

# Marco de muestreo

Andrés Gutiérrez

Comisión Económica para América Latina y el Caribe (CEPAL) - [andres.gutierrez@cepal.org](mailto:andres.gutierrez@cepal.org)

# Tabla de contenidos I

Marco de muestreo

Los censos y su incidencia en los marcos de muestreo

Construcción de Unidades Primarias de Muestreo (UPM)

Marco de muestreo

## Definición del marco de muestreo

Todo procedimiento de muestreo probabilístico requiere de un dispositivo que permita identificar y ubicar a todos y cada uno de las unidades pertenecientes a la población objetivo, las cuales posteriormente participarán en el proceso de selección aleatoria que definirá la muestra. Este dispositivo se conoce con el nombre de **marco de muestreo**.

# Conceptos fundamentales

**Diseño de Muestreo de Elementos:** Este enfoque se utiliza cuando se dispone de un marco de elementos (por ejemplo, una lista completa de individuos o unidades). Se seleccionan directamente los elementos individuales para formar la muestra. Es útil cuando el marco es completo y accesible.

**Marcos de Conglomerados:** Cuando no se dispone de un marco de elementos o su construcción es costosa, se utilizan marcos de conglomerados. Por ejemplo, al realizar una encuesta sobre personas que viven en una ciudad, se puede acceder a la división cartográfica de la ciudad (comunas, localidades o barrios) como conglomerados naturales. Luego, se seleccionan personas dentro de estos conglomerados.

# Tipos de Marcos de Muestreo

**Marcos de Lista:** Listados físicos o magnéticos, ficheros o archivos que identifican objetos participantes en el sorteo aleatorio.

- ▶ *Ejemplo:* El Instituto Nacional de Estadística y Censos (INEC) de Costa Rica realiza la Encuesta Nacional de Microempresas de los Hogares con base en la muestra de la Encuesta Nacional de Hogares (primera fase), en donde se identifican las actividades económicas de los respondientes y se enlistan los trabajadores autónomos.

# Tipos de Marcos de Muestreo

**Marcos de Lista:** Listados físicos o magnéticos, ficheros o archivos que identifican objetos participantes en el sorteo aleatorio.

- ▶ *Ejemplo:* El Instituto Nacional de Estadística y Censos (INEC) de Costa Rica realiza la Encuesta Nacional de Microempresas de los Hogares con base en la muestra de la Encuesta Nacional de Hogares (primera fase), en donde se identifican las actividades económicas de los respondientes y se enlistan los trabajadores autónomos.

**Marcos de Área:** Mapas físicos o magnéticos, fotografías aéreas, imágenes de satélite que delimitan regiones o unidades geográficas.

- ▶ *Aplicación Significativa:* Utilizados en encuestas de hogares para seleccionar áreas geográficas antes de elegir unidades individuales.

# Importancia de la Información Auxiliar:

1. *Virtud del Marco*: Contiene información auxiliar que facilita la aplicación de diseños muestrales y mejora la eficiencia.
2. *Tipos de Información Auxiliar*:
  - ▶ *Discreta*: Permite la desagregación en categorías como nivel socioeconómico, región, departamento, etc.
  - ▶ *Continua*: Características de interés continuas y positivas, mejorando la eficiencia del muestreo.



# Defectos Comunes en Marcos de Muestreo

## 1. *Sobre-cobertura:*

- ▶ Aparición de objetos no pertenecientes a la población objetivo.
- ▶ “No son todos los que están.”

## 2. *Sub-cobertura:*

- ▶ Ausencia de elementos de la población objetivo o falta de actualización.
- ▶ “No están todos los que son.”

## 3. *Duplicación:*

- ▶ Presencia de múltiples registros para un mismo objeto.
- ▶ Originado por la construcción descuidada del marco a partir de registros administrativos de múltiples fuentes.

## Marco de muestreo

El marco de muestreo es cualquier dispositivo o mecanismo usado para obtener acceso observacional a la población de interés, para identificar y seleccionar una muestra, de manera que respete el esquema de muestreo probabilístico y para establecer contacto con los elementos seleccionados, de manera presencial, por correo postal, por correo electrónico, o mediante procedimientos automatizados como los sistemas de captura CAPI (*Computer Assisted Personal Interviewing*) o CATI (*Computer Assisted Telephone Interviewing*)

## Características Deseables de un Marco de Muestreo y su Importancia:

- ▶ Que las unidades en el marco son identificadas con un serial.
- ▶ Que cualquier unidad puede ser ubicada (dirección, teléfono).
- ▶ Que se pueda ordenar sistemáticamente (geografía, tamaño).
- ▶ Que contenga información adicional para cada unidad.
- ▶ Que especifique el dominio geográfico o socioeconómico al cual pertenece cada unidad.
- ▶ Que cada elemento de la población está presente sólo una vez.
- ▶ Que no contenga elementos que no estén en la población.
- ▶ Que todos los elementos de la población de interés estén en el marco muestral.

**La calidad del marco puede ser medida mediante la relación que existe entre la población objetivo y la población del marco**

Los censos y su incidencia en los marcos de muestreo

# Esfuerzos Significativos en Actualización

- ▶ Diseños de encuestas de hogares requieren múltiples etapas de selección basadas en marcos de muestreo, comúnmente marcos de área obtenidos de la división geográfica del país.
- ▶ Institutos de estadística en América Latina dedican grandes esfuerzos para mantener actualizados sus marcos de muestreo.
- ▶ **Ejemplo:**

La **Encuesta Nacional de Hogares** en Costa Rica utiliza un marco muestral derivado de los censos nacionales de población y vivienda de 2011. En este marco, compuesto por áreas geográficas, se definen las Unidades Primarias de Muestreo (UPM) con 150 viviendas en zonas urbanas y 100 viviendas en zonas rurales.

## Desafíos en la Obtención de Marcos de Lista

- ▶ Gambino y Silva (2009) mencionan que, en la practica, obtener marcos de lista de hogares en la última etapa del muestreo puede ser desafiante, especialmente en áreas donde no es evidente observar exhaustivamente los hogares dentro del conglomerado.
- ▶ Por ejemplo, en la ruralidad, no solamente los caminos existentes sirven para delimitar los conglomerados, sino que también los accidentes topográficos y las señales naturales se utilizan para este fin.
- ▶ La delimitación en la ruralidad se torna compleja cuando han ocurrido cambios en la infraestructura del área y aparecen nuevas construcciones.

# Interés en Desagregaciones Geográficas Amplias

- ▶ En los estudios de un fenómeno social, las desagregaciones geográficas más amplias constituyen un interés natural para los usuarios de las encuestas.
- ▶ Las estadísticas nacionales que se publican a partir de las encuestas de hogares cobran mayor relevancia a nivel de regiones, estados o departamentos. Este tipo de desagregaciones se conocen con el nombre de *dominios de representación*, que a su vez son agregaciones de los *estratos de muestreo*.
- ▶ Estratos de inclusión forzosa se recomiendan para características sesgadas, como ingresos o gastos, reduciendo el error de muestreo.

# Ciclo Decenal de Censos

- ▶ Siguiendo las recomendaciones internacionales, los países realizan censos cada diez años, formando lo que se conoce como el periodo intercensal.
- ▶ Durante el censo, se registran todos los hogares del país y se enumeran a todos sus habitantes. Además, se observan algunas variables de interés que servirán como base de comparación para las cifras en la próxima década.
- ▶ Entre censos, se llevan a cabo encuestas de hogares para recopilar información económica y social.



## Particiones en el Censo

- ▶ Institutos Nacionales de Estadística (INE) emplean particiones geográficas y cartográficas del censo para seleccionar muestras de hogares.
- ▶ Estas particiones, denominadas **Unidades Primarias de Muestreo (UPM)**, pueden ser manzanas, agregaciones de manzanas, veredas o sectores censales.
- ▶ Algunos países utilizan información censal para estratificar socioeconómicamente los segmentos cartográficos del marco de muestreo.
- ▶ La estratificación aprovecha parámetros de interés con comportamientos estructurales distintos en subgrupos poblacionales, mejorando la precisión de las estimaciones.

# Ventajas de la Estratificación

- ▶ Permite crear índices de condiciones de vivienda y bienestar a partir del censo.
- ▶ Facilita la definición de grupos de viviendas mutuamente excluyentes, generando estratos cartográficos en todo el país.
- ▶ Mejora la precisión de indicadores sociales (desocupación, pobreza, ingreso medio, etc.).

## Ejemplo de estratificación

- ▶ La *Gran Encuesta Integrada de Hogares* en Colombia, realiza una estratificación en dos grupos, incluyendo capitales y áreas metropolitanas, para ello considera criterios económicos a nivel municipal, como nivel de urbanización y estructura de población basada en proporción de habitantes con necesidades básicas insatisfechas.
- ▶ El Instituto Nacional de Estadística y Geografía de México incorpora estratificación en su diseño de muestra maestra, basándose en indicadores generados con información del Censo de Población y Vivienda 2010. Para realizar la estratificación, fue necesario construir y seleccionar una serie de variables que lograran, en conjunto, separar el universo de UPM en agrupaciones que mejoraran las principales estimaciones de las diferentes encuestas usuarias del marco de muestreo (INEGI 2012)

## Desafíos con UPM de Tamaños Desiguales:

- ▶ Las UPM no tienen el mismo tamaño dentro de los estratos y pueden mostrar heterogeneidad y dispersión geográfica.
- ▶ En áreas rurales, una UPM puede contener un conjunto de viviendas con mucha variabilidad, desde pocas hasta muchas viviendas.
- ▶ La variabilidad en el tamaño de las UPM crea desafíos técnicos en la recolección de información y la estimación de errores de muestreo.

## Problemas Técnicos, Recomendaciones y Estrategias de mejora:

- ▶ La selección de UPM con tamaños desiguales puede resultar en problemas técnicos, como la variabilidad en el tamaño de la muestra final.
- ▶ La incertidumbre en el costo operativo aumenta si se seleccionan UPM con pocas viviendas, ya que se necesitará un proceso adicional para cumplir con la cuota de viviendas.
- ▶ Recomendación de actualizar la cartografía y los marcos de muestreo al menos cada diez años, siguiendo las recomendaciones internacionales.

### **Estrategias de Mejora:**

- ▶ Se sugiere definir UPM con un número uniforme de viviendas para mejorar la fluidez en los procesos logísticos y aumentar la precisión de las estimaciones.
- ▶ La creación de UPM con un mismo número de viviendas, especialmente si el equipo de planeación tiene flexibilidad, es considerada una práctica conveniente.

## Construcción de Unidades Primarias de Muestreo (UPM)

# Importancia del Tamaño de las UPM:

- ▶ La definición del marco de muestreo para las encuestas de hogares responde básicamente a un objetivo: la definición de las unidades primarias de muestreo.

- ▶ *¿Cuál debe ser el tamaño apropiado para las UPM?*

El tamaño adecuado de las UPM es esencial y debe abordarse desde diversas perspectivas, como muestreo, logística, presupuesto y cartografía.

- ▶ Por ejemplo, Valliant, Dever, y Kreuter (2013) mencionan un caso en el que, al variar las definiciones del tamaño de las UPM, se observan pérdidas o ganancias significativas de eficiencia en los estimadores de las encuestas de hogares.

# Principios para la Construcción de UPM

1. **Exclusión de Estructuras no Habitadas:** Las UPM deben excluir estructuras que no contengan hogares particulares ocupados.
2. **Contenencia Espacial:** Las nuevas UPM deben estar contenidas en un solo municipio y diferenciarse claramente entre áreas urbanas y rurales.
3. **Actualización Continua:** La actualización de viviendas con hogares ocupados en las UPM seleccionadas es necesaria para calcular probabilidades de inclusión de segunda etapa y factores de expansión.



# Principios para la Construcción de UPM

1. **Exclusión de Estructuras no Habitadas:** Las UPM deben excluir estructuras que no contengan hogares particulares ocupados.
2. **Contenencia Espacial:** Las nuevas UPM deben estar contenidas en un solo municipio y diferenciarse claramente entre áreas urbanas y rurales.
3. **Actualización Continua:** La actualización de viviendas con hogares ocupados en las UPM seleccionadas es necesaria para calcular probabilidades de inclusión de segunda etapa y factores de expansión.

**Tamaño Recomendado de las UPM:** En América Latina, el tamaño de las UPM suele estar en el rango de 75 a 225 viviendas.

## Coeficiente de Correlación Intraclass (ICC)

En Cochran (1977) y Gutiérrez (2016), en la construcción de las UPM, el parámetro predominante que se debe considerar es ICC, que para la realización de encuestas con selección en múltiples etapas puede ser aproximado mediante la siguiente expresión (Valliant, Dever, y Kreuter 2013)

$$ICC = \frac{SCE}{SCE + SCD}$$

En donde  $SCE$  es suma de cuadrados relativa de los totales de la característica de interés entre las UPM y  $SCD$  es la suma de cuadrados relativa de los totales de la característica de interés dentro las UPM

# Evaluación de Idoneidad de las UPM:

## **Relación entre Tamaño de UPM e ICC**

1. ICC Cercano a Cero: Se necesitan pocas UPM para inferencias precisas debido a la heterogeneidad interna y homogeneidad entre UPM.
2. ICC Cercano a Uno: Se requiere una muestra grande de UPM para inferencias precisas debido a la homogeneidad interna y heterogeneidad entre UPM.

## **Control de Tamaño de UPM**

1. Se recomienda combinar secciones pequeñas geográficamente o bloques cercanos para que todas las UPM tengan al menos un número mínimo de personas.
2. Consideración de límites geográficos, tamaño y extensión para uniformidad en cargas de trabajo y óptima precisión estadística.

## Desafíos en la contrucción de las UPM:

- ▶ **Actualización de Marcos de Muestreo:** Los ciclos censales implican la actualización de los marcos de muestreo utilizados en las encuestas de hogares. Esto garantiza que la información primaria se recolecte de manera precisa y actualizada.
- ▶ **Uniformidad en el Tamaño de UPMs:** Es importante evitar que las Unidades Primarias de Muestreo (UPMs) tengan tamaños desiguales dentro de los estratos. Por ejemplo, en áreas rurales, una sola UPM puede agrupar viviendas muy heterogéneas, lo que afecta la eficiencia del diseño de muestreo.
- ▶ **Consecuencias de la Distribución Desigual de Viviendas en UPMs:** La distribución desigual de viviendas en las UPMs tiene efectos negativos, como estimaciones de varianza más amplias y cifras oficiales menos precisas. Para lograr un umbral de error de muestreo aceptable, se requiere un tamaño de muestra más grande.

¡Gracias!

*Email:* [andres.gutierrez@cepal.org](mailto:andres.gutierrez@cepal.org)

# Referencias

- Cochran, W. G. 1977. *Sampling Techniques*. Third Edition. Wiley.
- Gambino, J. G., y PL. d N. Silva. 2009. «Chapter 16 - Sampling and Estimation in Household Surveys». En *Handbook of Statistics*, 29:407-39. Handbook of Statistics. Elsevier. [https://doi.org/10.1016/S0169-7161\(08\)00016-3](https://doi.org/10.1016/S0169-7161(08)00016-3).
- Gutiérrez, H. A. 2016. *Estrategias de muestreo: diseño de encuestas y estimación de parámetros*. Segunda edición. Ediciones de la U.
- INEGI. 2012. «Metodología de la Construcción del Marco Maestro de Muestreo 2012 y del Diseño de la Muestra Maestra 2012».
- Valliant, Richard, Jill A. Dever, y Frauke Kreuter. 2013. *Practical Tools for Designing and Weighting Survey Samples*. Springer New York.  
<https://doi.org/10.1007/978-1-4614-6449-5>.