

# Desagregación de Estimaciones en Áreas Pequeñas un enfoque bayesiano

Invalid Date



# OBJETIVOS DE DESARROLLO SOSTENIBLE



## Algunas metas del ODS2 (Hambre cero)

De aquí a 2030, poner fin al hambre y asegurar el acceso de todas las personas, en particular los pobres y las personas en situaciones de vulnerabilidad, incluidos los niños menores de 1 año, a una alimentación sana, nutritiva y suficiente durante todo el año.

- ▶ Prevalencia de la subalimentación.
- ▶ Prevalencia de la inseguridad alimentaria moderada o grave en la población, según la Escala de Experiencia de Inseguridad Alimentaria.

## Algunas metas del ODS8 (Empleo decente)

De aquí a 2030, lograr el empleo pleno y productivo y el trabajo decente para todas las mujeres y los hombres, incluidos los jóvenes y las personas con discapacidad, así como la igualdad de remuneración por trabajo de igual valor.

- ▶ Tasa de desempleo, desglosada por sexo, edad y personas con discapacidad.

# 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.

## ¿Qué es el coeficiente de variación?

El coeficiente de variación es una medida de error relativo a un estimador, se define como:

$$cve\left(\hat{\theta}\right)=\frac{se\left(\hat{\theta}\right)}{\hat{\theta}}$$

Muchas veces se expresa como un porcentaje, aunque no está acotado a la derecha, y por eso es conveniente a la hora de hablar de la precisión de una estadística que viene de una encuesta.

## Estándares de alerta en algunos países (encuestas de hogares)

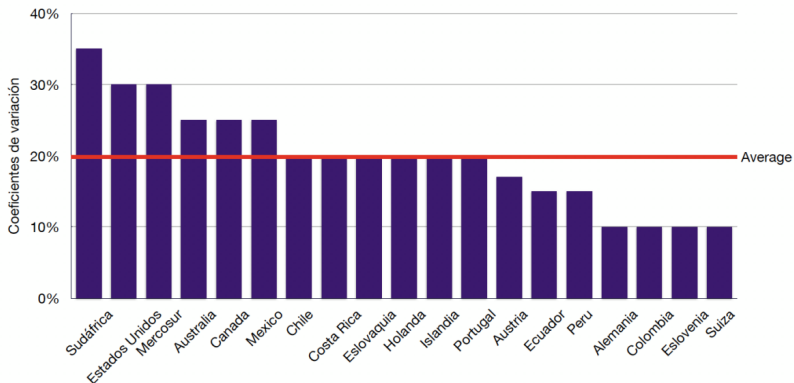


Figure 1: Alertas sobre los coeficientes de variación



## Algunas alertas definidas en la publicación

Cuando se sobrepasa el umbral del coeficiente de variación aparecen algunas de las siguientes alertas:

- ▶ No se publica
- ▶ Usar con precaución.
- ▶ Las estimaciones requieren revisiones, no son precisas y se deben usar con precaución.
- ▶ Poco confiable, menos preciso.
- ▶ No cumple con los estándares de publicación.
- ▶ Con reserva, referencial, cuestionable.
- ▶ Valores muy aleatorios, estimación pobre.

## Dominios de estudio y subpoblaciones de interés

Una encuesta se planea con el fin de generar información precisa y confiable en los dominios de estudio que se han predefinido. Sin embargo, existen subgrupos poblacionales que la encuesta no abordó en su diseño, y sobre los cuales se quisiera una mayor precisión.

- ▶ Incidencia de la pobreza desagregado por departamento o provincia (tamaño de muestra conocido y planificado).
- ▶ Tasa de desocupación desagregada por sexo (tamaño de muestra aleatorio, pero planificado).
- ▶ Tasa de asistencia neta estudiantil en primaria desagregada por quintiles de ingreso (tamaño de muestra aleatorio).

# Precisión de los estimadores

Debido a que una encuesta es una investigación parcial sobre una población finita, es necesario saber que:

- ▶ A partir de una encuesta, no se calculan indicadores, sino que se estiman con ayuda de los datos de la encuesta.
- ▶ Es necesario calcular el grado de error que se comete al no poder realizar una investigación exhaustiva. Este error es conocido como el error de muestreo.
- ▶ La precisión de un estimador está supeditada al intervalo de confianza.

Entre más angosto sea el intervalo, más precisión se genera y por ende se tiene un menor error de muestreo.

## El tamaño de muestra efectivo

- ▶ En las encuestas de hogares, con diseños de muestreo complejos, no existe una sucesión de variables que sean independientes e idénticamente distribuidas.
- ▶ La muestra  $y_1, \dots, y_n$  no es un vector en el espacio  $n$ -dimensional, donde se asume que cada componente del vector puede variar por sí mismo.
- ▶ La dimensión final del vector  $(y_1, \dots, y_n)$  es mucho menor que  $n$ , puesto que existe una forma jerárquica en la selección de los hogares y a la interrelación de la variable de interés con las UPMs

## El tamaño de muestra efectivo

El tamaño de muestra efectivo se define como sigue:

$$n_{efectivo} = \frac{n}{Deff}$$

En donde Deff es el efecto de diseño que depende de: 1. El número de encuestas promedio que se realizaron en cada UPM. 2. La correlación existente entre la variable de interés y las mismas UPMs.

Es posible considerar que, si el tamaño de muestra efectivo no es mayor a un umbral, entonces la cifra no debería ser considerada para publicación.

## Grados de libertad

En las subpoblaciones los grados de libertad no se consideran fijos sino variables.

$$gl = \sum_{h=1}^H v_h \times (n_{Ih} - 1)$$

Note que  $v_h$  es una variable indicadora que toma el valor uno si el estrato  $h$  contiene uno o mas casos de la subpoblación de interés,  $n_{Ih}$  es el número de UPMs en el estrato. En el caso más general, los grados de libertad se reducen a la siguiente expresión:

$$gl = \#UPMs - \#Estrato$$

## Uso de métodos SAE

# Justificación

- ▶ Los estimadores directos, basados solo en unidades de muestreo observadas para cada área pequeña, no son suficientemente confiables.
- ▶ Tamaño de muestra pequeño o incluso ninguna unidad observada (falta de información).
- ▶ El coeficiente de variación (CV) es demasiado alto para el indicador objetivo a nivel de área.



## Incremento del coeficiente de variación

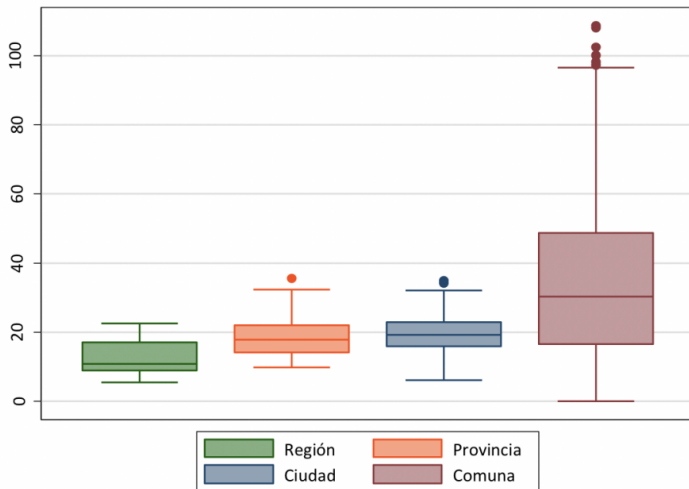


Figure 2: Distribución de los coeficientes de variación en Chile

# Justificación

Cuando los estimadores directos no son confiables para algunos dominios de interés, existen dos opciones:

- 1 Sobremuestreo: aumentar el tamaño de la muestra en los dominios de interés (aumento de los costos).
- 2 Aplicar técnicas estadísticas que permitan estimaciones confiables en esos dominios, métodos SAE.

## ¿Qué es un área pequeña?

- ▶ La mayoría de las encuestas nacionales están planificadas para entregar estimaciones confiables a nivel nacional y regional pero a niveles más bajos se reduce la precisión.
- ▶ Un área pequeña es un dominio para el cual el tamaño de muestra específico no es suficientemente grande para obtener estimaciones confiables.
- ▶ Habitualmente son dominios no planificados y su tamaño de muestra esperado es aleatorio y es más grande a medida que aumenta el tamaño de la población del área.

## ¿Qué es un área pequeña?

La subpoblación de interés puede ser una zona geográfica o subgrupos socioeconómicos.

- ▶ Geográfico: provincias, áreas del mercado de trabajo, municipios, sectores censales para medir por ejemplo la tasa de desempleo a nivel comunal.
- ▶ Dominio de subgrupos específicos: edad  $\times$  sexo  $\times$  raza dentro del ámbito geográfico de una zona, para medir por ejemplo la tasa de desempleo por sexo o edad específica en las zonas urbanas.

## Algunos métodos

- ▶ Los estimadores SAE se dividen en dos tipos principales dependiendo de cómo se aplican los modelos a los datos dentro de las áreas pequeñas: nivel de área y nivel de unidad.
- ▶ Los estimadores de área pequeña se basan en cálculos de nivel de área si los modelos vinculan la variable de interés y con variables auxiliares x específicas del área.

## Algunos métodos

- ▶ Se llaman modelos a nivel de unidad si se vinculan valores individuales para las variables auxiliares específicas de la unidad.
- ▶ Los estimadores basados en áreas pequeñas se calculan a nivel de área si los datos de la unidad no están disponibles.
- ▶ También pueden ser calculados si los datos de nivel de unidad están disponibles resumiéndolos en el nivel de área apropiado.

# Proceso de estimación

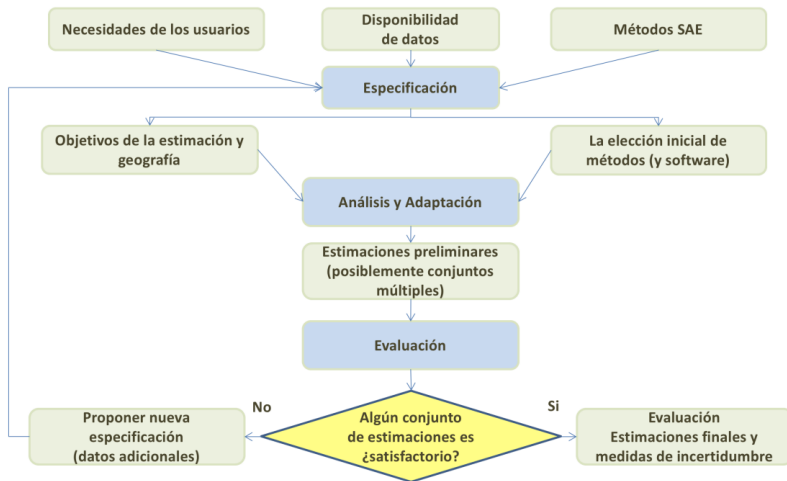


Figure 3: Producción de estadísticas con SAE

## Consideraciones

- ▶ Todos los métodos SAE requieren datos auxiliares a nivel del área pequeña desde el cual toman prestada la fuerza.
- ▶ La efectividad de los métodos SAE depende del grado de asociación entre la variable de interés y los datos auxiliares.
- ▶ La búsqueda de buenas variables auxiliares es crítica, incluida la construcción imaginativa de tales variables.
- ▶ Los datos auxiliares deben medirse de manera consistente a través de las áreas pequeñas, pero pueden incluir estimaciones de muestras grandes con error de muestreo conocido.



# Desafíos

- ▶ Aumento de las tasas de no respuesta.
- ▶ Aumento de costos, menos financiación.
- ▶ Aumento de la demanda de estimaciones para dominios pequeños como por raza, etnia o pobreza.
- ▶ Aumento de la demanda de estimaciones de áreas pequeñas.
- ▶ Aumento de la complejidad en los contenidos de los cuestionarios y por lo tanto la carga de respuesta.
- ▶ Aumento de la demanda de análisis secundarios, uso público y archivos de datos de uso restringido.

¡Gracias!

*Email:* andres.gutierrez@cepal.org