



NACIONES UNIDAS

CEPAL

Medición desagregada para indicadores de pobreza

División de Estadísticas

2019

Andrés Gutiérrez, PhD

Asesor Regional en Estadísticas Sociales

andres.gutierrez@un.org

Las desagregaciones en la medición de los Objetivos de Desarrollo Sostenible



OBJETIVOS DE DESARROLLO SOSTENIBLE

1 FIN DE LA POBREZA

2 HAMBRE CERO

3 SALUD Y BIENESTAR

4 EDUCACIÓN DE CALIDAD

5 IGUALDAD DE GÉNERO

6 AGUA LIMPIA Y SANEAMIENTO

7 ENERGÍA ASEQUIBLE Y NO CONTAMINANTE

8 TRABAJO DECENTE Y CRECIMIENTO ECONÓMICO

9 INDUSTRIA, INNOVACIÓN E INFRAESTRUCTURA

10 REDUCCIÓN DE LAS DESIGUALDADES

11 CIUDADES Y COMUNIDADES SOSTENIBLES

12 PRODUCCIÓN Y CONSUMO RESPONSABLES

13 ACCIÓN POR EL CLIMA

14 VIDA SUBMARINA

15 VIDA DE ECOSISTEMAS TERRESTRES

16 PAZ, JUSTICIA E INSTITUCIONES SÓLIDAS

17 ALIANZAS PARA LOGRAR LOS OBJETIVOS


OBJETIVOS DE DESARROLLO SOSTENIBLE

Algunas metas del ODS1 (fin de la pobreza)

- De aquí a 2030, erradicar para todas las personas y en todo el mundo la pobreza extrema (actualmente se considera que sufren pobreza extrema las personas que viven con menos de 1,25 dólares de los Estados Unidos al día).
- De aquí a 2030, reducir al menos a la mitad la proporción de hombres, mujeres y niños de todas las edades que viven en la pobreza en todas sus dimensiones con arreglo a las definiciones nacionales.

Principio fundamental de la desagregación de datos

Los indicadores de los Objetivos de Desarrollo Sostenible deberán desglosarse, siempre que sea pertinente, por ingreso, sexo, edad, raza, etnicidad, estado migratorio, discapacidad y ubicación geográfica, u otras características, de conformidad con los Principios Fundamentales de las Estadísticas Oficiales.

*****Resolución de la Asamblea General - 68/261*****

Limitaciones de las encuestas y el poder de la información auxiliar

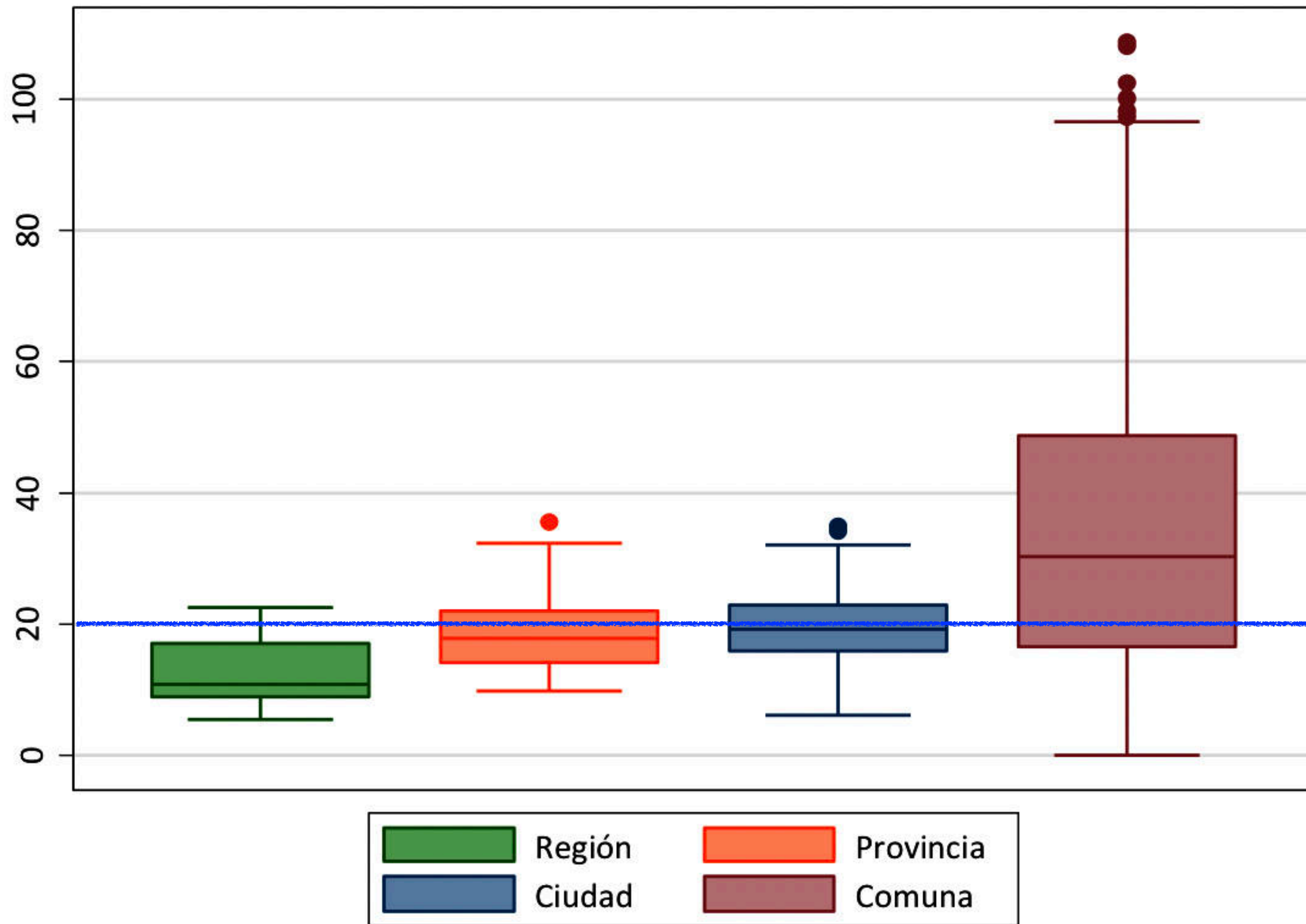
¿De qué se trata?

Cuando el tamaño de muestra de las encuestas no alcanza a soportar la inferencia estadística requerida para algunos subgrupos de interés, es necesario recurrir a información auxiliar externa (censos o registros administrativos) para que en **conjunto** (encuestas y datos externos) se construya un sistema inferencial preciso y exacto, que brinde información en las desagregaciones de interés.

¿Qué es un área pequeña?

Un área (o dominio) se denomina pequeña si el tamaño de muestra no es suficiente para soportar el proceso de inferencia directa (basado en el principio del diseño de muestreo de las encuestas) con una precisión adecuada

El término **no** se refiere al tamaño absoluto del subgrupo: departamentos, provincias o estados pueden ser áreas pequeñas si el tamaño de muestra es insuficiente.



Coeficientes de variación en diferentes desagregaciones

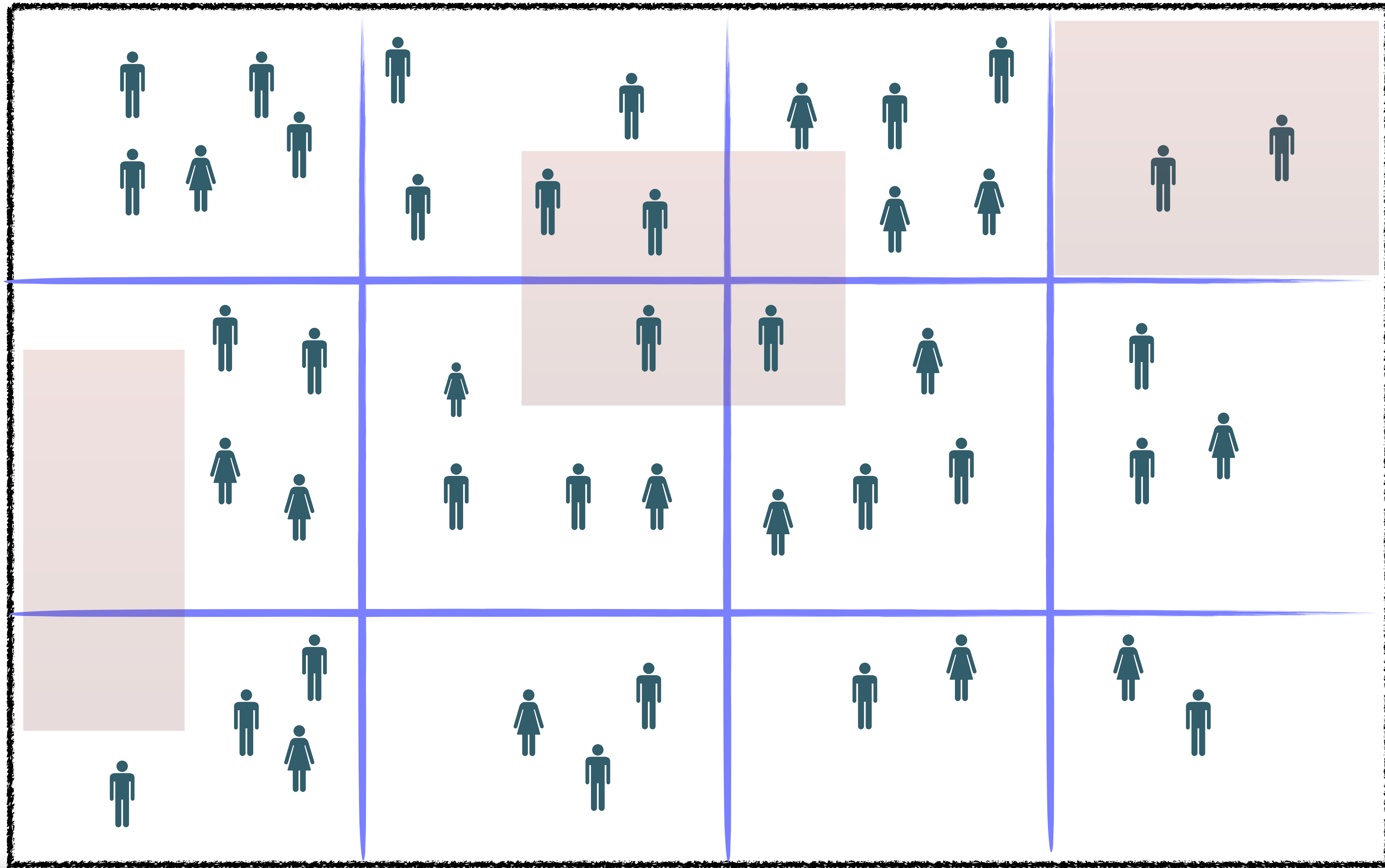
Fuente: INE - Chile

¿Qué es un área pequeña?

Los parámetros de interés puede ser requeridos en desagregaciones geográficas (que pueden ser visualizadas en un mapa) o en cruces de subgrupos sociodemográficos y de ingreso.

- Geográfico: provincias, departamentos, municipios y sectores censales.
- Subgrupos específicos: cruces entre edad × sexo × etnia × estatus migratorio.

En general, si los subgrupos no hacen parte de los dominios de diseño de las encuestas, su tamaño de muestra no se planifica de antemano y por ende será aleatorio, lo que aumenta la incertidumbre de la estimación.



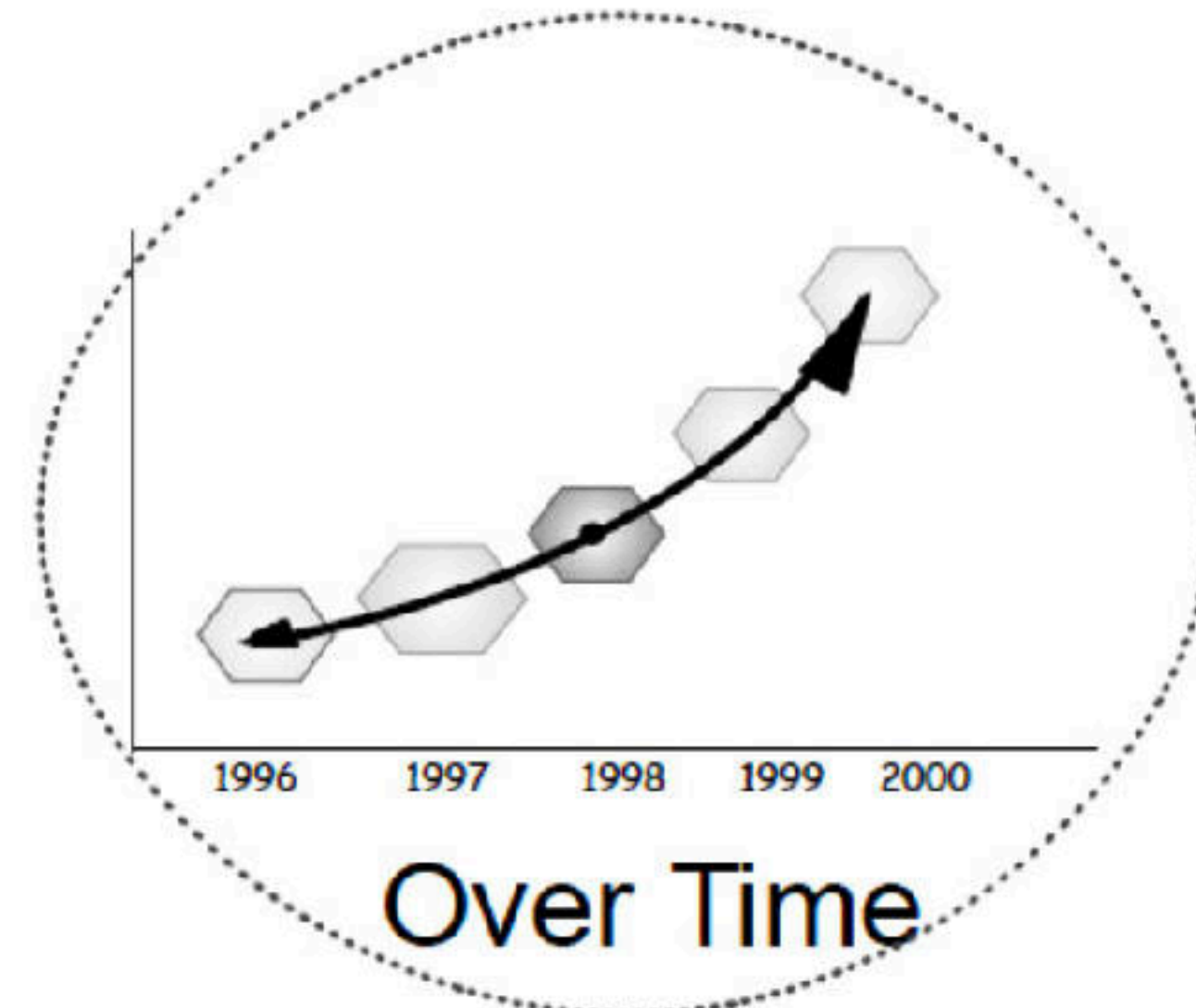
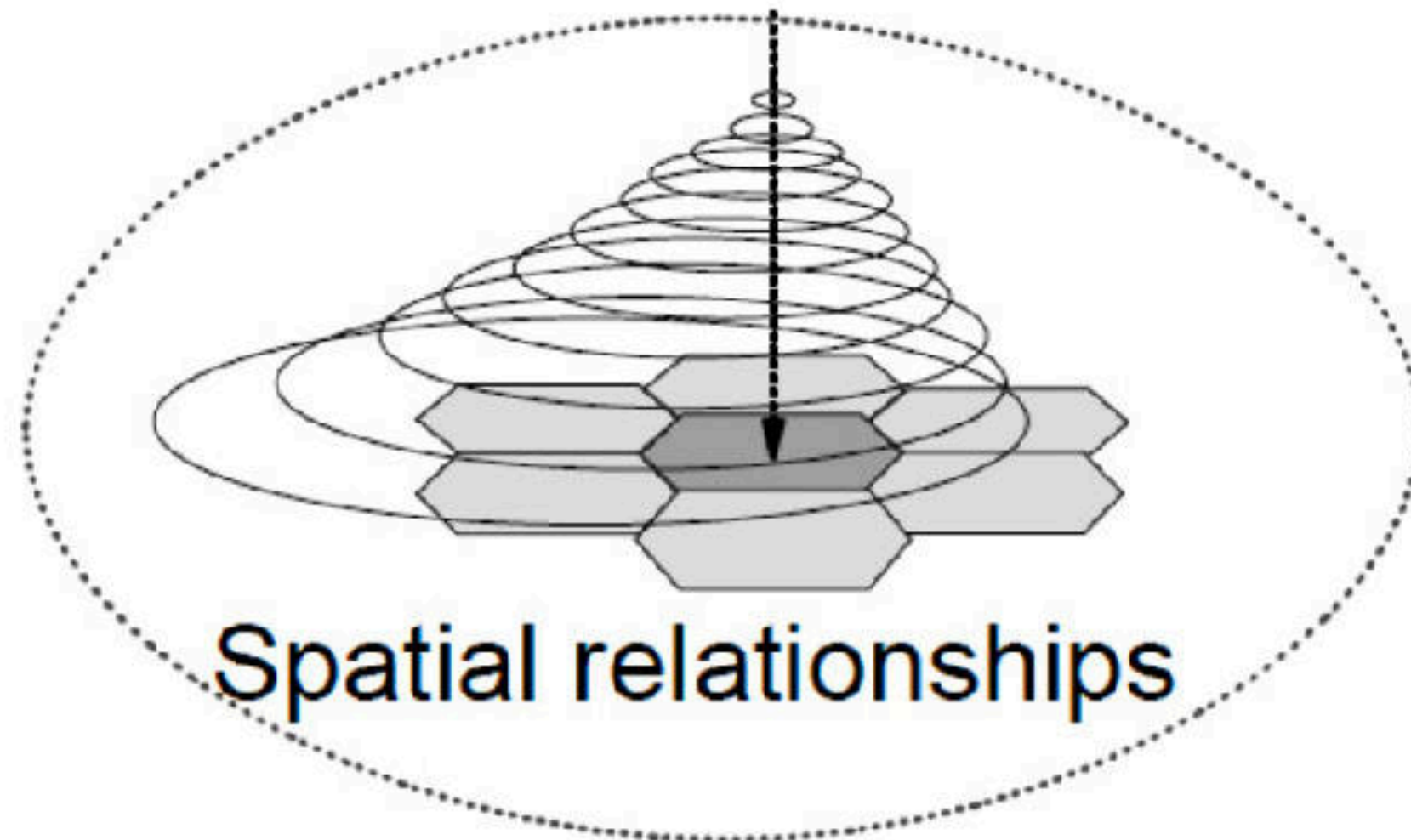
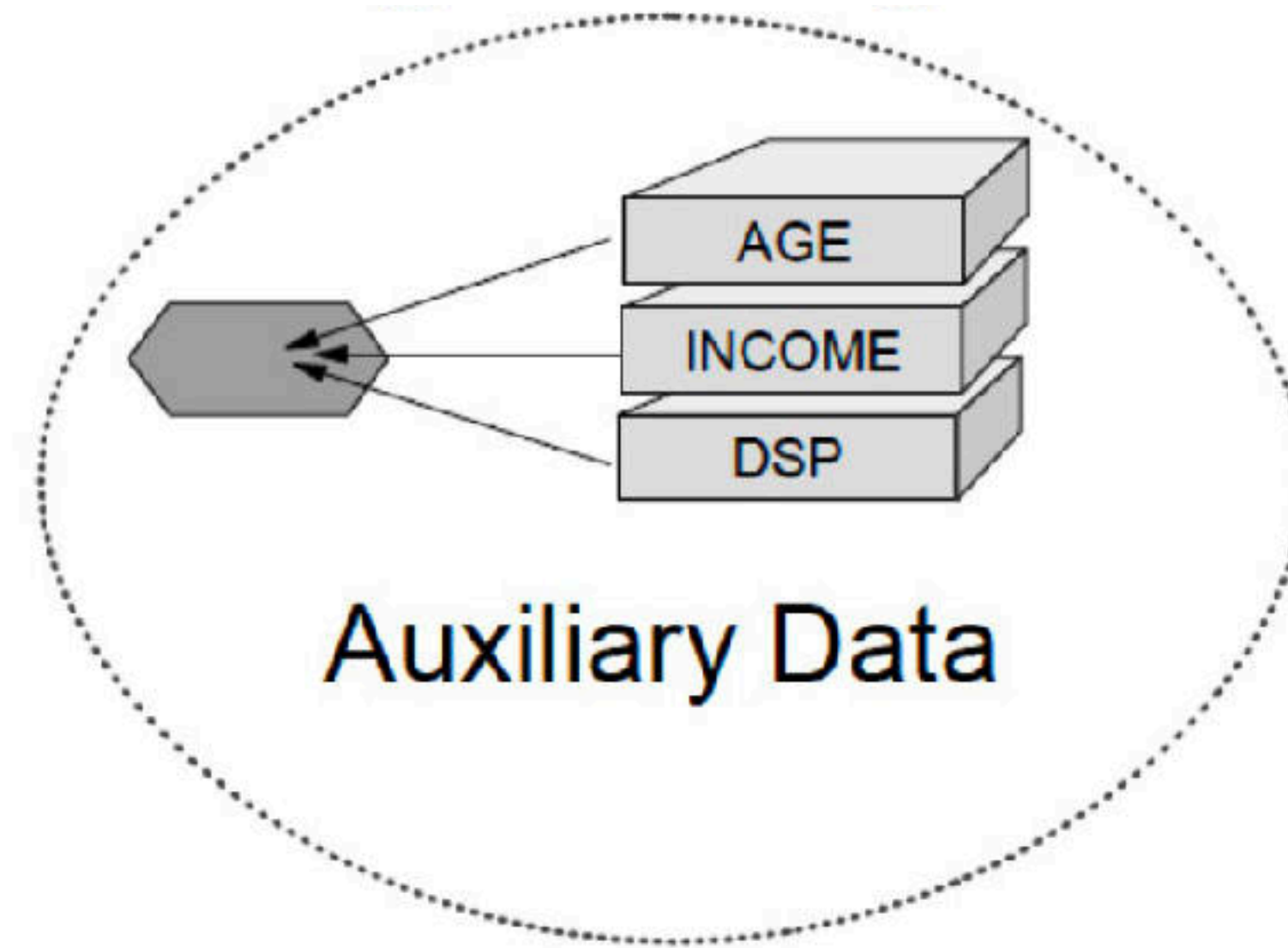
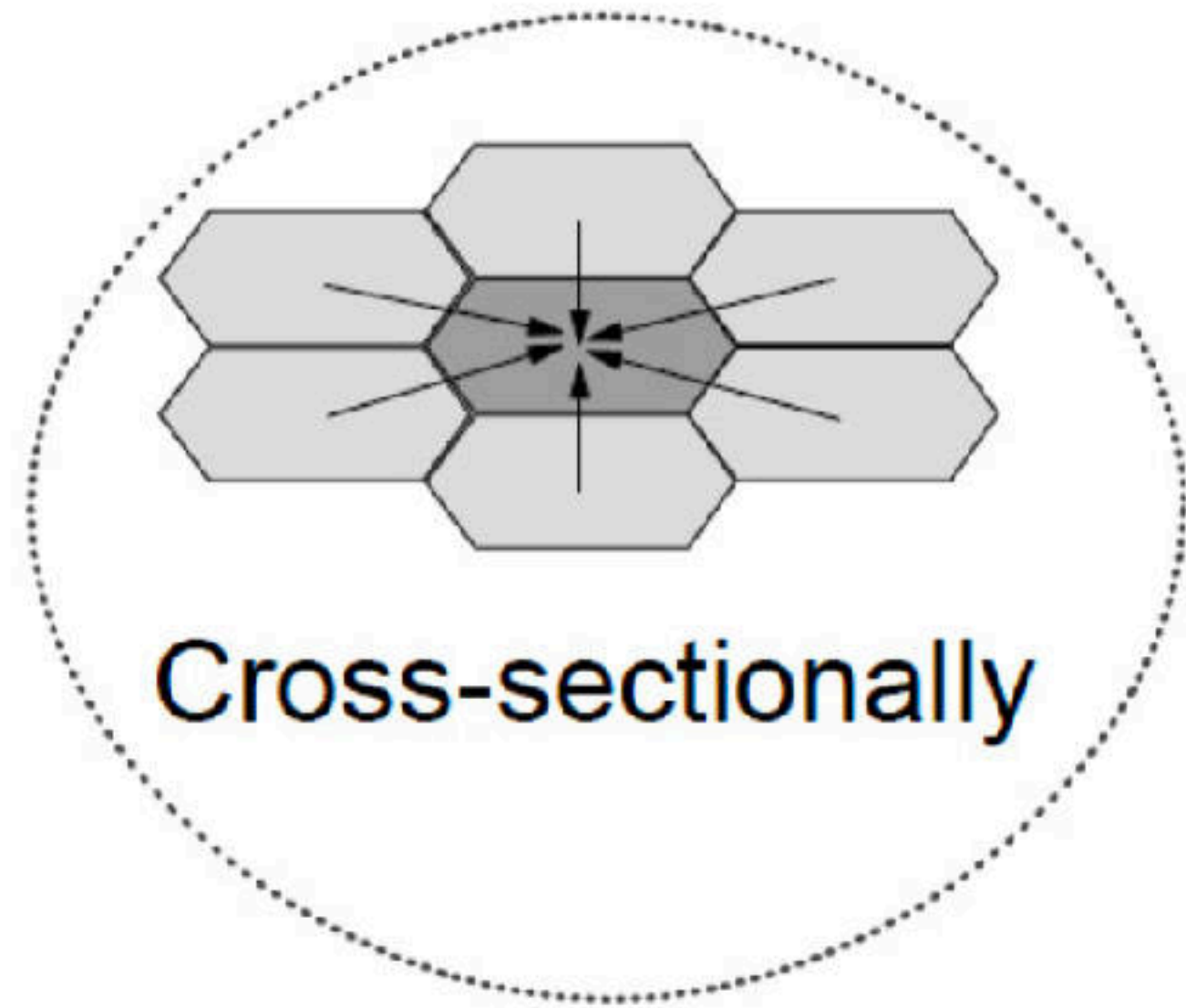
Diferentes tipos de estimación: áreas planeadas y de tamaño aleatorio

Fuente: adaptación de Methodology of Modern Business Statistics (2014)

Solución

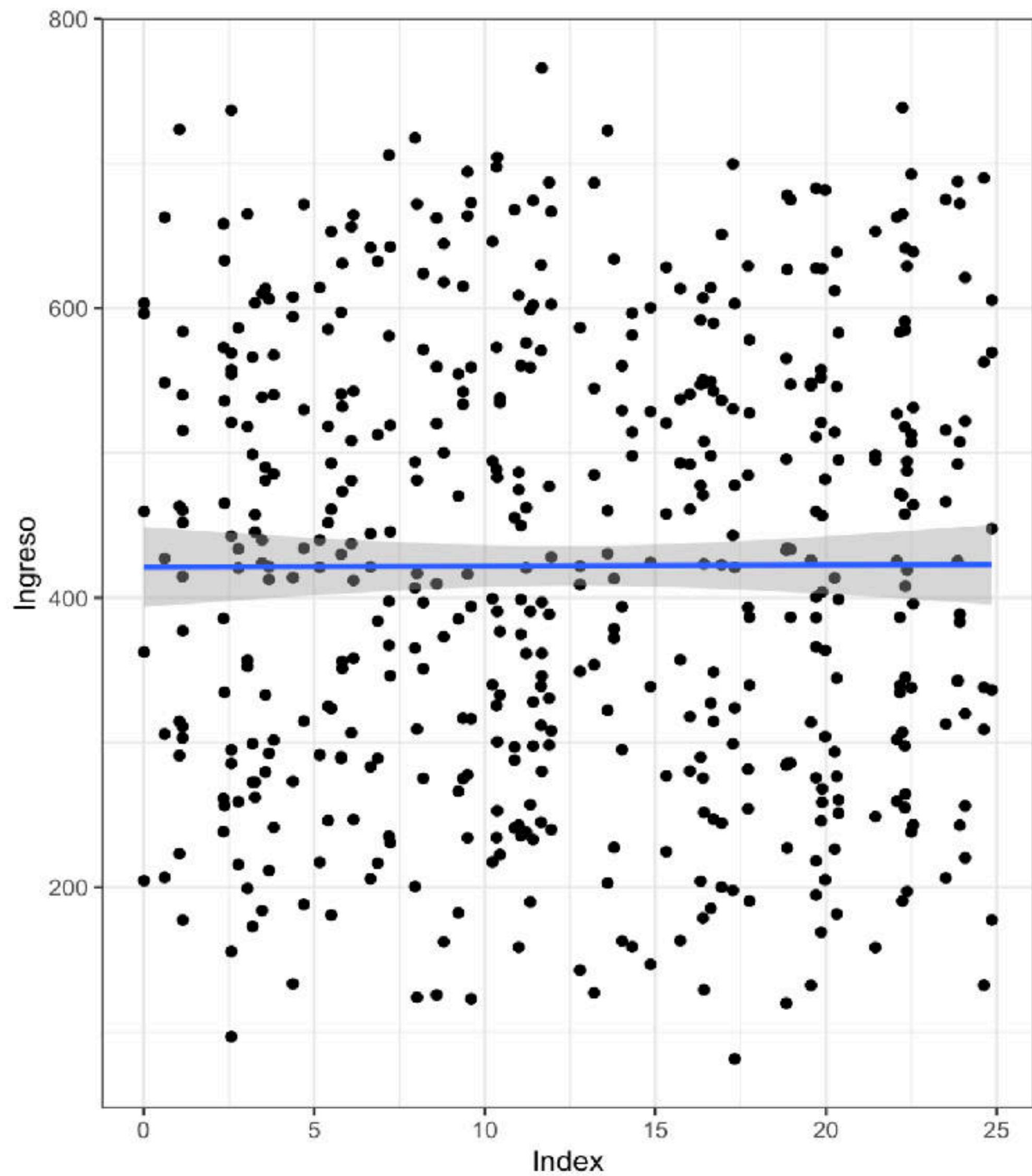
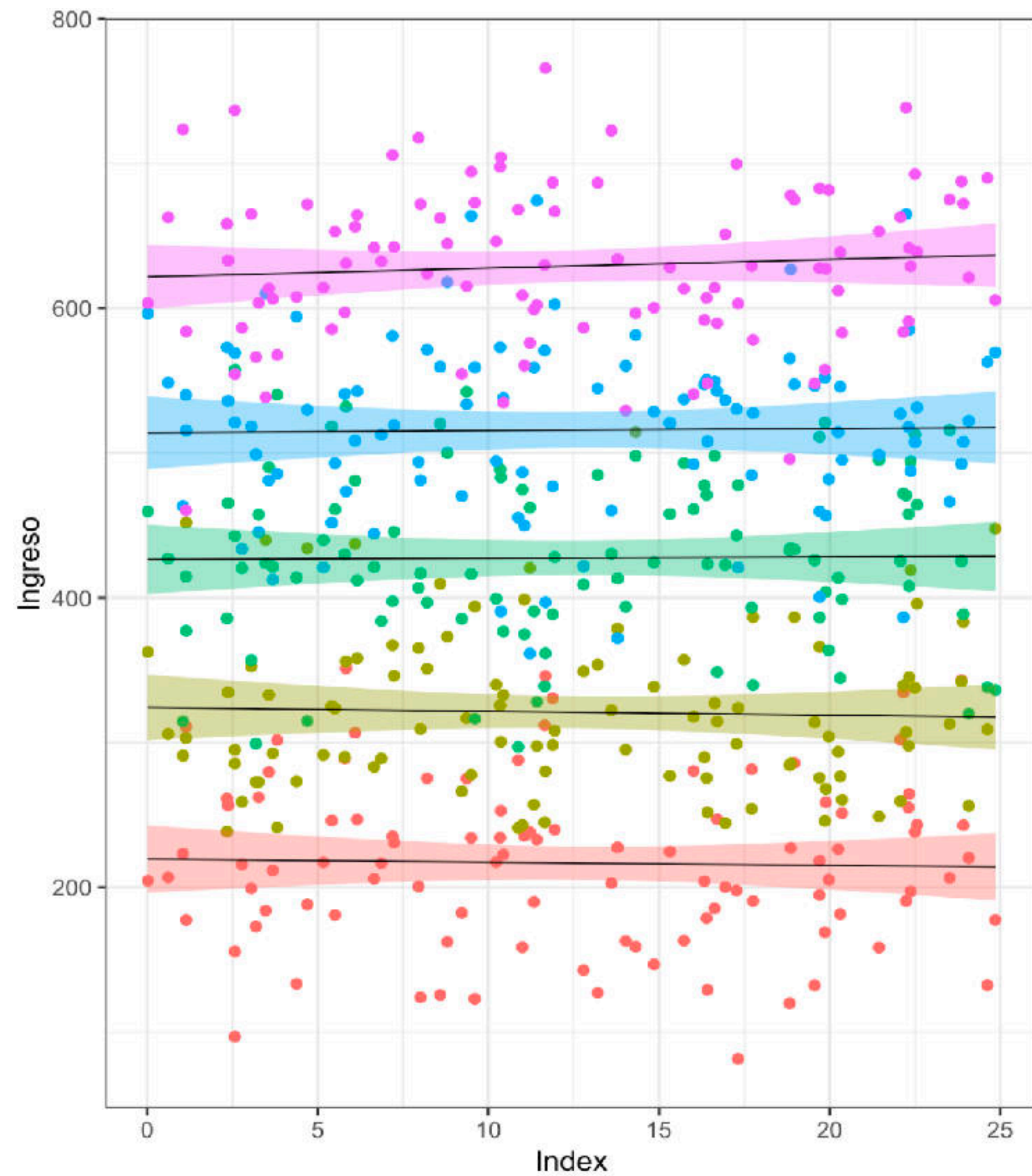
Cuando el tamaño de muestra no permite obtener estimaciones directas confiables para algunos dominios de interés, es posible abordar las siguientes opciones:

1. Aumentar el tamaño de muestra: esta opción tienen un efecto adverso en los costos que la pueden hacer inviable.
2. Utilizar metodologías estadísticas que involucren el uso de información auxiliar externa para obtener estimaciones (no directas) confiables en los subgrupos de interés.



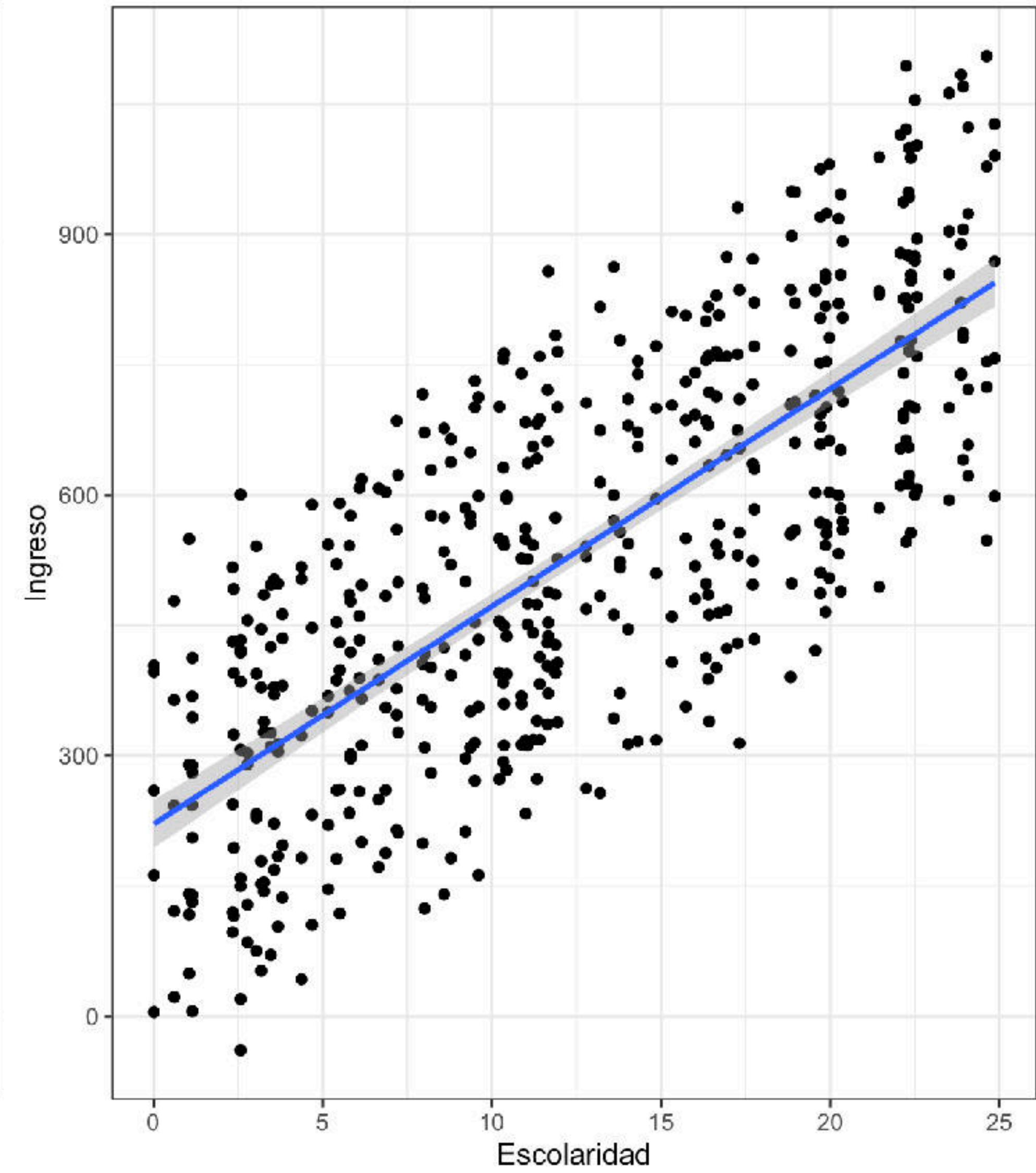
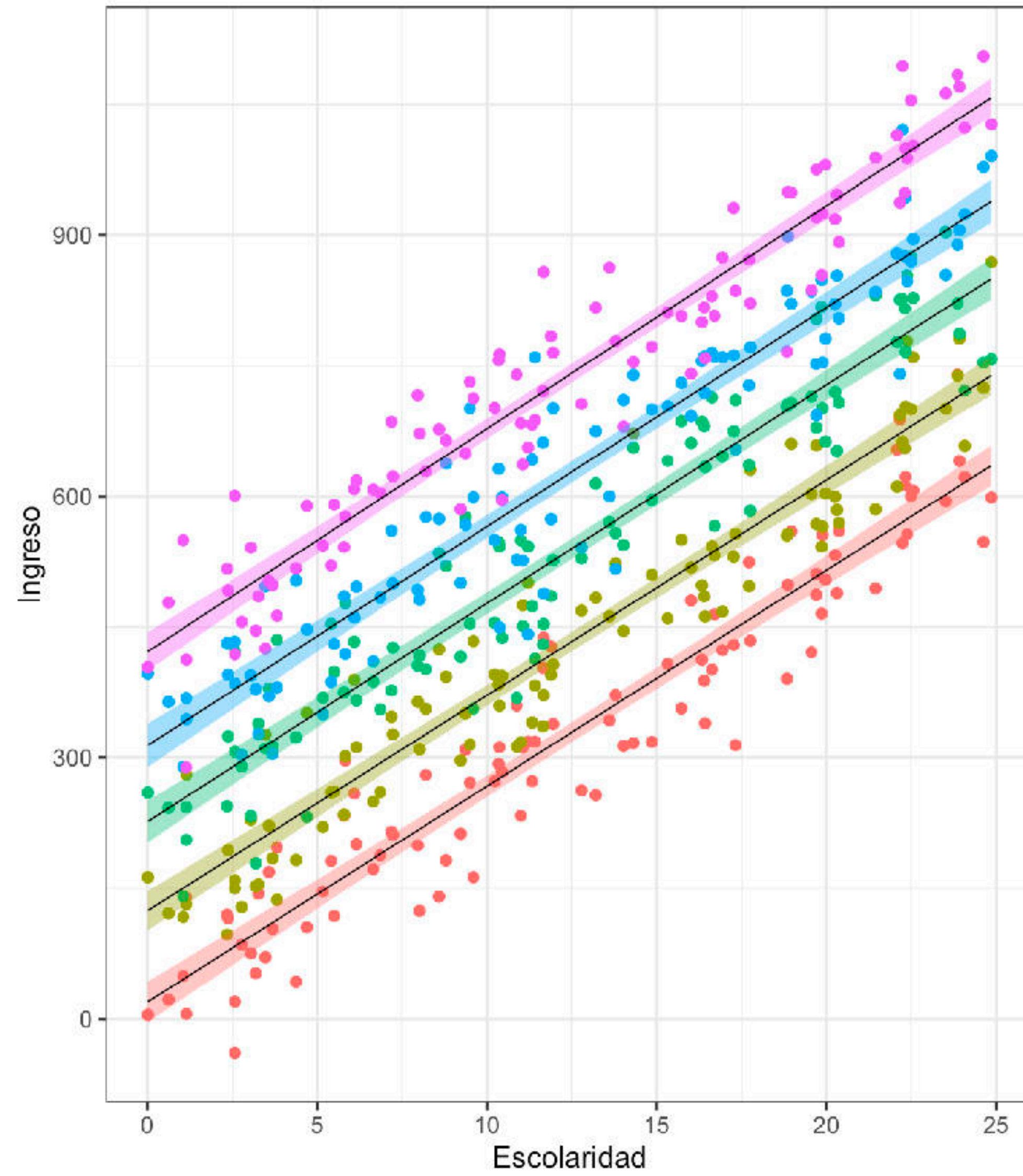
Borrowing strength: diferentes formas de apoyo en SAE

Fuente: Methodology of Modern Business Statistics (2014)



Fuerza prestada de otras áreas en una encuesta transversal

Fuente: elaboración propia



Fuerza prestada en presencia de información auxiliar

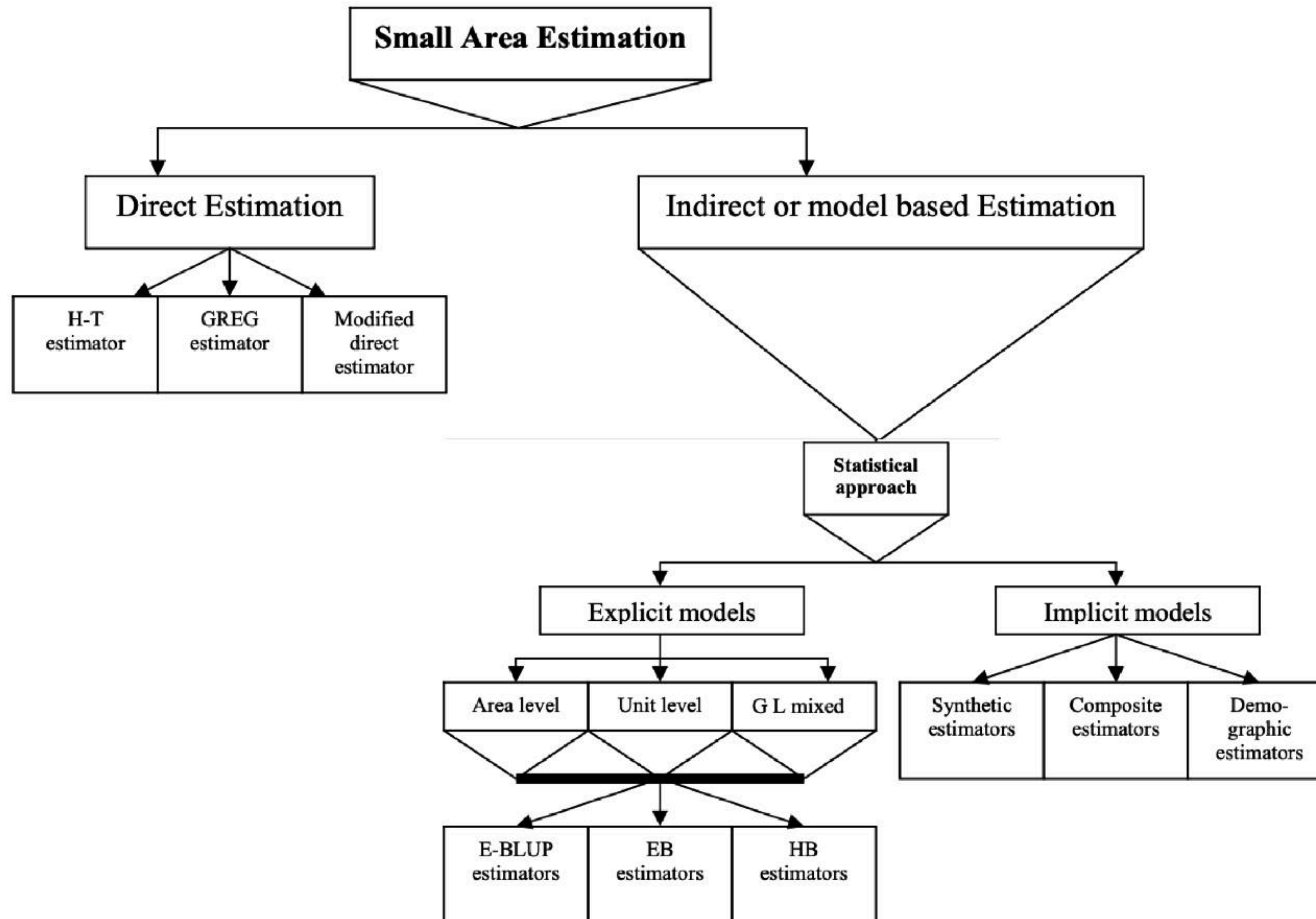
Fuente: elaboración propia

Metodologías SAE

Los estimadores de SAE se dividen en tres tipos principales:

1. Estimadores sintéticos indirectos
2. Estimadores basados en modelos de área
3. Estimadores basados en modelos de unidad

La escogencia del método que se debe utilizar en la estimación de los dominios de interés se realiza dependiendo del nivel en el que se encuentre la información auxiliar (a nivel de dominio o agregación - a nivel de hogar o persona)



Jerarquía de posibilidades y métodos en SAE

Fuente: adaptación de Rahman (2008)

Modelo de área (Fay-Herriot)

Fay & Herriot (1979) analizaron los ingresos per cápita para áreas pequeñas con menos de 1000 habitantes.

Modelo de vínculo:

$$\delta_d = \mathbf{x}'_d \boldsymbol{\beta} + u_d \quad u_d \sim_{ind} N(0, \sigma_u^2)$$

Modelo de muestreo:

$$\hat{\delta}_d^{DIR} = \delta_d + e_d = \mathbf{x}'_d \boldsymbol{\beta} + u_d + e_d \quad e_d \sim_{ind} N(0, \psi_d)$$

Modelo de unidad (Elbers-Lanjouw-Lanjouw)

Elbers et. al. (2003) analizaron indicadores de pobreza y desigualdad en Ecuador a nivel de parroquia (municipio) a través del siguiente modelo.

Modelo de vínculo:

$$Y_{di} = \log(E_{di} + c)$$

Modelo de muestreo:

$$Y_{di} = \mathbf{x}'_{di}\boldsymbol{\beta} + u_d + e_{di} \quad u_d \sim_{iid} (0, \sigma_u^2) \quad e_{di} \sim_{ind} (0, \sigma_e^2 k_{di}^2)$$

Modelo de unidad (*Empirical Best/Bayes*)

Molina & Rao (2010) estimaron las tasas y brechas de pobreza en el cruce de género y provincia en España por medio del siguiente modelo.

Modelo de vínculo:

$$\tilde{\delta}_d^B(\boldsymbol{\theta}) = E_{\mathbf{y}_{dr}}[\delta_d(\mathbf{y}_d) \mid \mathbf{y}_{ds}; \boldsymbol{\theta}]$$

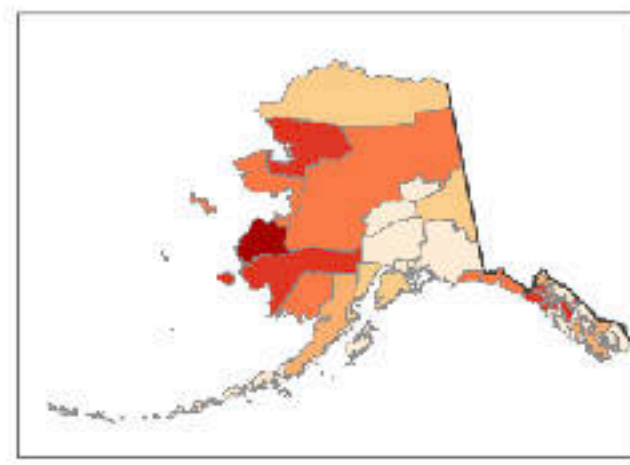
$$\mathbf{y}_{dr} \mid \mathbf{y}_{ds} \sim_{ind} N(\boldsymbol{\mu}_{dr|s}, \mathbf{V}_{dr|s})$$

Estimación condicional:

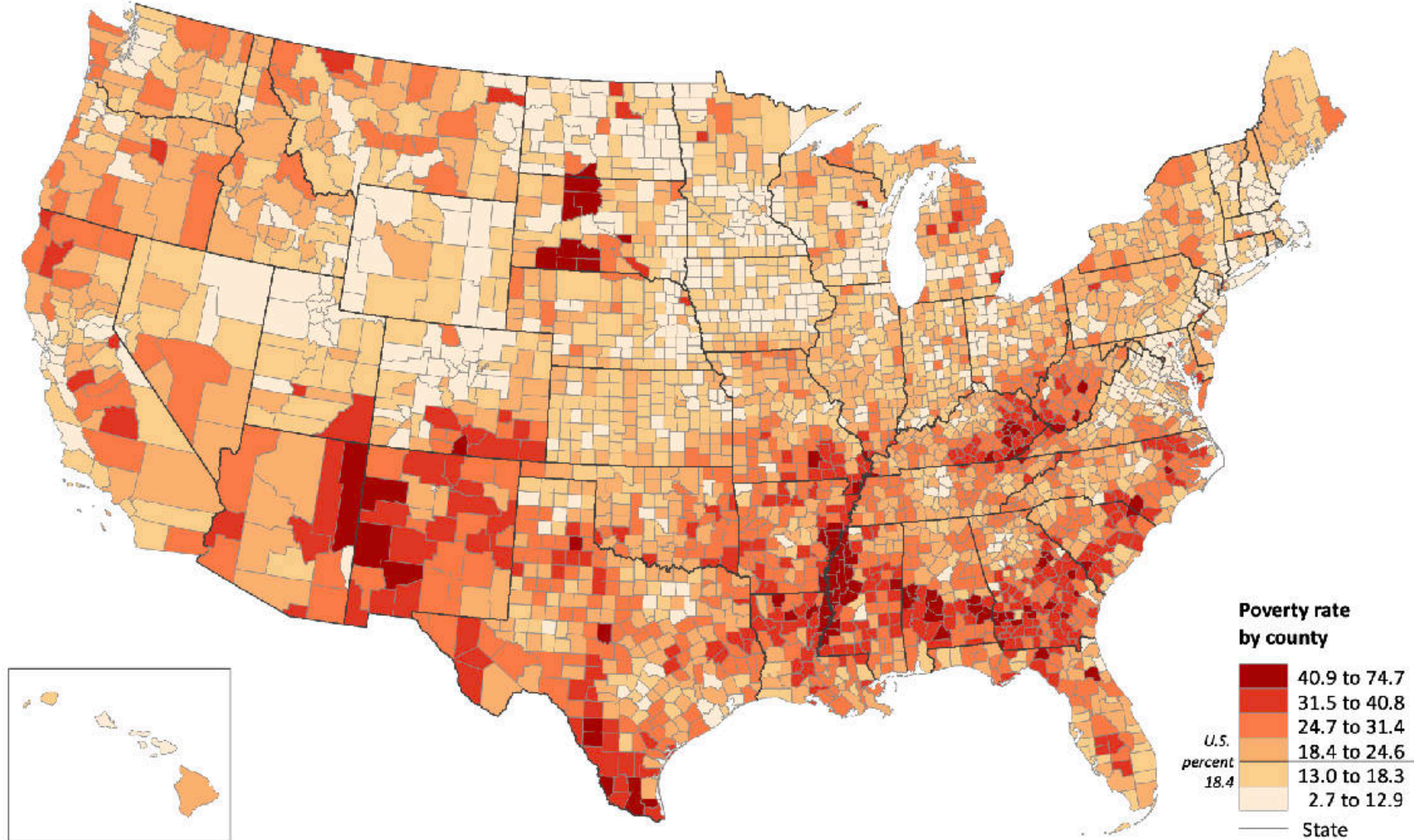
$$\boldsymbol{\mu}_{dr|s} = \mathbf{X}_{dr}\boldsymbol{\beta} + \gamma_d(\bar{y}_{da} - \bar{\mathbf{x}}_{da}^T\boldsymbol{\beta})\mathbf{1}_{N_d-n_d}$$

$$\mathbf{V}_{dr|s} = \sigma_u^2(1 - \gamma_d)\mathbf{1}_{N_d-n_d}\mathbf{1}_{N_d-n_d}^T + \sigma_e^2 \text{diag}_{i \in r_d}(k_{di}^2)$$

Algunos mapas de pobreza en América



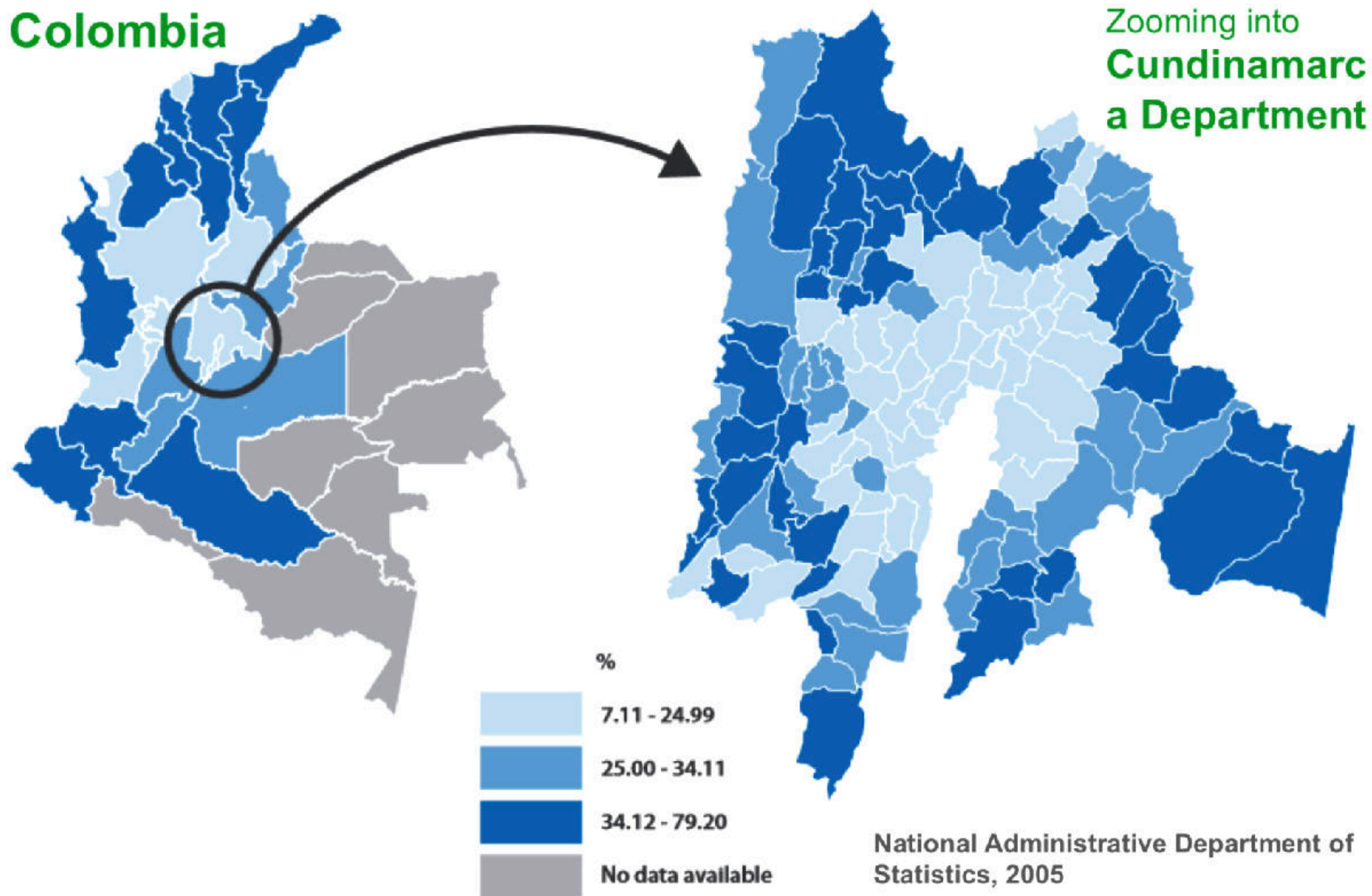
Poverty Rates of the Population Under 18
by County: 2017



Tasas de pobreza por condado en EE. UU. Modelo de área: Fay-Herriot

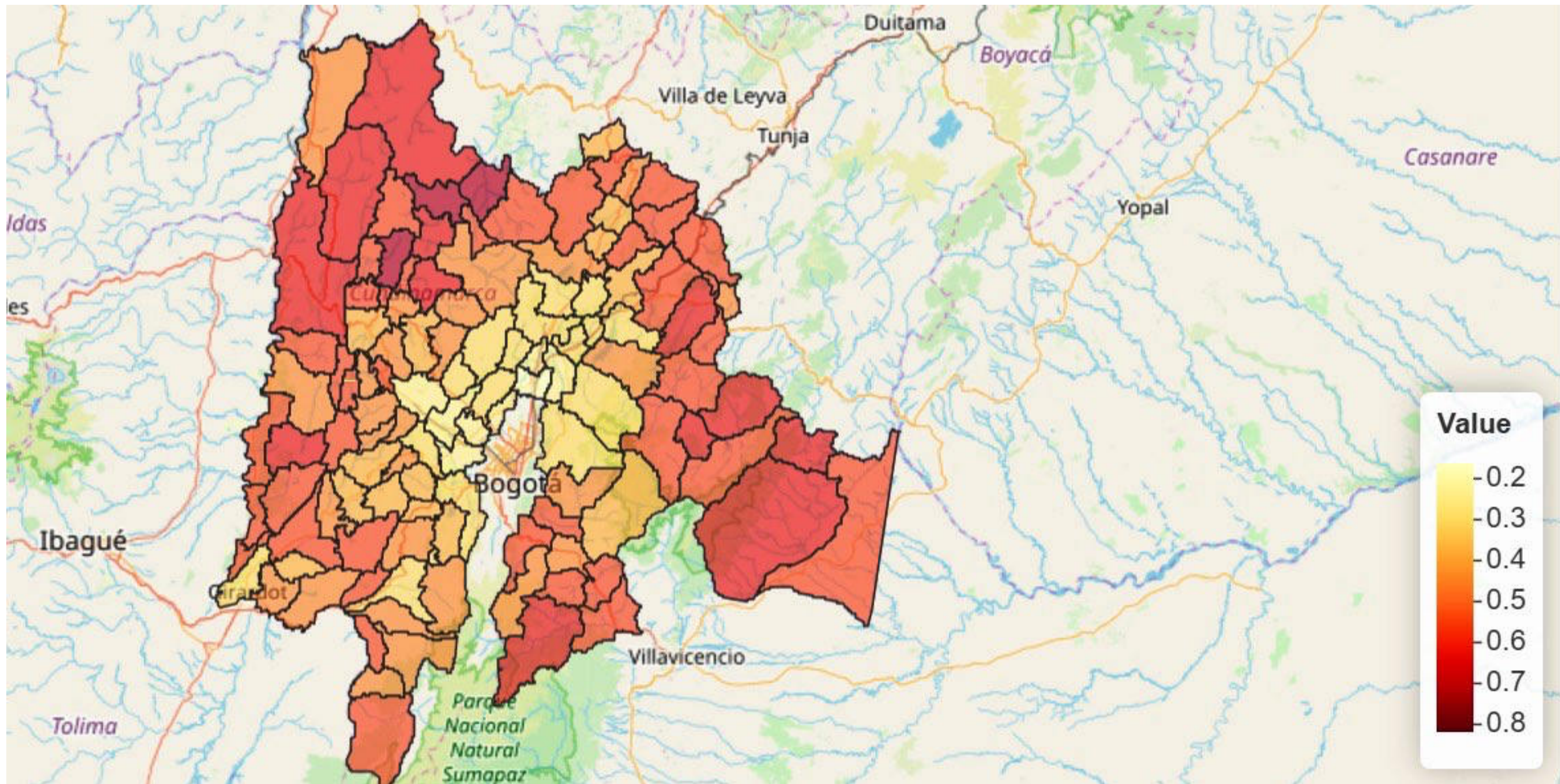
Fuente: US Census Bureau

Share of households per « Basic Unmet Needs » index, Colombia



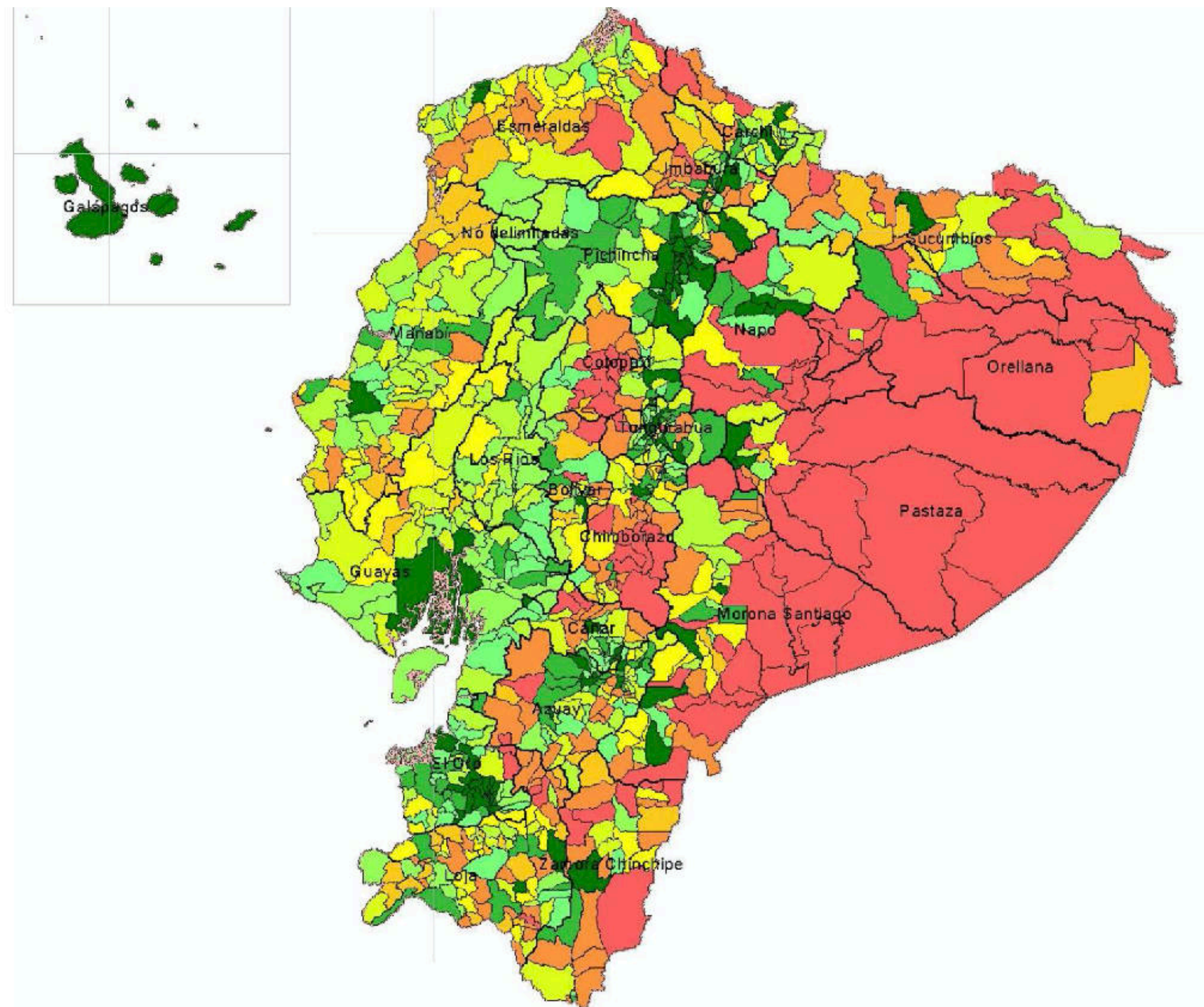
Desagregación de un indicador en áreas pequeñas en Colombia

Fuente: UNSD



Pobreza multidimensional en municipios de Cundinamarca **Modelo de área: Fay-Herriot espacial**

Fuente: DNP

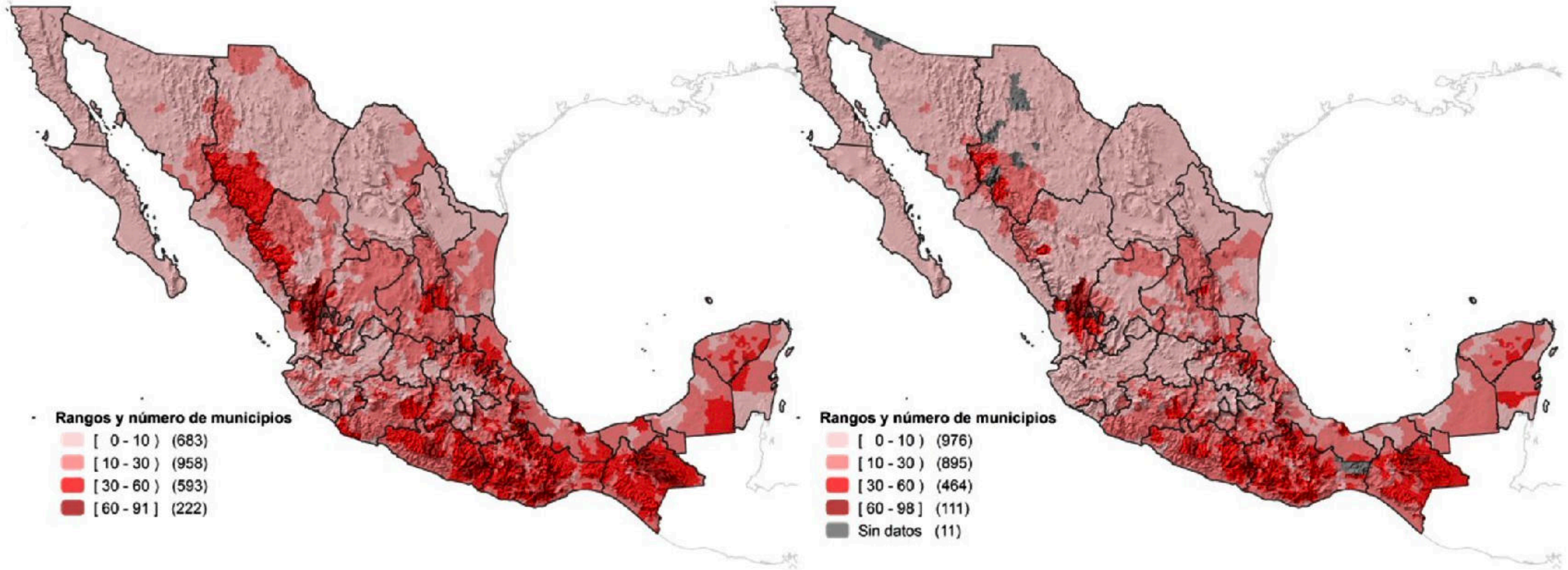


Tasas de pobreza en parroquias Ecuador Modelo de área: ELL - World Bank

Fuente: PovMap

2010

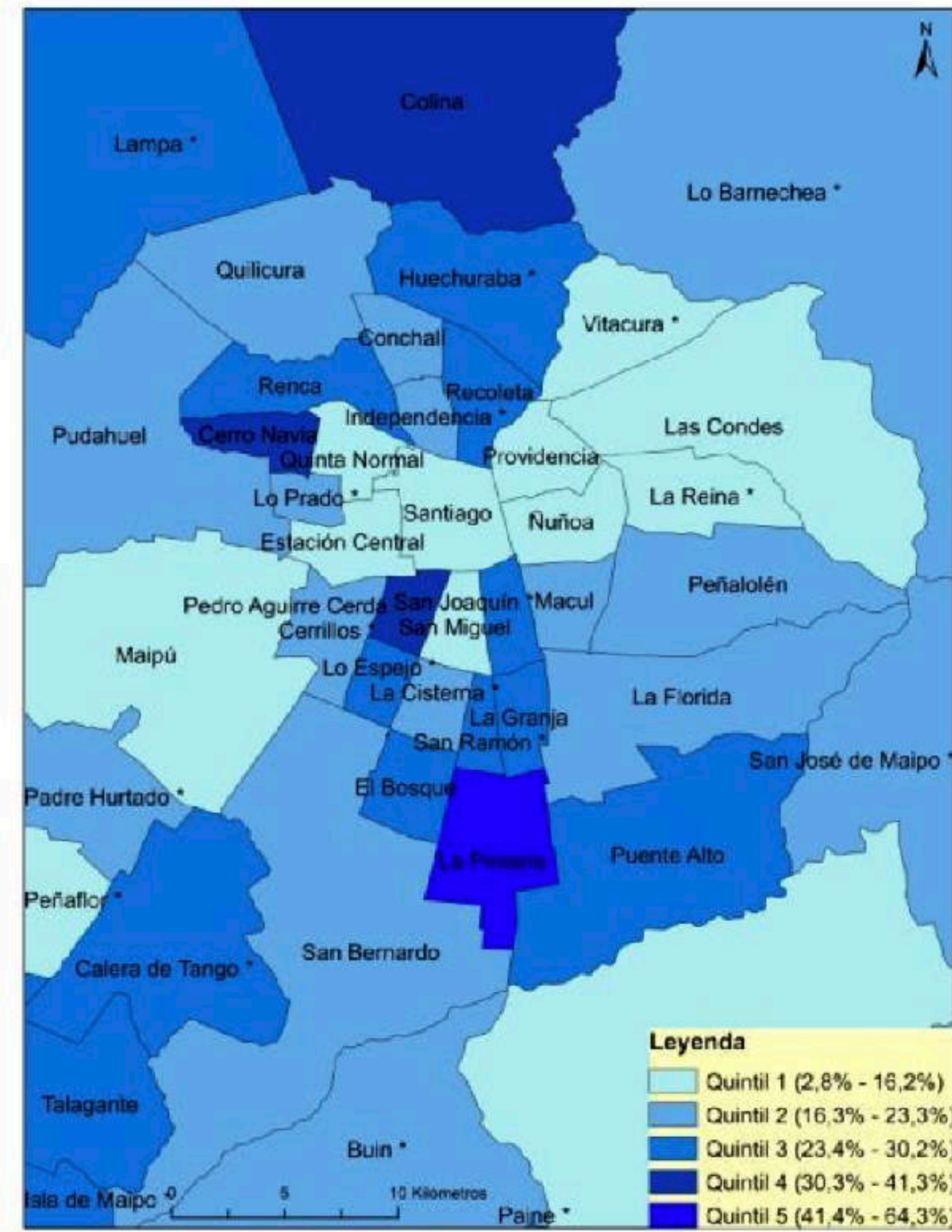
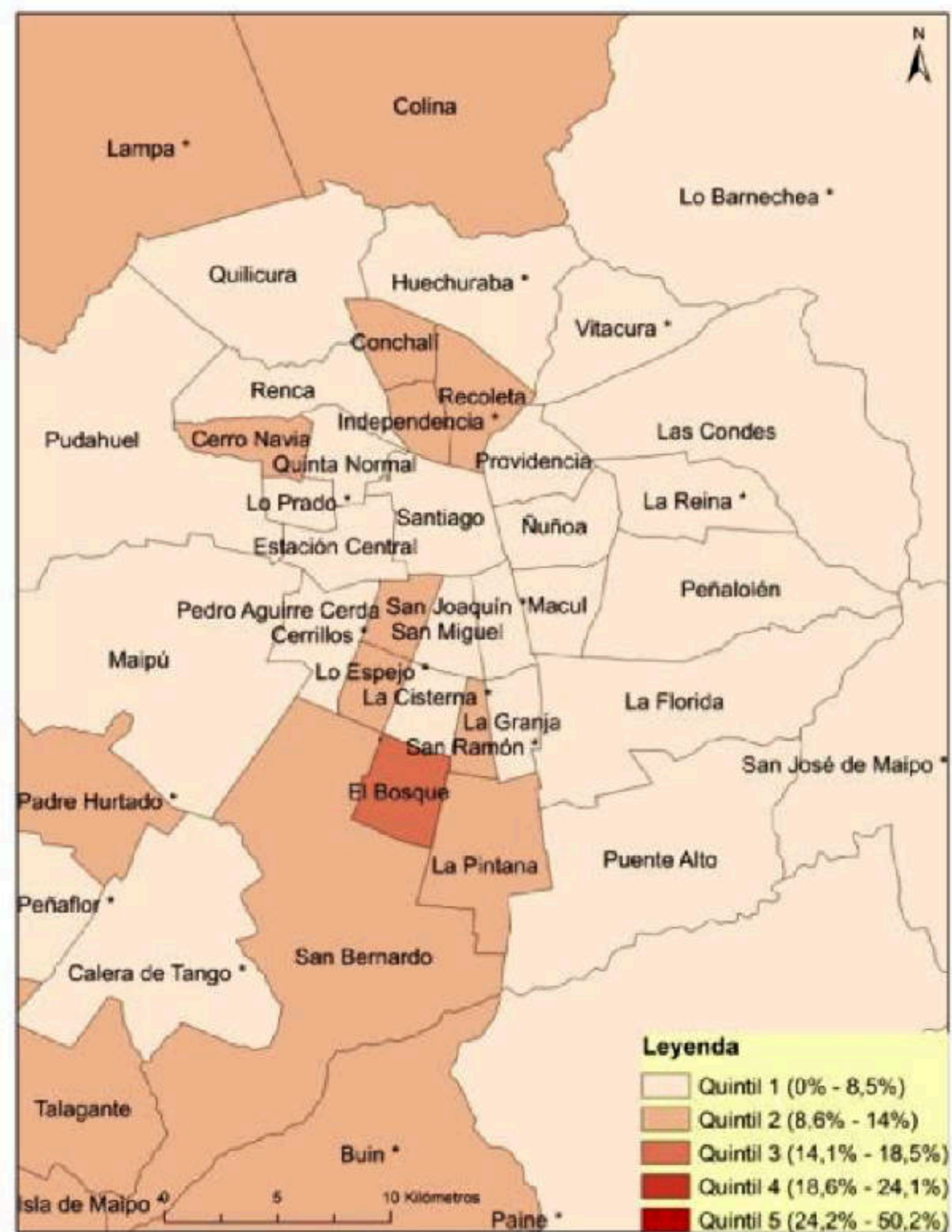
2015



Tasas de pobreza extrema por municipio en México

Modelo de unidad: *Empirical Best (Bayes) Predictor*

Fuente: CONEVAL



Pobreza y pobreza multidimensional por comuna en el Gran Santiago

Modelo de área: Fay-Herriot

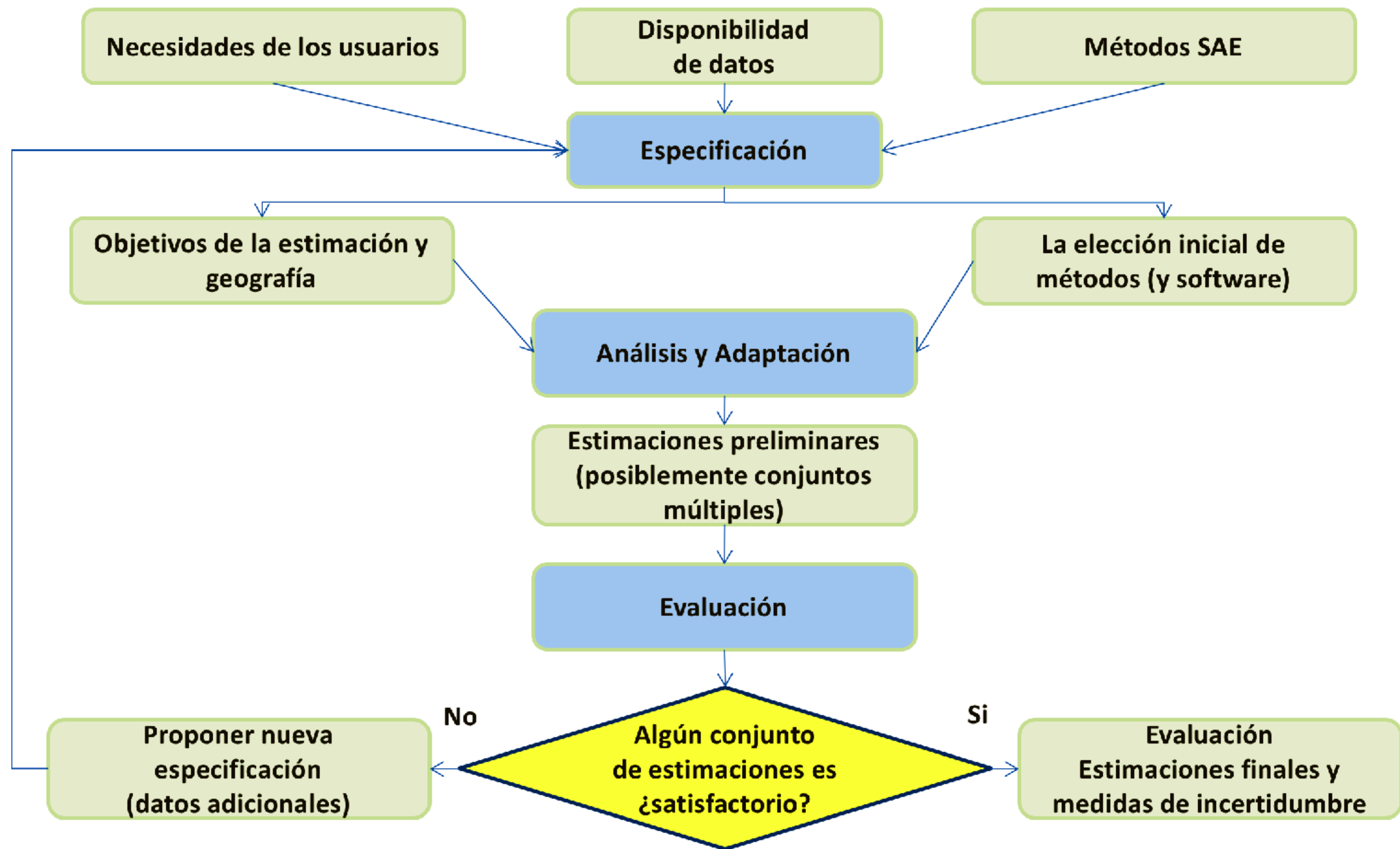
Fuente: MDS

Hacia un sistema integrado de producción estadística basado en SAE

¿Por qué SAE?

Dorfman (2018) argumenta que las metodologías de estimación de áreas pequeñas responden a:

- **Demanda** de información por parte de los hacedores de política, muchas veces para asignar recursos públicos en las áreas pequeñas.
- **Presupuestos limitados** (de tiempo, dinero y energía) que son insuficientes para recolectar suficientes datos que induzcan información precisa.



Proceso de estimación para la metodología SAE

Fuente: adaptación del INE (Chile) a partir de JRSS-A (2018)

Primera fase: especificación

- **Definir las necesidades de los usuarios:** especificación de los objetivos de estimación y la naturaleza de los indicadores (lineales o no lineales).
- **Revisar la disponibilidad de los datos:** identificar qué datos son requeridos ayuda a planificar la carga de trabajo de los funcionarios de las ONE.
- **Actualización de las metodologías SAE:** la definición de las necesidades y la definición de información disponible podría incrementar la dificultad de las metodologías de estimación.

Segunda fase: análisis y adaptación

- **Definir la tripleta inicial de estimaciones:** empieza por revisar las estimaciones directas, sintéticas y compuestas. En esta etapa se escogen los estimadores con mayor precisión para ser utilizados posteriormente en los modelos de unidad o área.
- **Uso de modelos SAE:** cuando se hace necesario se incorporan modelos estocásticos para construir predictores de los indicadores de interés con base en los dominios pequeños para los cuales se requieren las estimaciones. Se distinguen los modelos de Fay-Herriot, del banco mundial y de Battese-Harter-Fuller.

Tercera fase: evaluación

- **Análisis de los primeros conjuntos de estimaciones:** a través de la revisión de los supuestos de los modelos escogidos, como transformaciones sobre la variable de interés o el análisis de residuales.
- **Propuesta de nuevas especificaciones o datos adicionales:** de acuerdo a los hallazgos encontrados en la identificación del modelo, es posible que se requiera la introducción de más covariables en el modelo o de nuevas interacciones entre las covariables.

¿Fuerza prestada o debilidad?

Al utilizar SAE se supone que el aumento de la **demanda** y el decremento del **presupuesto** pueden reconciliarse.

¿Pero hasta qué punto?

- Se debe evitar el ciclo nocivo de recortar los presupuestos y aumentar la demanda, supeditado a un equipo de estadísticos que asegura la eficacia de la metodología.
- Si un país entra en este ciclo nocivo, ya no se tomaría fuerza prestada de las áreas grandes.
- Al debilitar las encuestas, las metodologías de SAE expandirían la debilidad de las áreas grandes a las áreas pequeñas.

La respuesta pendiente: criterio de balance

Así como las encuestas y los censos requieren un presupuesto, los programas de SAE también. Para estas operaciones hay criterios de precisión y exclusión:

- En las encuestas: el coeficiente de variación estimado.
- En los censos: las tasas de cobertura.

Se debería encontrar respuesta a la siguiente pregunta:

¿Cuál será el criterio estadístico que indique si el presupuesto de una encuesta es insuficiente para producir estimaciones de áreas pequeñas satisfactorias?

Ventajas de la estimación SAE

Entre otras, se consideran las siguientes ventajas en los sistemas de estimación que se basan en SAE.

- Estabilidad en el tiempo.
- Menores errores de estimación/predicción.
- *Benchmarked*.
- Basadas en un modelo ajustado a los datos.
- Robustez (áreas grandes que fueron muestreadas).

Propuesta de evaluación

Dorfman (2018) afirma que la falta de datos, justo en donde se necesitan, afecta la verificación de los supuestos de los modelos basados en SAE y propone las siguientes sugerencias a las ONE:

- **Muestra de muestras:** cada encuesta que sea usada para realizar estimación SAE debería asignar una pequeña parte de sus recursos para seleccionar una muestra suplementaria de algunas áreas pequeñas.
- **Medición de precisión y cobertura:** con base en la *muestra suplementaria* es posible evaluar si la estimación SAE se acerca al verdadero parámetro y si el intervalo de confianza SAE lo cubre.

Plan de capacitación, entrenamiento y acompañamiento

**Modelamiento
estadístico
(GLM - Mixtos)**

GIS & mapping

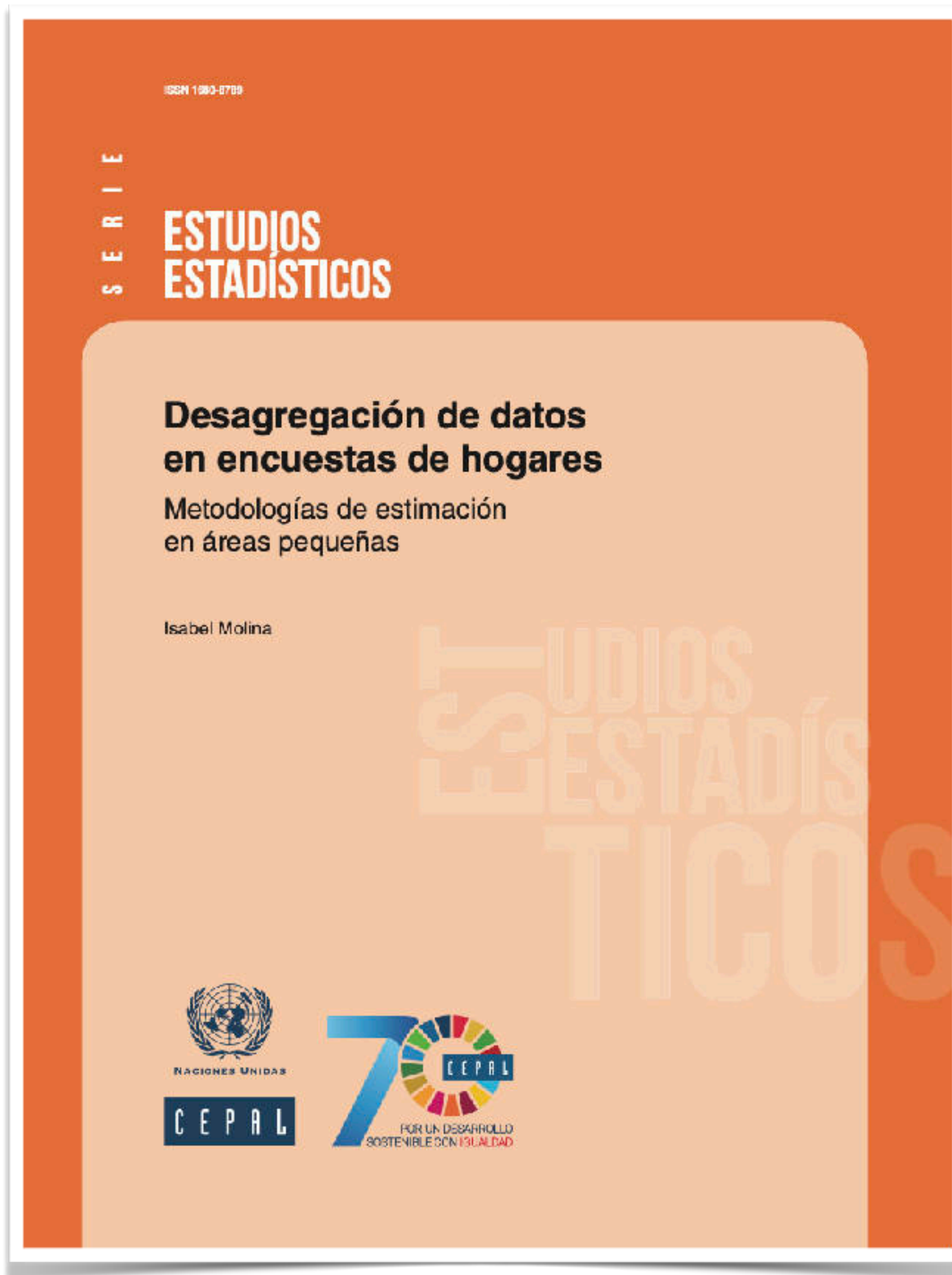
**Estimación en
áreas pequeñas**

**Muestreo y análisis
de encuestas**

**Optimización y
software
(Bayes - Bootstrap)**

Síntesis de los procesos involucrados en la producción de datos con SAE

Fuente: adaptación de Kolenikov (2014)



NACIONES UNIDAS

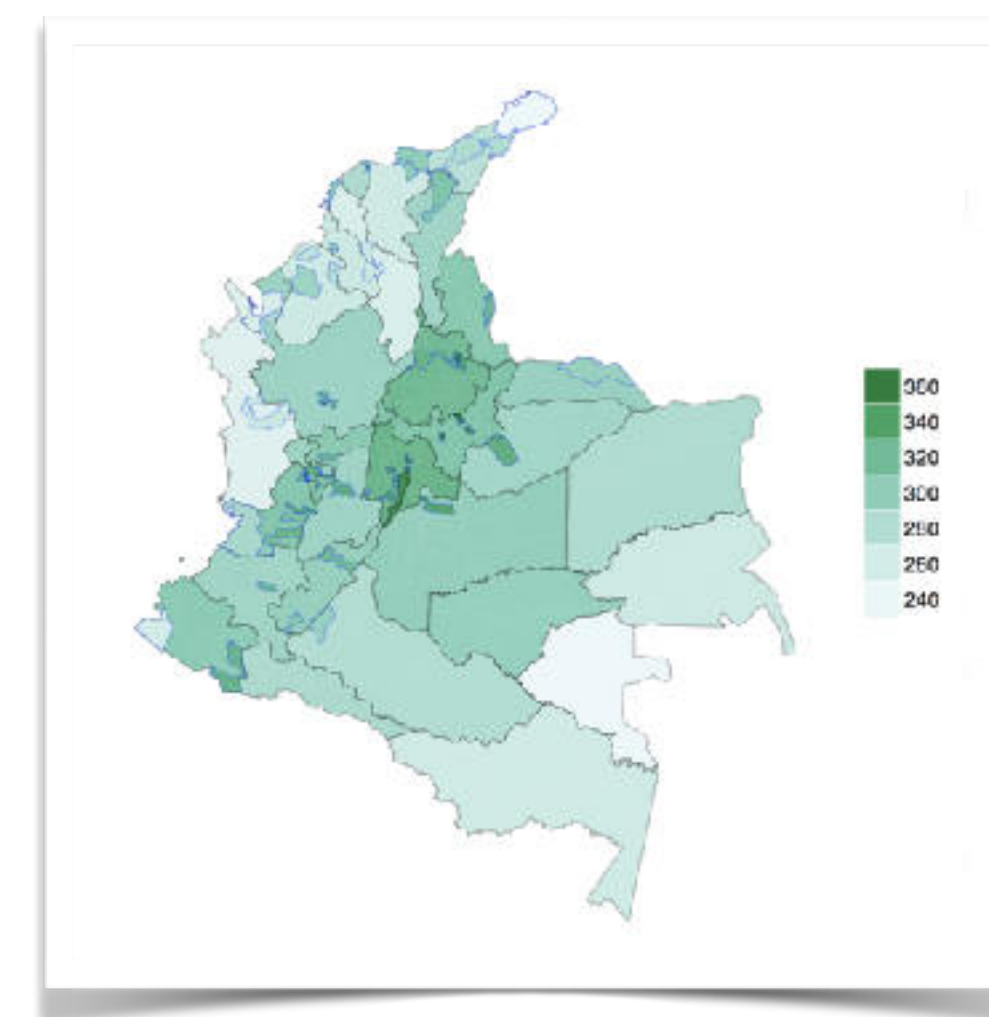
CEPAL

No dejar a nadie atrás

Integración de las fuentes de información para la estimación desagregada de estadísticas del trabajo

Unidad de Estadísticas Sociales
División de Estadísticas

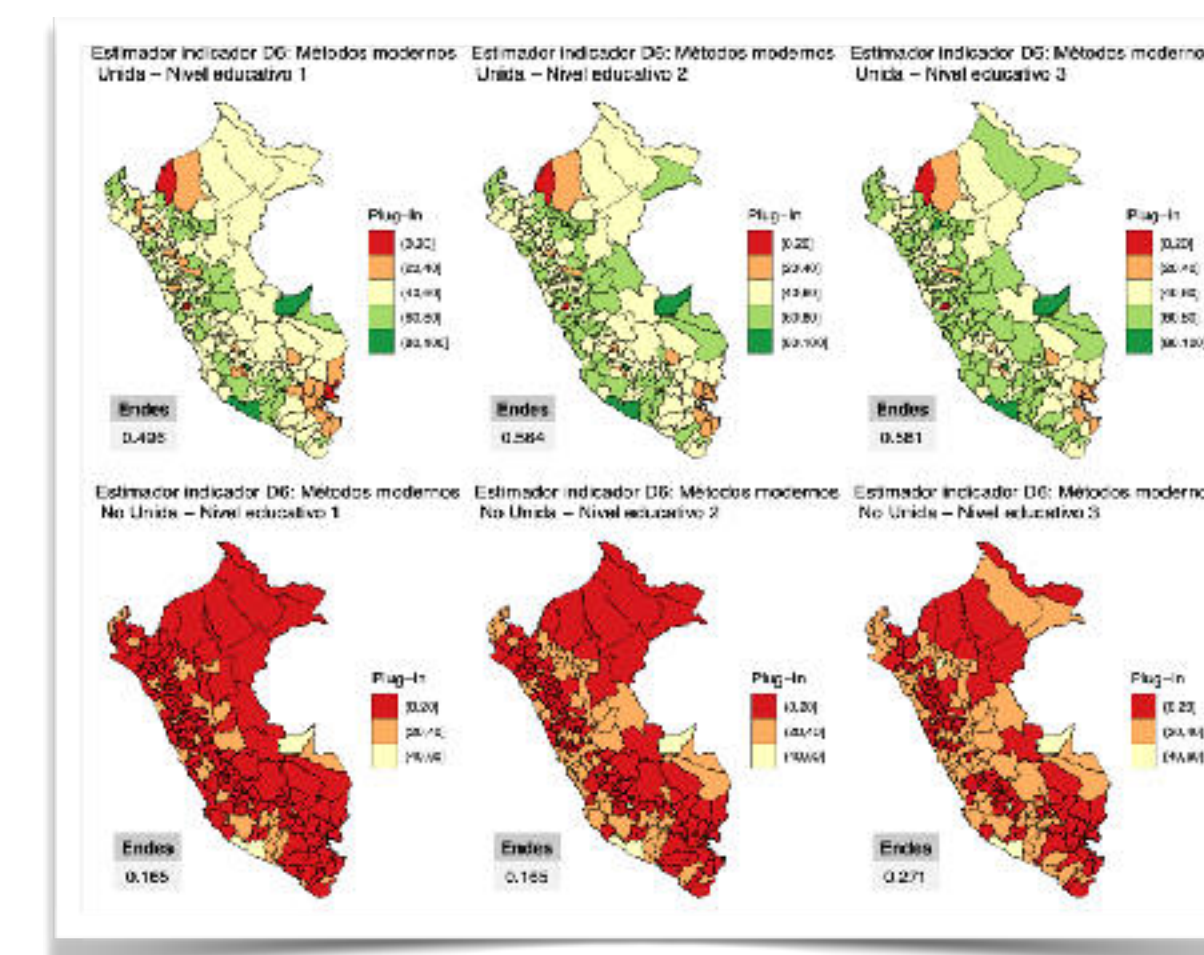
Andrés Gutiérrez, PhD
Experto Regional en Estadísticas Sociales
andres.gutierrez@un.org



```

15 sigma <- 10
16 beta1 <- 0
17 beta0 <- 200
18 x1 <- runif(N, 0, 25)
19
20 y1 <- beta0 + 20 + beta1 * x1 + rnorm(N, 0, sigma)
21 y2 <- beta0 + 120 + beta1 * x1 + rnorm(N, 0, sigma)
22 y3 <- beta0 + 220 + beta1 * x1 + rnorm(N, 0, sigma)
23 y4 <- beta0 + 320 + beta1 * x1 + rnorm(N, 0, sigma)
24 y5 <- beta0 + 420 + beta1 * x1 + rnorm(N, 0, sigma)
25 ID <- rep(LETTERS[1:5], each = N)
26
27 test <- data.frame(Index = c(x1),
28                   Ingreso = c(y1, y2, y3, y4, y5), ID = ID)
29
30 p1 <- ggplot(data = test, aes(Index, Ingreso)) +
31   geom_point() + geom_smooth(method=lm) + theme_bw()
32

```



Desarrollo de un sistema de estimación SAE en ALC

Fuente: elaboración propia

Asistencia en SAE

1. Ecuador: tasas de desnutrición a nivel cantonal (ODS 2).
 - Modelos FH, BHF.
 2. Chile: tasas de pobreza desagregada a nivel comunal (ODS 1).
 - Modelo FH.
 3. Chile: tasas de victimización desagregada a nivel comunal (ODS 5 y 16).
 - Modelo FH, EBP.
 4. Perú: indicadores de planificación familiar a nivel provincial (ODS 3).
 - Modelo EBP.
 5. Colombia: tasas de acceso a la justicia a nivel municipal (ODS 16).
 - Modelo FH.
- Capacitaciones en Uruguay, El Salvador, Colombia y Chile.

Capacitación teórica y entrenamiento en la ONE

Acompañamiento en estimaciones puntuales con encuestas de hogares

xavier.mancero@cepal.org
andres.gutierrez@cepal.org

¡Gracias!
