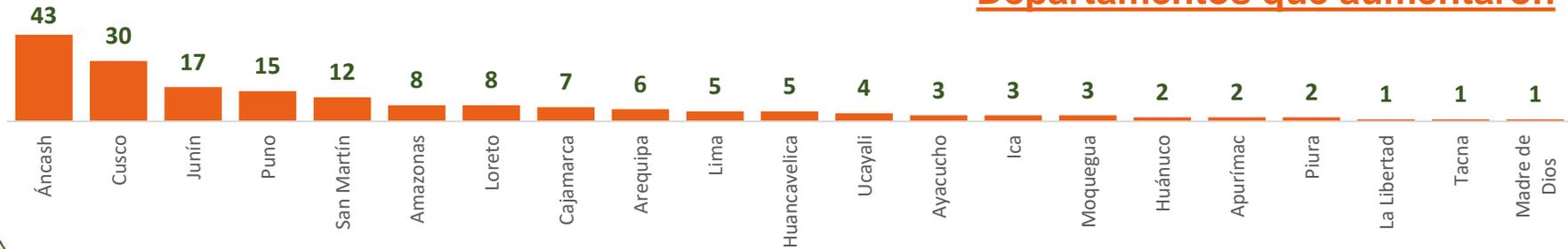
Variación de las tasas de pobreza

- Se redujo la pobreza de 20 a más puntos (331 distritos)
- Se redujo la pobreza entre 10 y 19 puntos (502 distritos)
- Se redujo la pobreza 5 y 9 puntos (278 distritos)
- No hubo cambios (585 distritos)
- Aumentó la pobreza entre 5 y 9 puntos (87 distritos)
- Aumentó la pobreza entre 10 y 19 puntos (69 distritos)
- Aumentó la pobreza de 20 a más puntos (22 distritos)

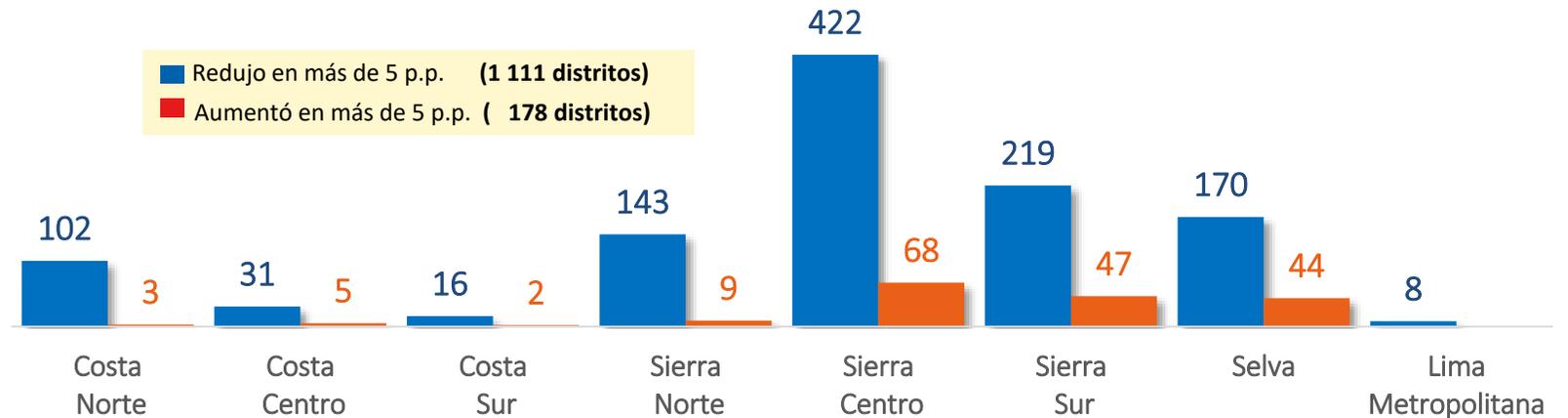
Departamentos que redujeron



Departamentos que aumentaron



Dominio Geográfico



Etapas de la Elaboración del Mapa de Pobreza Monetaria

- ✓ Identificar y analizar las fuentes de datos relevantes para determinar cuáles tienen potencial para usarse en la construcción del mapa de pobreza monetaria.

Encuesta Nacional de Hogares 2017-2018

Censos Nacionales de Población y Vivienda 2017

Cartografía digital a nivel distrital

Bases de Población

- Oficina de Normalización Previsional (ONP), 2016
- Registro de Usuarios Afiliados al Aseguramiento Universal de Salud (SUSALUD), 2019
- Sistema Integral de Salud (SIS), 2017
- Sistema de Información de Apoyo a la Gestión de la Institución Educativo (SIAGIE), 2016-2017
- Evaluación Censal de Estudiantes (ECE), 2011-2016
- Plan de Salud Escolar (PSE), 2016
- Registro Nacional de Usuarios de Programas Sociales (RNU), 2017
- Programa Nacional CUNAMAS, 2018
- IV Censo Nacional Agropecuario (CENAGRO), 2012
- Paneles Solares u algún otro mecanismo que genere energía eléctrica (MINEM), 2017

Bases Institucionales

- Registro Nacional de Municipalidades (RENAMU), 2018
- Registro Nacional de Identificación y Estado Civil (RENIEC), 2017
- Directorio Central de Empresas y Establecimientos, 2017
- Programa Nacional de Becas y Crédito Educativo (PRONABEC), 2012-2018
- Programa Nacional de Saneamiento Rural, 2018
- Censo Escolar, 2017
- Sistema de Denuncias Policiales (SIDPOL), 2017
- Organismo Supervisor de la Inversión en Energía y Minería (OSINERMIN), 2017
- Censo Penitenciario, 2016
- Mapa de Pobreza Provincial y Distrital, 2013

- ✓ Asegurar la coincidencia de las diversas fuentes de datos a través de códigos geográficos y/o variables de diseño de encuestas (como estratificación y agrupamiento).

Se realiza la equivalencia de los marcos de manzanas y centros poblados entre la ENAHO 2017-2018 y el CPV 2017.

Inconsistencia entre Manzanas

- Cambio de posición.
- Cambio de orientación y forma.
- Distinta ubicación espacial, distinto sistema de coordenadas.
- Desfase en escala y orientación.

Inconsistencia Tabular

- Las manzanas mantienen su forma y/o ubicación ; sin embargo, su codificación cambia.

- ✓ Identificar y listar preguntas del censo y la encuesta que a priori son lo suficientemente similares.
- ✓ Estandarizar aquellas preguntas comunes al censo y la encuesta, a nivel de las categorías de respuesta.

ENAH0 2017

103. EL MATERIAL PREDOMINANTE EN LOS PISOS ES:	
¿Parquet o madera pulida?	1
¿Láminas asfálticas, vinílicos o similares?	2
¿Losetas, terrazos o similares?	3
¿Madera (entablados)?	4
¿Cemento?	5
¿Tierra?	6
¿Otro material? _____	7
(Especifique)	

ENAH0 2018

103. EL MATERIAL PREDOMINANTE EN LOS PISOS ES:	
¿Parquet o madera pulida?	1
¿Láminas asfálticas, vinílicos o similares?	2
¿Losetas, terrazos o similares?	3
¿Madera (pona, tornillo, etc.)?	4
¿Cemento?	5
¿Tierra?	6
¿Otro material? _____	7
(Especifique)	

CPV 2017

5. EN LA VIVIENDA, ¿EL MATERIAL DE CONSTRUCCIÓN PREDOMINANTE EN LOS PISOS ES DE:	
(Lea cada alternativa y rellene solo un óvalo ●)	
1 <input type="radio"/>	Parquet o madera pulida?
2 <input type="radio"/>	Láminas asfálticas, vinílicos o similares?
3 <input type="radio"/>	Losetas, terrazos, cerámicos o similares?
4 <input type="radio"/>	Madera (pona, tornillo, etc.)?
5 <input type="radio"/>	Cemento?
6 <input type="radio"/>	Tierra?
7 <input type="radio"/>	Otro material?

- ✓ Imputar los datos faltantes de las variables de interés.

Se utilizan los siguientes criterios:

- Si la variable es numérica, los datos faltantes son remplazados por su valor medio o mediana y si es categórica es decir la variable tiene 2 o más atributos se remplazan por la moda.
- Para que las variables reflejen la diferencia que existe entre las subdivisiones de los departamentos, provincias y demás, es necesario que los datos faltantes también reflejen esa diferencia; por lo que la imputación debe de realizarse con observaciones que se encuentren a la mínima subdivisión a la cual pertenece la observación con datos faltantes (por ejemplo, la manzana).
- Para que el valor medio, mediana o moda de una variable que hay dentro de una subdivisión pueda utilizarse para la imputación no debe de tener más del 20% de sus observaciones vacías, caso contrario se deberá subir a la siguiente subdivisión.

- ✓ Verificar que las variables no tengan valores constantes; es decir, que los valores de las variables no se concentren en su totalidad en un solo valor.

Se eliminan las variables que están concentradas en un solo valor. El umbral establecido es de 95%

- ✓ Verificar la similitud de las distribuciones (características estadísticas) de las variables comunes entre el censo y la encuesta.

Se realizan las siguientes pruebas:

- La media censal debe estar dentro del intervalo de confianza calculada en la encuesta
- Prueba t de dos muestras
- Prueba de Kolmogorov-Smirnov de dos muestras
- Densidad de kernel
- Prueba chi-cuadrada de homogeneidad

- ✓ Seleccionar variables que tengan suficiente apoyo empírico para ser incluidas como variables predictivas candidatas en los modelos.

Modelo de Regresión Lasso

- El modelo LASSO (Least Absolute Shrinkage and Selection Operator) es una técnica similar a OLS que contrae o regulariza el valor de los coeficientes (β) “forzándolos” hacia el valor cero.
- Y por lo tanto, reduce la variabilidad de las estimaciones y produce modelos más simples.

En el modelo de regresión Lasso se minimiza:

$$\hat{\beta} = \arg \min_{\beta} \left\{ \frac{1}{n} \sum_{i=1}^n (y_i - \mathbf{x}_i \beta')^2 + \lambda \sum_{j=1}^p \omega_j |\beta_j| \right\}$$

Donde

- $\lambda > 0$ es el parámetro de penalización de lasso
- \mathbf{x} contiene las p variables predictoras
- Los ω_j son las ponderaciones a nivel de los parámetros

LASSO adaptativo

Tiene las mismas ventajas que LASSO, pero además, evita el sobreajuste al utilizar pesos para la penalización de los coeficientes.

Procedimiento

1. Estimar el modelo de consumo (modelo beta) por OLS
2. Estimar el modelo de residuos (modelo alfa) para estimar las varianzas de los errores (heterocedasticidad)
3. Reestimar modelo de consumo por GLS para obtener los coeficientes (β) y los otros parámetros (σ^2_e , σ^2_η)
4. Simular niveles de consumo en el censo

Software PovMap

- Software PovMap creado por el Banco Mundial, y disponible gratuitamente
- Tiene una interfaz gráfica de usuario que simplifica su uso. Sin embargo, los profesionales más avanzados que deseen una mayor funcionalidad y opciones tendrán que programarlos ellos mismos.

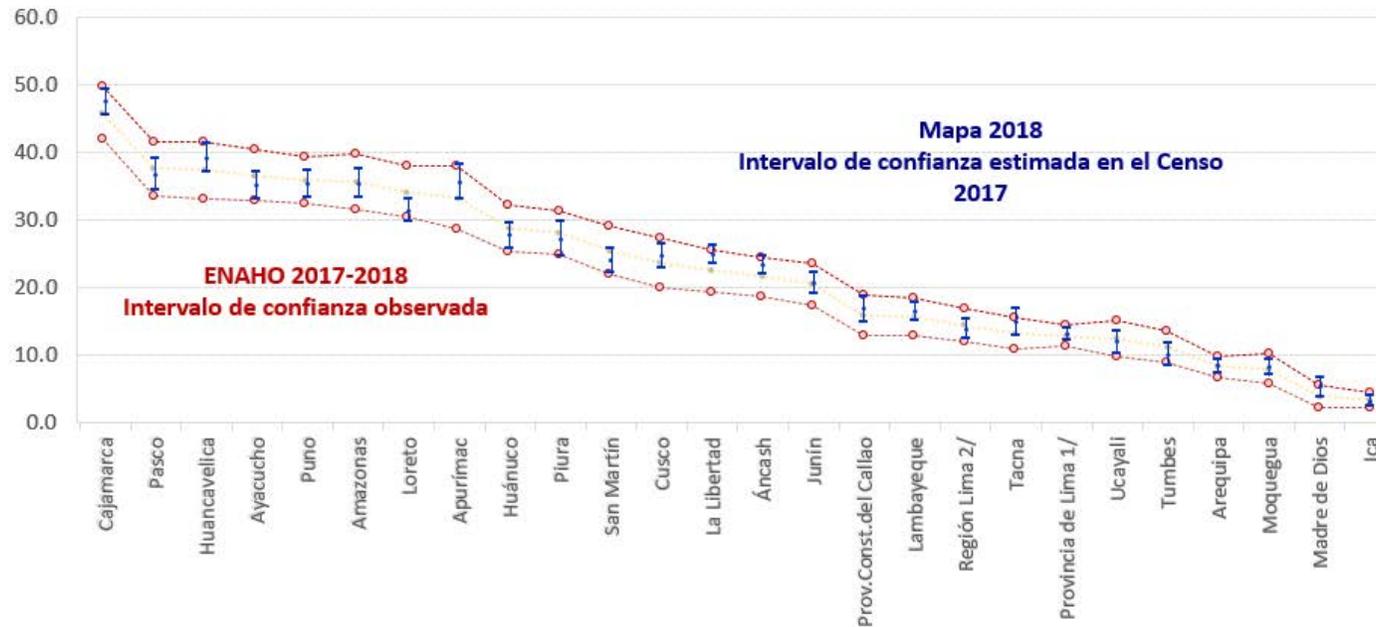
Implementación en Stata (Paquete SAE)

- Se ha creado una familia de comandos Stata que implementan los métodos de estimación de áreas menores (metodología ELL), y proporcionan a los usuarios una alternativa válida, modular y flexible a PovMap.
- Este paquete usa Mata para evitar las limitaciones de memoria que inevitablemente aparecen cuando se trabajan con grandes cantidades de datos.

SELECCIÓN Y VALIDACIÓN DE MODELOS

- ✓ Comparar gasto estimado con el observado en la encuesta en los diferentes niveles de inferencia.

PERÚ: INCIDENCIA DE LA POBREZA TOTAL ESTIMADA EN EL CENSO 2017 Y LA INCIDENCIA OBSERVADA DE LA ENAHO 2017-2018 POR REGIONES



- ✓ Comparar incidencia de pobreza estimada con la observada en la encuesta en sus diferentes niveles de inferencia.
- ✓ Validar la sensibilidad de las estimaciones en las submuestras.

Se agruparon **distritos**, con menor tamaño poblacional y alto coeficiente de variación, para **obtener mayor precisión** en la estimación de la pobreza monetaria.

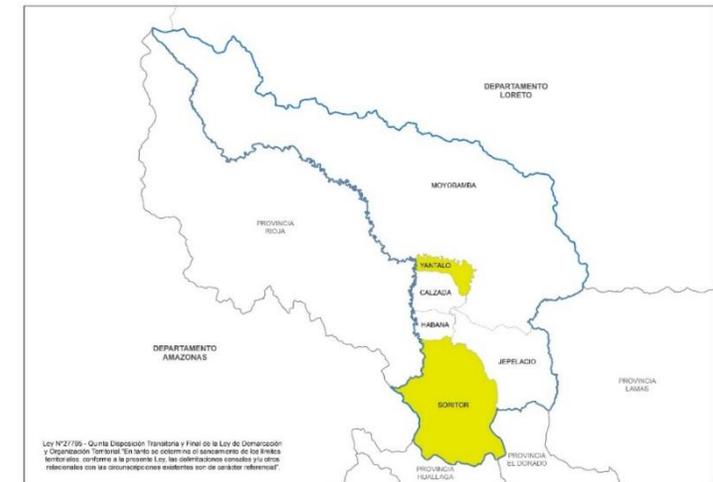
Criterios de selección de distritos a agrupar:

- Menos de mil hogares
- Coeficiente de variación estimado mayor de 25%
- Carácter urbano/rural del distrito

La agrupación se realizó dentro de cada provincia con distritos del mismo grupo robusto

El grupo robusto, de estimación de pobreza, se formó previamente mediante pruebas de hipótesis de diferencia de medias.

Características	Grupo	
	61	61
Departamento	San Martín	San Martín
Provincia	Moyobamba	Moyobamba
Distrito	Soritor	Yantalo
Hogares	5 992	800
Área urbana	72.8	78.0
Área rural	27.2	22.0
Incidencia inicial		
Incidencia	33.0	25.8
C.V.	14.4	26.3
Incidencia agrupada		
Incidencia	31.1	31.1
C.V.	14.9	14.9

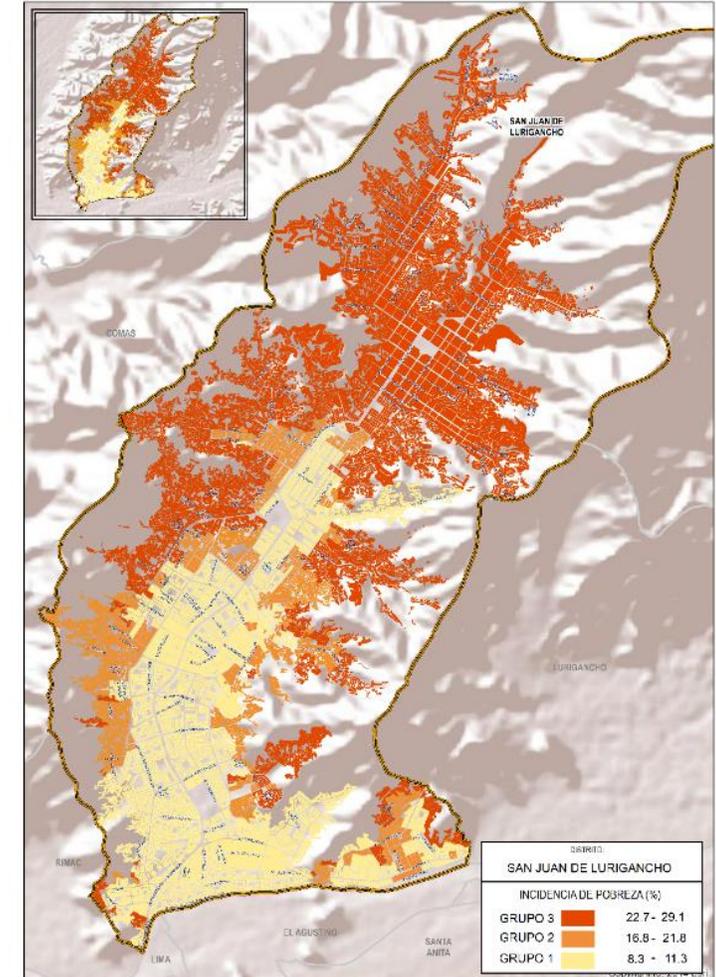


Se desagregaron **distritos**, con **gran población y alta heterogeneidad** en las condiciones de vida, en zonas según niveles de pobreza.

Criterios de selección de distritos a desagregar:

- 20 mil a más hogares
- Pobreza estimada mayor de 3%
- Coeficiente de variación estimado menor de 20%

La desagregación del distrito se realizó a nivel de núcleos urbanos.



Distrito de San Juan de Lurigancho en Lima.





www.inei.gob.pe
infoinei@inei.gob.pe



José Antonio Llanos Solórzano
Dirección Nacional de Censos y Encuestas
INEI - PERÚ
jose.llanos@inei.gob.pe