

Diseño y mecanismo de selección de la muestra

Andrés Gutiérrez

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

Tabla de contenidos I

Diseño y mecanismo de selección de la muestra

Diseños de muestreo

El diseño de muestreo estándar en una encuesta de hogares

Coordinación de muestras en encuestas de hogares

Diseño y mecanismo de selección de la muestra

Introducción

- ▶ La muestra será un vehículo adecuado para representar las características más importantes de la población en estudio, en la forma en que justamente las variables se incorporan en el formulario de la encuesta.
- ▶ Gutiérrez (2016) afirma que el muestreo es un procedimiento que responde a la necesidad de información estadística precisa sobre la población y los conjuntos de elementos que la conforman.
- ▶ Las encuestas de hogares son herramientas fundamentales para el seguimiento del desarrollo social y económico.
- ▶ El muestreo probabilístico permite obtener una muestra representativa de la población objetivo.

Principios del muestreo probabilístico:

- ▶ **Aleatorización:** las unidades incluidas en la muestra son seleccionadas mediante un proceso probabilístico.
- ▶ **Inclusión:** todas las unidades de la población tienen una probabilidad no nula de ser incluidas en la muestra.

Observaciones

- ▶ Las encuestas han sido fundamentales en la evolución de las mediciones de indicadores sociales.
- ▶ Permiten el seguimiento y monitoreo de cifras importantes para la sociedad.
- ▶ Facilitan la investigación sobre la efectividad de políticas públicas y la mejora de condiciones sociales y económicas.
- ▶ Una muestra bien seleccionada puede representar con precisión a una población mucho más grande.

Muestra Representativa:

- ▶ Una muestra representativa se considera un modelo reducido de la población.
- ▶ La validez de la muestra radica en que las categorías aparezcan con las mismas proporciones que en la población.
- ▶ En algunos casos, es importante sobrerrepresentar ciertas categorías o seleccionar unidades con probabilidades desiguales.
- ▶ La muestra no debe ser simplemente un modelo reducido; debe ser una herramienta para obtener estimaciones válidas: exactas, confiables, precisas y consistentes.

Diseños de muestreo

Diseños de muestreo

- ▶ Un diseño de muestreo es un procedimiento para seleccionar una muestra de una población objetivo.
- ▶ El procedimiento de muestreo le asigna una probabilidad de selección conocida a cada posible muestra.
- ▶ La asignación de probabilidades se realiza de manera teórica, la pericia del equipo técnico deberá establecer cuál es la mejor forma de selección, y sobre esta escoger el mejor algoritmo de muestreo.
- ▶ Dada las probabilidades, una única muestra es escogida mediante un mecanismo aleatorio que siga a cabalidad esta configuración estocástica inducida por el diseño de muestreo.

Diseños de muestreo

- ▶ Las probabilidades deben ser distintas de cero puesto que, de lo contrario, no se podría garantizar una inferencia insesgada.
- ▶ Estas mismas probabilidades se utilizan para crear los factores de expansión que definen todo el proceso de estimación, junto con el cálculo de los errores de muestreo.

Diferencia entre diseño de muestreo y algoritmo de muestreo

- ▶ El diseño de muestreo indica qué probabilidad de selección tendrán las posibles muestras en el soporte de muestreo, definido como el conjunto de todas las posibles muestras.
- ▶ **Algoritmo de muestreo:** se define como el proceso de selección de una única muestra que respeta las probabilidades del diseño de muestreo.
- ▶ Por ejemplo, si se ha decidido que el diseño de muestreo sea en etapas, el equipo técnico deberá documentar exhaustivamente cada etapa de muestreo. Luego, es igual de importante explicar qué algoritmos de selección serán utilizados en cada etapa de muestreo.

Probabilidades de inclusión en diseños de muestreo para encuestas de hogares

- Cada esquema particular de muestreo se define una única función que asocia a cada hogar k con una probabilidad de inclusión en la muestra s , definida de la siguiente manera:

$$\pi_k = Pr(k \in s)$$

Si el diseño de muestreo es de tamaño fijo, estas probabilidades de inclusión de los hogares cumplirán con las siguientes propiedades

1. $\pi_k > 0$
2. $\sum_U \pi_k = n$

Observaciones

- ▶ La primera propiedad garantiza que ningún hogar será excluido de la selección inicial.
- ▶ En segunda medida, el tamaño de la muestra de hogares estará inducido por la magnitud de las probabilidades de inclusión. Por esta razón, una encuesta con un tamaño de muestra grande asignará una mayor probabilidad de inclusión a todos los hogares, que una encuesta de tamaño de muestra más modesto.

Muestreo aleatorio simple

El Muestreo aleatorio simple supone que es posible realizar una enumeración de todas las posibles muestras de tamaño fijo y escoger una de ellas mediante una selección aleatoria que asigne la misma probabilidad a cada una.

- ▶ Es necesario tener información suficiente y exhaustiva de la ubicación e identificación de todas las unidades de interés.
- ▶ Su uso es común en las etapas finales de selección de las encuestas, en donde los hogares o personas se seleccionan con la misma probabilidad.
- ▶ Las probabilidades de inclusión en el muestreo aleatorio simple sin reemplazo son todas iguales y están dadas por la siguiente expresión:

$$\pi_k = Pr(k \in s) = \frac{\binom{1}{1} \binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$$

Muestreo aleatorio simple

Cuando se usa el estimador de Horvitz-Thompson en este diseño de muestreo para estimar un total poblacional, y suponiendo que $S_{y_U}^2$ denota la varianza de la característica de interés en la población finita, entonces las expresiones del estimador puntual y su varianza, respectivamente, toman la siguiente forma:

$$\hat{t}_{y,\pi} = \sum_s \frac{y_k}{\pi_k}$$
$$Var(\hat{t}_{y,\pi}) = \frac{N^2}{n} \left(1 - \frac{n}{N}\right) S_{y_U}^2$$

Muestreo sistemático:

- ▶ Es una variante del MAS.
- ▶ Se ordena el marco muestral con un patrón predefinido.
- ▶ Se selecciona un primer hogar al azar.
- ▶ Se seleccionan los demás hogares con un intervalo de salto constante, $a = \lfloor N/n \rfloor$.
- ▶ Probabilidades de inclusión en el muestreo sistemático

$$\pi_k = Pr(k \in s) = \frac{1}{a} \approx \frac{n}{N}$$

Muestreo sistemático:

Una muestra sistemática podría ser:

$$s = \{2, 12, 22, 32, 42\}.$$

En donde el primer hogar elegido en la UPM fue el segundo y con saltos sistemáticos de diez hogares se va encuestando los restantes hogares en la lista.

Muestreo proporcional al tamaño

El muestreo proporcional al tamaño (PPS) es un diseño de muestreo en el que las unidades de la población tienen probabilidades de selección proporcionales a una medida de tamaño.

- ▶ Este tipo de muestreo utiliza como insumo una característica de información auxiliar cuantitativa, también conocida como medida de tamaño (en inglés conocida como *measure of size*).
- ▶ El marco de muestreo deberá contener el valor correspondiente a la medida de tamaño para cada una de sus unidades.
- ▶ Este muestreo es utilizado con frecuencia en las etapas iniciales de selección de las muestras, particularmente en la selección de las UPM que harán parte de la muestra.

Muestreo proporcional al tamaño

- ▶ Las probabilidades de inclusión en la muestra para las UPM serán desiguales y proporcionales a la medida de tamaño.
- ▶ Es que hace más eficiente la estimación de los indicadores de interés.
- ▶ Es útil cuando la medida de tamaño está correlacionada con la variable de interés.
- ▶ La medida de tamaño no necesariamente tiene que estar definida como el conteo simple de hogares o personas dentro de las UPM, también puede definirse como una función de estos conteos; por ejemplo, la raíz cuadrada, o incluso como una función compuesta de conteos de subpoblaciones

Muestreo proporcional al tamaño

Sea N_i la medida de tamaño de la i -ésima UPM U_i , n_I el número de UPM que serán seleccionadas en cada estrato y N la sumatoria (o total) del número de hogares en todas las UPM del estrato (es decir, el número de hogares en el estrato) se tiene que las probabilidades de inclusión a la muestra s_I están dadas por la siguiente expresión:

$$\pi_i = Pr(U_i \in s_I) = n_I * \frac{N_i}{N}$$

Muestreo estratificado

El muestreo estratificado es un diseño de muestreo en el que la población se divide en grupos homogéneos (estratos) y se selecciona una muestra de cada estrato.

- ▶ **Inferencias precisas en subgrupos:** El muestreo estratificado permite realizar estimaciones precisas en subgrupos específicos de la población. Estos subgrupos suelen estar definidos por características geográficas, como regiones de un país.
- ▶ **Flexibilidad en el diseño:** Los diseños estratificados son flexibles y permiten adaptar las estrategias de muestreo según las características de interés. Por ejemplo, en encuestas sobre el uso del tiempo, se pueden definir estratos según la cantidad de horas dedicadas a actividades no remuneradas por hombres y mujeres, considerando diferencias entre zonas rurales y urbanas.

Muestreo estratificado

- ▶ **Independencia entre estratos:** Cada estrato tiene su propio marco de muestreo independiente y mutuamente excluyente. Esto facilita la ejecución de diferentes estrategias de muestreo dentro de cada estrato.
- ▶ **Ejemplo común en América Latina:** En países de América Latina, los estratos suelen formarse cruzando áreas geográficas grandes con divisiones socioeconómicas. Además, la división territorial entre urbano y rural es común en investigaciones sociales.

Las probabilidades de inclusión en el muestreo estratificado

Las probabilidades de inclusión definidas por este diseño de muestreo variarán en función de cada estrato h ($h = 1, \dots, H$). Por ejemplo, si se hubiese planeado un diseño aleatorio simple en cada estrato, entonces las probabilidades de inclusión estarían dadas por la siguiente expresión

$$\pi_k = Pr(k \in s_h) = \frac{n_h}{N_h}$$

En donde s_h define la muestra seleccionada en el estrato h , N_h sería el número de hogares en ese estrato y n_h el tamaño de la muestra de hogares asociado a ese estrato.

Muestreo de conglomerados

El muestreo de conglomerados es un diseño que surge como alternativa cuando no es posible generar una muestra directa de hogares a partir de un marco de muestreo que enumere cada uno de los hogares en un país. En lugar de intentar una muestra aleatoria simple, que sería costosa, se aprovecha el principio de aglomeración de las poblaciones humanas (que forman hogares y se aglomeran en segmentos, ciudades, regiones, etc.).

Muestreo de conglomerados

- ▶ **Unidades homogéneas:** Se forman unidades homogéneas (conglomerados) de población, como segmentos, ciudades o regiones. Estos conglomerados se seleccionan como muestras, y luego se realiza un censo de hogares dentro de cada uno.
- ▶ **Ventajas económicas:** Este enfoque limita el operativo de campo a un número manejable de conglomerados, lo que es más eficiente en términos presupuestarios.

Muestreo de conglomerados

- ▶ **Inconvenientes estadísticos:** A pesar de su conveniencia logística, este diseño tiene mayores errores de muestreo que un diseño simple. La variación interna dentro de los conglomerados suele ser baja, mientras que la variación entre conglomerados es alta, lo que afecta la inferencia de la encuesta.
- ▶ **Aumento del tamaño de la muestra:** Para mejorar la eficiencia, se podría aumentar el tamaño de la muestra de conglomerados, sin embargo, este aumento puede llegar a ser tan grande que, en algunos estratos, se deberían seleccionar todas las UPM

Muestreo en varias etapas

El **muestreo en varias etapas** es un enfoque que combina los principios del muestreo de conglomerados con submuestreos dentro de los conglomerados o UPM.

- ▶ **Esquemas en dos etapas:** En América Latina, es común utilizar esquemas de selección en dos etapas. En la primera etapa, se selecciona una muestra de UPM (conglomerados). En la segunda etapa, se selecciona una muestra de hogares dentro de las UPM seleccionadas en la primera etapa.
- ▶ **Esquemas en más de dos etapas:** También es posible encontrar esquemas en más de dos etapas. Por ejemplo, se seleccionan municipios en la primera etapa, luego UPM dentro de los municipios seleccionados en la segunda etapa, y finalmente se selecciona una muestra de hogares en las UPM pertenecientes a los municipios seleccionados.

Muestreo en varias etapas

- ▶ **Proceso jerárquico y sistemático:** Si un municipio es incluido en la muestra, se puede realizar un proceso jerárquico y sistemático hasta llegar a la unidad de observación. Por ejemplo, en una ciudad seleccionada, se puede hacer un submuestreo de sus secciones cartográficas, luego seleccionar sectores cartográficos (contenidos en las secciones) y, finalmente, seleccionar hogares o personas.
- ▶ **Medida de tamaño proporcional:** Si el esquema de muestreo incluye la selección de municipios en la primera etapa, el diseño de muestreo debe ser proporcional a una medida de tamaño, como el número de habitantes de los municipios. Esto garantiza que las ciudades más importantes (con más habitantes) estén siempre representadas en el estudio.
- ▶ **Submuestreo de personas dentro del hogar:** En algunas encuestas, puede haber un submuestreo de personas dentro de cada hogar seleccionado.

Características

- ▶ **La independencia:** que implica que no hay ninguna correlación en el diseño de muestreo de las unidades primarias de muestreo. Esto quiere decir que en cada UPM se puede ejecutar con independencia cualquier estrategia de muestreo que se crea apropiada para seleccionar la submuestra de hogares.
- ▶ **La invarianza:** que implica que sin importar qué diseño de muestreo se ejecutó en la primera etapa para seleccionar las UPM, la segunda etapa de selección podrá ejecutarse de manera independiente de la primera etapa. Es decir, el submuestreo de los hogares es independiente del muestreo de las UPM.

Las probabilidades de inclusión (PPS-MAS)

Un esquema de selección usado en las encuestas de hogares de América Latina es el relacionado con los diseños auto-ponderados, lo cuales, en la primera etapa de muestreo se seleccionan n_I de N_I UPM con probabilidad proporcional al número de hogares N_i que la habitan; es decir:

$$\pi_i = Pr(U_i \in s_I) = n_I \frac{N_i}{N} \quad i = 1, 2, \dots, N_I.$$

Sea $n_0 = 10$ hogares por UPM, entonces en la segunda etapa, la probabilidad de que el k -ésimo hogar sea seleccionado en la submuestra s_i de la UPM U_i que fue seleccionada en la muestra de la primera etapa s_I , está dada por la siguiente expresión:

$$\pi_{k|i} = Pr(k \in s_i | U_i \in s_I) = \frac{n_0}{N_i}$$

Las probabilidades de inclusión (PPS-MAS)

En los esquemas auto-ponderados, la probabilidad de inclusión de los hogares es siempre la misma para todos los hogares.

$$\pi_k = \pi_{k|i} * \pi_i = \frac{n_0}{N_i} \frac{n_I * N_i}{N} = \frac{n_0 * n_I}{N} = \frac{n}{N}$$

Nótese que $n = n_0 * n_I$ corresponderá al número total de hogares que serán seleccionados.

Las probabilidades de inclusión (PPS-MAS)

Una particularidad de las encuestas de hogares es que, casi siempre, las personas y los hogares comparten las mismas probabilidades de inclusión.

$$\pi_k^{per} = Pr(persona \in hogar | hogar \in muestra) = 1$$

Por lo anterior, se tiene que la probabilidad de inclusión de las personas en la muestra es idéntica a la del hogar, puesto que

$$1 * \pi_{k|i} * \pi_i = 1 * \frac{n}{N} = \frac{n}{N}.$$

Muestreo en dos fases

El muestreo en dos fases es un diseño de muestreo que se utiliza cuando se requiere información auxiliar para la selección de la muestra.

- ▶ En la **primera fase**, se selecciona una muestra de tamaño grande, conocida como *muestra maestra* y se recopila información sobre variables auxiliares. Estas variables pueden ayudar a estratificar la muestra, recopilar información adicional o permitir comparaciones a lo largo del tiempo
- ▶ En la **segunda fase**, con la ayuda de la información obtenida en la primera fase, se selecciona una submuestra mediante un diseño de muestreo conveniente, mucho más eficiente y apropiado para estimar el fenómeno en estudio.

Ejemplo de Muestreo en dos fases

- ▶ Imagina que necesitas estimaciones precisas para diferentes subgrupos poblacionales.
- ▶ Si no tienes un marco de muestreo confiable o actualizado para un muestreo estratificado, puedes aplicar el esquema de muestreo en dos fases.
- ▶ Primero, seleccionas una muestra aleatoria simple de tamaño moderado.
- ▶ Luego, realizas un empadronamiento de los individuos en la muestra, preguntándoles sobre su membresía en los subgrupos de interés.
- ▶ En una segunda fase, con ayuda de la información recolectada en la primera fase, se realiza un diseño estratificado.

Ejemplo México:

El INEGI ha implementado un interesante diseño de muestreo en dos fases para llevar a cabo encuestas de hogares importantes, como:

- ▶ La Encuesta Nacional de Ocupación y Empleo.
- ▶ La Encuesta Nacional sobre la Confianza del Consumidor.
- ▶ La Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública
- ▶ La Encuesta Nacional de Gasto de los Hogares.

Ejemplo México:

1. En la *primera fase*, se crea una muestra maestra de viviendas. Esta muestra es de gran tamaño y representa una variedad de hogares.
 - ▶ El objetivo es mantener actualizada la información sobre las viviendas particulares dentro de esta muestra.
2. En la *segunda fase*, se utiliza la información recopilada en la primera fase.
 - ▶ Se seleccionan submuestras específicas para encuestas particulares.
 - ▶ El diseño de estas submuestras se basa en la cobertura, tamaño y distribución de las encuestas continuas y periódicas del INEGI.

Ejemplo Costa Rica:

En Costa Rica, la Encuesta Nacional de Microempresas de los Hogares (ENAMEH) sigue un diseño en dos fases para recopilar información sobre las personas que poseen micro o pequeños negocios donde se desarrolla alguna actividad económica por cuenta propia.

Ejemplo Costa Rica:

1. Primera Fase: Se inicia con la Encuesta Nacional de Hogares, que identifica aquellos hogares cuyos miembros están involucrados en actividades económicas relacionadas con emprendimientos y microempresas. A partir de este listado exhaustivo, se construye una muestra maestra.
2. Segunda Fase: Se seleccionan todas las personas al frente de estas microempresas. A estas personas se les aplica un cuestionario para obtener información detallada sobre sus características y actividades económicas.

Ejemplo Chile:

Se realiza el Estudio Nacional de la Discapacidad que asume un marco de muestreo reducido, en una primera fase, basado en la encuesta de hogares CASEN, en la cual se identifican los hogares que tienen miembros con alguna condición de discapacidad. En una segunda fase, se realiza una selección de hogares y mediante un cuestionario estructurado se indagan las características de las personas con esta condición.

Muestreo balanceado

- ▶ El método del cubo (Tillé 2006) permite la selección de una muestra aleatoria para la cual el inverso de las probabilidades de inclusión reproduce de forma exacta el total poblacional de las variables de balanceo.
- ▶ Gutiérrez (2016) afirma que este es un procedimiento general y riguroso que permite la extracción de muestras probabilísticas balanceadas y la posterior estimación de las cantidades de interés, enmarcados bajo métodos de inferencia basados en el diseño de muestreo.

Muestreo balanceado

El **método del cubo** es una técnica de muestreo que se compone de dos fases: la **fase de vuelo** y la **fase de aterrizaje**.

1. Fase de Vuelo:

- ▶ En esta fase, se deben redondear las **probabilidades de inclusión** a cero (0) o uno (1) para satisfacer las restricciones de forma exacta.

2. Fase de Aterrizaje:

- ▶ En la **segunda fase**, se maneja adecuadamente el redondeo utilizando **programación lineal**.
- ▶ Por ejemplo, se puede aplicar el **método simplex** sujeto a una función de costo relacionada con la **varianza del estimador**.

Muestreo balanceado

En Perú, la Encuesta Demográfica y de Salud Familiar utiliza este tipo de muestreo en la selección de las UPM. De esta manera, como variables de balanceo se podrían definir las siguientes:

- ▶ Una columna de unos para que exista balanceo en el número de UPM.
- ▶ El vector de probabilidades de inclusión iniciales.
- ▶ Total de personas por grupos de edad y sexo (a partir de la información de los censos de población y vivienda).

El diseño de muestreo estándar en una encuesta de hogares

El diseño de muestreo estándar en una encuesta de hogares

- ▶ La mayoría de encuestas de hogares son multipropósito, lo que significa que se recopilan datos sobre múltiples variables de interés.
- ▶ Para lograr la precisión requerida en todas las estimaciones, el tamaño de muestra puede ser más exigente.
- ▶ La definición de los dominios de representatividad debe estar alineada con los objetivos de la encuesta y las unidades de muestreo.
- ▶ Mantener la comparabilidad de cifras entre levantamientos periódicos es esencial.

El diseño de muestreo estándar en una encuesta de hogares

El diseño de muestreo de una encuesta de hogares es generalmente **probabilístico estratificado y bietápico**:

1. Se realiza una estratificación por zona: urbano/rural, por región, departamento o estado y por los estratos socioeconómicos.
2. De forma independiente, dentro de cada estrato se realiza un muestreo bietápico.
 - ▶ En la primera etapa, se seleccionan áreas cartográficas, conocidas como unidades primarias de muestreo (UPM) siguiendo un diseño de muestreo proporcional al número de viviendas, hogares o personas del conglomerado.
 - ▶ En la segunda etapa, se escoge aleatoriamente un número fijo de hogares dentro de cada UPM siguiendo un diseño de muestreo aleatorio simple.

Elementos en el diseños de muestreo de una encuesta de hogares

- ▶ Las encuestas de hogares estimarán el total de cada UPM t_i .
- ▶ Suponga que la población de hogares U se divide en N_I UPM llamados conglomerados y denotadas como $U_I = \{U_1, \dots, U_{N_I}\}$.
- ▶ N_I es el número total de UPM dentro del país.
- ▶ La i -ésima UPM U_i $i = 1, \dots, N_I$ contiene N_i hogares.

Proceso de selección en una encuesta de hogares

- ▶ Una muestra s_I de UPM es seleccionada de U_I de acuerdo a un diseño de muestreo $p_I(s_I)$. El tamaño de la muestra de UPM se denota como n_I .
- ▶ Para cada UPM U_i $i = 1, \dots, n_I$ en la muestra seleccionada s_I , se realiza de forma independiente un submuestreo de hogares, de tal forma que en cada UPM existirá una muestra s_i de hogares de acuerdo a un diseño de muestreo $p_i(s_i)$.

Proceso de selección en una encuesta de hogares

1. Primera etapa:

- ▶ Se selecciona una muestra de UPM.
- ▶ Se estratifican las UPM por variables predefinidas.
- ▶ La probabilidad de selección de una UPM se basa en el número de personas o viviendas (medida de tamaño).
- ▶ Se selecciona un número mayor de UPM en los estratos más grandes.

2. Segunda etapa:

- ▶ Se selecciona una muestra de hogares dentro de cada UPM.
- ▶ Todos los hogares tienen la misma probabilidad de ser seleccionados.
- ▶ Se requiere un listado exhaustivo de todos los hogares en las UPM seleccionadas.
- ▶ Se realiza un empadronamiento previo para actualizar el número de hogares e identificarlos.

Actualización del marco de muestreo:

- ▶ La población cambia con el tiempo (nacimientos, muertes, migraciones).
- ▶ Se deben crear y eliminar hogares.
- ▶ La migración interna y externa afecta la distribución de la población.
- ▶ El marco de muestreo se desactualiza con el tiempo.
- ▶ Se debe ajustar el peso de muestreo de las UPM según la evidencia de cambios en el número de hogares.

Coordinación de muestras en encuestas de hogares

Introducción

- ▶ Se debe realizar una planificación rigurosa al definir un nuevo marco de muestreo.
- ▶ Esta planificación debe determinar las UPM que se seleccionarán a lo largo del periodo intercensal.
- ▶ La planificación debe atender a los parámetros estadísticos de cada encuesta.

Tipos de coordinación

Coordinación Positiva y Negativa: Se establece que una muestra es *positivamente coordinada* con otra si comparten todos sus elementos. Por otro lado, dos muestras son *negativamente coordinadas* si no tienen elementos en común.

Ejemplo de Esquema Rotacional: En un esquema rotacional $2(2)2$, dos muestras de periodos consecutivos tendrán un traslape del 50% y estarán parcialmente coordinadas. Sin embargo, si dos muestras están distanciadas por dos periodos consecutivos, no tendrán ningún traslape y deberán estar negativamente coordinadas.

Selección Secuencial y Números Aleatorios

Para lograr el objetivo, se pueden utilizar esquemas de selección secuenciales que asignen números aleatorios a cada UPM en el marco de muestreo.

- *Números aleatorios permanentes*: a cada unidad $i \in U_I$ se le asignara el siguiente número:

$$\xi_i^P \sim Uniforme(0, 1)$$

Evidentemente, en este caso los números aleatorios permanentes no son equidistantes.

- *Números aleatorios colocados*: a partir de los números aleatorios ξ_i creados en el paso anterior, es posible utilizar su rango para definir su ordenamiento y mediante la siguiente función crear números aleatorios equidistantes:

$$\xi_i^C = \frac{\text{Rango}(\xi_i^P) - \varepsilon}{N}$$

En donde ε es un único valor aleatorio entre cero y uno.

Ejemplo

Tabla 1: Ejemplo reducido de la conformación de números aleatorios permanentes y colocados.

Unidad	ξ_P	ξ_C
1	0.2875	0.1717
2	0.7883	0.6717
3	0.4089	0.2717
4	0.8831	0.7717
5	0.9404	0.9717
6	0.0455	0.0717
7	0.5281	0.4717
8	0.8924	0.8717
9	0.5514	0.5717
10	0.4566	0.3717

Coordinación de muestras aleatorias simples

Selección de Muestra Aleatoria Simple:

Para elegir una muestra aleatoria simple s de tamaño n , se debe ordenar el marco de muestreo en orden ascendente según los números ξ_i^P . La muestra s estará compuesta por los primeros n registros del marco ordenado (o los últimos n registros).

Coordinación de Muestras

- ▶ Se escogen dos constantes a_1 y a_2 en el intervalo $(0, 1)$.
- ▶ La muestra s_1 como las primeras n_1 unidades a la derecha (o izquierda) de a_1 y la muestra s_2 como las primeras n_2 unidades a la derecha (o izquierda) de a_2 .
- ▶ Si se desea coordinar positivamente, se establece $a_1 = a_2$.
- ▶ Si se desea coordinar negativamente, sumar 0.5 (en módulo uno) a la constante a_1 para obtener $a_2 = (a_1 + 1/2) \bmod 1$.
- ▶ Si se quieren coordinar negativamente Q muestras diferentes, se aconseja añadir la cantidad de $1/Q$ (en módulo uno) a la constante a_1 .

Tabla de ejemplo

La tabla muestra la selección de dos muestras negativamente coordinadas de tamaño $n_1 = n_2 = 3$, con $a_1 = 0$ y $a_2 = 0.5$.

Tabla 2: Ejemplo de la selección de dos muestras aleatorias simples coordinadas negativamente.

Unidad	xi_P	s_1	s_2
6	0.0455	1	0
1	0.2875	1	0
3	0.4089	1	0
10	0.4566	0	0
7	0.5281	0	1
9	0.5514	0	1
2	0.7883	0	1
4	0.8831	0	0
8	0.8924	0	0
5	0.9404	0	0

Método de Poisson secuencial para muestras proporcionales

El método de Poisson secuencial (Ohlsson 1995), que define los siguientes números aleatorios permanentes para cada UPM:

$$\xi_i^{pps} = \frac{\xi_i^P}{N_I \times p_i}$$

En donde N_I es el número de UPM en el marco de muestreo y $p_i = Ni/N$ es la proporción de hogares en la i -ésima UPM.

Coordinación de muestras proporcionales

- ▶ Al ordenar el marco mediante los números ξ_i^{pps} y seleccionar los primeros elementos, se obtendrá una muestra secuencial de Poisson.
- ▶ En cuanto a la coordinación de muestras, es posible aplicar los mismos principios de la sección anterior.

Tabla de ejemplo

Tabla 3: Ejemplo de la selección de dos muestras secuenciales de Poisson coordinadas negativamente.

Unidad	xi_P	N_I	xi_pps	s_1	s_2
6	0.0455	198	0.0405	1	0
1	0.2875	173	0.2928	1	0
3	0.4089	184	0.3913	1	0
10	0.4566	179	0.4494	0	0
9	0.5514	195	0.4981	0	0
7	0.5281	155	0.6001	0	1
2	0.7883	162	0.8568	0	1
5	0.9404	190	0.8715	0	1
8	0.8924	166	0.9463	0	0
4	0.8831	159	0.9780	0	0

Algoritmo de Pareto (IBGE, Brasil)

Se calcula el número aleatorio permanente para cada UPM:

$$\xi_i^{par} = \frac{\xi_i^P / (1 - \xi_i^P)}{\pi_i / (1 - \pi_i)}$$

En donde $\pi_i = n_I * p_i$ es la probabilidad de inclusión de la i -ésima UPM y deberá garantizarse que sea menor que uno.

Al ordenar el marco mediante los números ξ_i^{par} y seleccionar los primeros elementos, se obtendrá una muestra secuencial de Poisson.

Tabla de ejemplo

Tabla 4: Ejemplo de la selección de dos muestras de Pareto coordinadas negativamente.

Unidad	ξ_P	ξ_{P_par}	s_1	s_2
6	0.0455	0.0937	1	0
1	0.2875	0.9662	1	0
3	0.4089	1.5148	1	0
10	0.4566	1.9165	0	1
9	0.5514	2.4720	0	1
7	0.5281	3.1199	0	1
2	0.7883	9.7679	0	0
4	0.8831	20.3317	0	0
8	0.8924	21.0230	0	0
5	0.9404	32.9658	0	0

¡Gracias!

Email: andres.gutierrez@cepal.org

Referencias

- Gutiérrez, H. A. 2016. *Estrategias de muestreo: diseño de encuestas y estimación de parámetros*. Segunda edición. Ediciones de la U.
- Ohlsson, E. 1995. «Coordination of Samples using Permanent Random Numbers». *Chapter 9 in Business Survey Methods*, 153-69.
- Tillé, Yves. 2006. *Sampling Algorithms*. Springer Series en Statistics. Springer-Verlag. <https://doi.org/10.1007/0-387-34240-0>.