Cálculo del tamaño de muestra

Andrés Gutiérrez

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

Tabla de contenidos I

Tabla de contenidos II

Cálculo del tamaño de muestra

Introducción

- El tamaño de muestra es un tema crucial en el diseño y análisis de encuestas de hogares.
- La selección de la muestra debe depender del propósito específico de la encuesta.
- No solo se refiere a la selección de hogares, sino también a la inclusión de personas en la muestra.
- En el caso de encuestas de propósitos múltiples, que abordan diversos temas anualmente, el tamaño de muestra debe ser útil, pertinente y apropiado para todos los indicadores que se desean medir simultáneamente.

Confiabilidad y Precisión

Definición de Intervalo de Confianza

- Se establece un intervalo de confianza para el parámetro θ, basado en su estimador insesgado $\hat{\theta}$, asumiendo una distribución normal de media θ y varianza $Var(\hat{\theta})$.
- La fórmula del intervalo de confianza es

$$IC(1-\alpha) = \left[\hat{\theta} - z_{1-\alpha/2}\sqrt{Var(\hat{\theta})}, \hat{\theta} + z_{1-\alpha/2}\sqrt{Var(\hat{\theta})}\right]$$

  donde $z_{1-\alpha/2}$ se refiere al cuantil $(1-\alpha/2)$ de una variable aleatoria con distribución normal estándar
- En diseños de muestreo complejo, se reemplaza el percentil de la distribución normal estándar por el percentil de una distribución $t - student$ con $N_I - H$ grados de libertad, donde hay N_I unidades primarias de muestreo y H estratos.

Margen de Error y Error Estándar

- Se define el *margen de error* (ME) como la cantidad agregada y sustraída al estimador insesgado.

$$ME = z_{1-\alpha/2}\sqrt{Var(\hat{\theta})}$$

- El *error estándar* (EE) se define como la raíz cuadrada de la varianza del estimador.

$$EE = \sqrt{Var(\hat{\theta})}$$

Margen de Error Relativo y Coeficiente de Variación

- Se introduce el *margen de error relativo* (MER), que considera la precisión y el sesgo del estimador.

$$MER = z_{1-\alpha/2} \frac{\sqrt{Var(\hat{\theta})}}{E(\hat{\theta})}$$

- Se define el *coeficiente de variación* (CV) o *error estándar relativo*, una medida que considera la precisión y el sesgo del estimador.

$$CV = \frac{\sqrt{Var(\hat{\theta})}}{E(\hat{\theta})}$$

Consideraciones sobre el Tamaño de Muestra:

- El tamaño de muestra necesario dependerá del tipo de error que se busca minimizar: margen de error, margen de error relativo, coeficiente de variación, etc.
- Se destaca que el tamaño de muestra requerido para minimizar el margen de error no será el mismo que para minimizar el coeficiente de variación, dependiendo de la población en consideración.

El efecto de diseño en la determinación del tamaño de muestra

Importancia del Tamaño de Muestra en Encuestas de Hogares

- Permite controlar el costo del estudio y asegurar la calidad estadística de los resultados desde la fase de diseño.
- La calidad se mide en términos de error muestral, con indicadores como margen de error, margen de error relativo y coeficiente de variación.

Paradigmas y Retos en el Tamaño de Muestra

- El supuesto de independencia entre observaciones, fundamental en la teoría estadística, no se cumple en encuestas de hogares debido a estratificación y aglomeración.
- Fórmulas clásicas basadas en la independencia de observaciones conducen a tamaños de muestra insuficientes.

Efecto de Diseño y Correlación

- El efecto de diseño (DEFF) se introduce como un factor de ajuste en función de la correlación entre la variable de interés y las unidades primarias de muestreo.
- DEFF se calcula para obtener una aproximación de la varianza bajo un diseño muestral complejo y se utiliza para la determinación del tamaño de muestra.
- La relación DEFF se utiliza para ajustar el tamaño de muestra necesario bajo un diseño complejo en comparación con un muestreo aleatorio simple.

Expresiones del Efecto de Diseño

- Cuando se selecciona una muestra de una población de interés utilizando un diseño de muestreo de conglomerados o en varias etapas, no se puede asumir que las observaciones son independientes.
- Las fórmulas clásicas para determinar el tamaño de muestra, basadas en el muestreo aleatorio simple, no sean aplicables. Sin embargo, podemos incorporar este efecto de aglomeración mediante la relación de las varianzas en el efecto de diseño:

$$DEFF(\hat{\theta}) = \frac{Var_p(\hat{\theta})}{Var_{MAS}(\hat{\theta})}$$

Expresiones del Efecto de Diseño

- Ahora, es posible escribir la varianza del estimador bajo el diseño de muestreo complejo como

$$Var_p(\hat{\theta}) = DEFF(\hat{\theta})\ Var_{MAS}(\hat{\theta}) \quad (1)$$

$$= DEFF(\hat{\theta})\ \frac{N^2}{n}\left(1 - \frac{n}{N}\right) S^2_{y_U} \quad (2)$$

- Si al implementar un muestreo aleatorio simple el tamaño de muestra n_0 es suficiente para conseguir la precisión deseada, entonces el valor del tamaño de muestra que tendrá en cuenta el efecto de aglomeración para un diseño complejo estará cercano a $n \approx n_0 \times DEFF$.

Recomendaciones para los Valores del DEFF

- DEFF indica cuánto más grande debe ser el tamaño de muestra para lograr la misma confiabilidad que una muestra aleatoria simple.
- UN (2008) afirma que, claramente indeseable tener un plan de muestreo con valores mucho mayores que 2.5 o 3.0 para los indicadores clave de la encuesta.

Cálculo del Tamaño de Muestra para una Proporción

La calidad del estimador se puede medir en términos de la amplitud del intervalo de confianza de al menos $(1-\alpha)\times 100\%$;esto es,

$$1-\alpha \geq \Pr\left(|\hat{P}-P| < ME\right)$$

Ahora, el estimador de Horvitz-Thompson de la proporción $\hat{P}$ es insesgado para P y su distribución asintótica es gausiana con varianza dada por

$$Var\left(\hat{P}\right) = DEFF\frac{1}{n}(1-\frac{n}{N})P(1-P)$$

Cálculo del Tamaño de Muestra para una Proporción

Al despejar el tamaño muestral n de la anterior expresión, se tiene que

$$n \geq \frac{P(1-P)}{\frac{ME^2}{DEFF\ z^2_{1-\alpha/2}} + \frac{P(1-P)}{N}}$$

Cálculo del Tamaño de Muestra para un Promedio

- Si el interés recae en la estimación de un promedio $\bar{y}_U$, el tamaño de muestra necesario para que la amplitud relativa del intervalo de confianza no supre un margen de error relativo previamente establecido (MER) es de

$$n \geq \frac{S^2_{y_U} DEFF}{\frac{MER^2 \bar{y}^2_U}{z^2_{1-\alpha/2}} + \frac{S^2_{y_U} DEFF}{N}}$$

- Valores grandes del efecto de diseño inducirán un mayor tamaño de muestra

Algunos Escenarios de Asignación del Tamaño de Muestra

Algunos Escenarios de Asignación del Tamaño de Muestra

- Las encuestas de hogares se basan en un marco de muestreo de áreas que abarca toda la población de un país.
- Este marco incluye Unidades Primarias de Muestreo (UPM), que contienen los hogares donde residen las personas sujetas a ser encuestadas.
- La agrupación natural de las personas en hogares complica los cálculos, especialmente considerando que la población de interés es un subconjunto de los habitantes de los hogares.
- Es común que el marco de muestreo consista en una lista de UPM, lo que implica calcular el tamaño de muestra tanto para las personas como para los hogares.

Notación

Suponemos una población U de N elementos sobre la que se desea seleccionar una muestra s de n elementos en los cuales se quiere medir una característica de interés.

- N es el tamaño de la población U.
- n es el tamaño de la muestra s.
- N_I es el número de UPM en el marco de muestreo.
- n_I es el número de UPM que se selecciona en la muestra de la primera etapa s_i.
- N_{II} es el número de hogares existentes en el país.
- n_{II} es el número de hogares seleccionados en la muestra de la segunda etapa s_{II}.

Notación

- $\bar{n}$ es el número promedio de personas que se van a seleccionar en cada UPM.
- $\bar{n}_{II}$ es el número promedio de hogares que se van a seleccionar en cada UPM.
- ρ_y es el coeficiente de correlación intraclase, calculado para la variable de interés sobre las UPM.
- b es el número promedio de personas por hogar.
- r es el porcentaje promedio de personas en el hogar susceptibles de ser observadas para la característica de interés.
- $z_{1-\alpha/2}$ es el percentil $(1-\alpha/2)$ asociado a una distribución normal estándar y a la confianza que se requiera en la inferencia.

Condicionantes Iniciales para el Diseño de Muestreo

En aras de mantener la uniformidad en los cálculos, consideremos los siguientes supuestos iniciales:

- **Tamaño de la Población (N)**: La población tiene un tamaño de **50 millones**.
- **Hogares en la Población (N_{II})**: Hay **12 millones** de hogares en la población.
- **Objetivo de la Muestra**: Se busca obtener una muestra con un nivel de confianza del **90%**.
- **Unidades Primarias de Muestreo (N_I)**: El país se divide en **30 mil UPM**, que a su vez están conformadas por segmentos cartográficos (agregaciones de manzanas).

Tamaño de muestra para UPM, hogares y personas

Tamaño de muestra para UPM, hogares y personas

- La unidad de observación en encuestas de hogares es frecuentemente la persona, incluso si la variable de interés está a nivel de hogar.
- Se destaca la importancia de basar los cálculos en el tamaño de muestra de las personas para una inferencia apropiada.
- Por ejemplo, para tener una inferencia apropiada al estimar el ingreso medio percápita es necesario definir a la población objetivo como todas las personas que componen un hogar.

Pasos para el Cálculo del Tamaño de Muestra

1. **Definición de Población de Interés**
   - Esencial especificar si la unidad de análisis son las personas o los hogares.
   - Parámetros clave: r (porcentaje de personas con la característica) y b (número promedio de personas por hogar).
   - Si la unidad de análisis son todas las personas del hogar, entonces $r = 1$, de otra forma $r < 1$.
   - b dependerá del dominio de representatividad en el que se requiera el cálculo.

Pasos para el Cálculo del Tamaño de Muestra

2. **Número Promedio de Hogares**
   - Determinar $\bar{n}_{II}$, el número promedio de hogares a encuestar en cada UPM.
   - Iterar este proceso para evaluar la calidad del diseño y seleccionar un valor óptimo.
3. **Número Promedio de Personas Encuestadas**

   Calcular $\bar{n}$, el número promedio de personas a encuestar, dependiendo de $\bar{n}_{II}$, r, y b de la siguiente manera:

$$\bar{n} = \bar{n}_{II} \times r \times b$$

Pasos para el Cálculo del Tamaño de Muestra

4. **Cálculo del Efecto de Diseño**
   - Definir la correlación intraclase ρ_y.
   - Calcular $DEFF$ como función de ρ_y y $\bar{n}$; esto es

$$DEFF \approx 1 + (\bar{n} - 1)\rho_y$$

Pasos para el Cálculo del Tamaño de Muestra

5. **Tamaño de Muestra de Personas**

   - Para estimar un promedio $\bar{y}_U$:

$$n \geq \frac{S^2_{y_U} DEFF}{\frac{MER^2 \bar{y}^2_U}{z^2_{1-\alpha/2}} + \frac{S^2_{y_U} DEFF}{N}}$$

   - Para estimar una proporción P:

$$n \geq \frac{P\,(1-P)\;DEFF}{\frac{MER^2 P^2}{z^2_{1-\alpha/2}} + \frac{P\,(1-P)\;DEFF}{N}}$$

Pasos para el Cálculo del Tamaño de Muestra

6. **Tamaño de Muestra de Hogares**

   Calcular n_{II}, el número total de hogares necesarios para entrevistar a las personas definidas en el paso anterior:

$$n_{II} = \frac{n}{r \times b}$$

7. **Número de UPM** Calcular n_I, el número de UPM necesarias:

$$n_I = \frac{n}{\bar{n}} = \frac{n_{II}}{\bar{n}_{II}}$$

Ejemplo de Tamaño de Muestra para Proporción de Personas Pobres

El parámetro de interés es el porcentaje de personas pobres, aquellos cuyos ingresos están por debajo de un umbral preestablecido. El objetivo es realizar inferencias con un margen de error relativo máximo del 5%. Según estudios previos en este país, se estima que la proporción de personas pobres es aproximadamente 4%. La población objetivo abarca a todos los habitantes del país ($r = 100\%$). Además, se ha estimado que el tamaño promedio del hogar es de alrededor de 3.5 personas. Por último, la correlación intraclase entre la característica de interés y las unidades primarias de muestreo se sitúa en **0.034**.

Estracción del Contexto

- Parámetro de interés: Porcentaje de personas pobres.
- Margen de error relativo máximo del 5%.
- Proporción estimada de personas pobres (P): 4%.
- Población objetivo: Todos los habitantes del país ($r = 100\%$).
- Tamaño promedio del hogar (b): 3.5 personas.
- Correlación intraclase con unidades primarias de muestreo (ρ_y): 0.034.

Resultados del Ejercicio

Hogares promedio por UPM $(\bar{n}_{II})$	Personas promedio por UPM $(\bar{n})$	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})	Muestra de personas (n)
10	35	2.2	1598	15982	55936

Escenarios Alternativos

Tabla 2: Tabla de muestreo para la estimación de proporción de personas pobres.

Hogares promedio por UPM ($\bar{n}_{II}$)	Personas promedio por UPM ($\bar{n}$)	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})	Muestra de personas (n)
5	18	1.6	2315	11575	40512
10	35	2.2	1598	15982	55936
15	52	2.8	1359	20386	71351
20	70	3.4	1239	24787	86756
25	88	3.9	1167	29186	102152
30	105	4.5	1119	33582	117538
35	122	5.1	1085	37976	132915
40	140	5.7	1059	42366	148282
45	158	6.3	1039	46754	163640

Ejemplo: ingreso promedio por persona

Suponga que se desea estimar el ingreso promedio por hogar con un margen de error relativo máximo del 2%. La variable de interés (ingreso) es continua y se estima que la media oscila alrededor de $\bar{y}_U = 1180$ dólares con una varianza de $S^2_{y_U} = 1845.94^2$. En este caso, la población objetivo son todos los habitantes del hogar, por lo cual $r = 100\%$. La composición del hogar se calcula en $b = 3.79$ personas por hogar. Por último, según levantamientos anteriores, la correlación intraclase de la característica de interés es $\rho_y = 0.035$. Nótese que la correlación intraclase cambia con respecto a la característica que se desee medir.

Contexto del Ejemplo

- Estimación del ingreso promedio por hogar.
- Margen de error relativo máximo del 2%.
- Media estimada del ingreso ($\bar{y}_U$): \$1180.
- Varianza estimada del ingreso ($S^2_{y_U}$): 1845.94^2.
- Población objetivo: Todos los habitantes del hogar ($r = 100\%$).
- Tamaño promedio del hogar (b): 3.79 personas.
- Correlación intraclase con unidades primarias de muestreo (ρ_y): 0.035.

Resultados del Ejercicio

Hogares promedio por UPM $(\bar{n}_{II})$	Personas promedio por UPM $(\bar{n})$	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})	Muestra de personas (n)
15	57	3	859	12892	48861

Escenarios Alternativos

Tabla 4: Tabla de muestreo para la estimación del ingreso promedio por persona en el ejemplo.

Hogares promedio por UPM ($\bar{n}_{II}$)	Personas promedio por UPM ($\bar{n}$)	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})	Muestra de personas (n)
5	19	1.6	1422	7108	26938
10	38	2.3	1000	10001	37902
15	57	3.0	859	12892	48861
20	76	3.6	789	15783	59816
25	95	4.3	747	18672	70766
30	114	4.9	719	21560	81711
50	190	7.6	662	33098	125443
100	379	14.2	619	61857	234439

Ejemplo: tasa de desocupación en adultos mayores

Suponga que la incidencia de la desocupación está alrededor de $P = 5.5\%$ en la población objetivo, que son las personas económicamente activas (PEA) mayores de 60 años; en este país, se ha estimado que en promedio hay $r = 4.6\%$ de adultos mayores por hogar que pertenecen a la PEA, cuyo tamaño promedio es de alrededor de $b = 5$ personas. Además, se quiere hacer inferencia con un margen de error relativo máximo del 15%. Por último, según levantamientos anteriores, la correlación intraclase de la característica de interés es $\rho_y = 0.7$.

Contexto del Ejemplo

- La población objetivo es la de personas económicamente activas (PEA) mayores de 60 años.
- La incidencia de la desocupación en esta población es del 5.5%.
- Se estima que, en promedio, hay un $r = 4.6\%$ de adultos mayores por hogar que pertenecen a la PEA, con un tamaño promedio de hogar de alrededor de $b = 5$ personas.
- Margen de error relativo máximo del 15%.
- Correlación intraclase con unidades primarias de muestreo (ρ_y): 0.7

Resultados del diseño de muestreo

Hogares promedio por UPM ($\bar{n}_{II}$)	Personas promedio por UPM ($\bar{n}$)	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})	Muestra de personas (n)
20	4.6	3.5	1581	31617	7272

Escenarios alternativos

Tabla 6: Tabla de muestreo para la estimación de la tasa de desocupación en adultos mayores.

Hogares promedio por UPM ($\bar{n}_{II}$)	Personas promedio por UPM ($\bar{n}$)	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})	Muestra de personas (n)
5	1.1	1.1	1985	9926	2283
10	2.3	1.9	1716	17157	3946
15	3.5	2.7	1626	24387	5609
20	4.6	3.5	1581	31617	7272
25	5.8	4.3	1554	38848	8935
30	6.9	5.1	1536	46074	10597
50	11.5	8.3	1500	74983	17246
100	23.0	16.4	1472	147222	33861

Tamaño de muestra para UPM y hogares

Tamaño de muestra para UPM y hogares

- En ciertos casos, la variable de interés y la unidad de observación están a nivel de hogar, como cuando todas las variables relevantes se miden a nivel de la vivienda o del hogar.
- Se puede ajustar el algoritmo previamente descrito para seleccionar solo viviendas u hogares en la muestra, sin necesidad de un submuestreo de personas.
- Algunas cantidades, como r y b, desaparecen ya que no son relevantes para la población de hogares.
- Expresiones como el coeficiente de correlación intraclase (ρ_y), el efecto de diseño y los tamaños de muestra deben ser redefinidos en el contexto de los hogares.

Pasos del nuevo algoritmo

- **Definir el número promedio de hogares:** El número promedio de hogares que se desea encuestar en cada una de las UPM está dado por $\bar{n}_{II}$
- **Calcular el efecto de diseño:** Se debe definir la correlación intraclase ρ_y de la variable de interés a nivel del hogar con el agrupamiento por Unidad Primaria de Muestreo (UPM).
- **Tamaño de muestra de hogares:** Se calcula a partir de expresiones adaptadas para muestreos complejos y el tamaño de muestra necesario para alcanzar un margen de error relativo máximo.

$$n_{II} \geq \frac{S_y^2 DEFF}{\frac{MER^2 \bar{y}^2}{z_{1-\alpha/2}^2} + \frac{S_{y_U}^2 DEFF}{N_{II}}}$$

Pasos del nuevo algoritmo

el tamaño de muestra para una proporción estará dada por

$$n_{II} \geq \frac{P\ (1-P)\ DEFF}{\frac{MER^2P^2}{z^2_{1-\alpha/2}} + \frac{P\ (1-P)\ DEFF}{N_{II}}}$$

- **Cálculo del número de UPM:** Se determina cuántas UPM deben ser seleccionadas en el muestreo.

$$n_I = \frac{n_{II}}{\bar{n}_{II}}$$

Recomendaciones

- Es importante adaptar el algoritmo según el contexto específico del estudio y las características de la variable de interés.

-Es necesario evaluar diferentes escenarios de muestreo y ajustar el tamaño de muestra en función de los objetivos y limitaciones del estudio.

Ejemplo: gasto promedio del hogar

Suponga que se desea estimar el promedio de gasto anual en dólares en los hogares del país con un margen de error relativo máximo admisible del 3.5%. La variable de interés (gasto) es continua y se estima que la media oscila alrededor de $\bar{y}_U = 1407$ dólares con una varianza de $S^2_{y_U} = 2228^2$. En este ejemplo se supone que el país está dividido en $N_I = 10$ mil UPM y la correlación intraclase de la característica de interés, medida a nivel del hogar, con las UPM es de $\rho_y = 0.173$.

Contexto del Ejemplo

- Estimar el promedio de gasto anual en dólares en los hogares del país con un margen de error relativo máximo admisible del 3.5%.
- La variable de interés es el gasto, una variable continua con una media estimada de $\bar{y}_U = 1407$ dólares y una varianza de $S^2_{y_U} = 2228^2$.
- El país está dividido en $N_I = 10$ mil Unidades Primarias de Muestreo (UPM).
- La correlación intraclase (ρ_y) de la variable de interés, medida a nivel del hogar con las UPM, es 0.173.

Resultados del ejercicio

Hogares promedio por UPM $(\bar{n}_{II})$	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})
12	2.9	1338	16056

Escenarios alternativos

Tabla 8: Tabla de muestreo para la estimación del gasto promedio del hogar en el ejemplo.

Hogares promedio por UPM $(\bar{n}_{II})$	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})
2	1.2	3246	6493
4	1.5	2102	8407
6	1.9	1720	10320
8	2.2	1529	12233
10	2.6	1414	14145
12	2.9	1338	16056
14	3.2	1283	17967
16	3.6	1242	19877
18	3.9	1210	21787
20	4.3	1185	23695

Recomendaciones

- Se desestiman escenarios con efectos de diseño mayores a 3, siguiendo las recomendaciones internacionales.
- Se subraya la flexibilidad del diseño de muestreo, adaptándolo según las necesidades logísticas y presupuestarias del estudio.

Ejemplo: proporción de hogares sin agua potable

Suponga que se desea obtener una muestra con un margen de error relativo máximo admisible del 10% sobre la variable de interés (necesidades básicas insatisfechas en agua) y el parámetro de interés es el porcentaje de hogares con esta carencia. En este país, se estima que la proporción de hogares con esta condición asciende a $P = 7.5\%$. En este ejemplo se supone que la correlación intraclase de la característica de interés con las UPM es de $\rho_y = 0.045$.

Contexto del Ejemplo

- Estimar la proporción de hogares sin agua potable con un margen de error relativo máximo admisible del 10%.
- Variable de interés: Necesidades básicas insatisfechas en agua.
- La proporción de hogares con esta condición es $P = 7.5\%$.
- La correlación intraclase (ρ_y) de la variable de interés con las Unidades Primarias de Muestreo (UPM) es 0.045.

Resultados del ejercicio

Hogares promedio por UPM $(\bar{n}_{II})$	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})
10	1.3	436	4360

Escenarios alternativos

Tabla 10: Tabla de muestreo para la estimación de la proporción de hogares sin agua potable en el ejemplo.

Hogares promedio por UPM ($\bar{n}_{II}$)	DEFF	Muestra de UPM (n_I)	Muestra de hogares (n_{II})
5	1.1	758	3790
10	1.3	436	4360
15	1.5	328	4924
20	1.6	274	5490
25	1.8	242	6057
30	2.0	221	6624
35	2.2	205	7190
40	2.3	194	7757
45	2.5	185	8323

Tamaño de muestra para UPM y personas

Tamaño de muestra para UPM y personas

- En algunos casos, la variable de interés está a nivel de persona, y el cuestionario no incluye preguntas sobre el hogar.
- Se dispone de un inventario detallado de personas en cada Unidad Primaria de Muestreo (UPM).
- Se evita la selección de hogares y se dirige directamente a la selección de personas en función del inventario detallado de cada UPM.

Cálculo del tamaño de muestra

- **Definir la población de interés:** Se expresa explícitamente a través de la variable r, que denota el porcentaje de personas con la característica de interés en la población.
- **Número promedio de personas:** Se define $\bar{n}$ como el número promedio de personas que se desean encuestar por cada UPM, con el objetivo de evaluar distintos escenarios de muestreo.
- **Calcular el efecto de diseño:** $DEFF$ se establece como una función de la correlación entre la variable de interés y la conformación de las UPM ($DEFF \approx 1 + (\bar{n} - 1)\rho_y$).

Cálculo del tamaño de muestra

- **Tamaño de muestra de personas:** Se emplean expresiones de tamaño de muestra para muestreos complejos para garantizar precisión en la inferencia. Las expresiones coinciden con el primer escenario.
- **Tamaño de muestra final:** Se calcula el número total de personas necesarias para observar a quienes forman parte de la población objetivo (n/r).
- **Cálculo del número de UPM:** Se determina el número de UPM necesarias para el muestreo a partir de la relación $n_I = \frac{n}{\bar{n}}$.

Ejemplo: ingreso promedio en personas empleadas

Suponga que se desea estimar el ingreso promedio en las personas empleadas con un margen de error relativo máximo admisible del 2%. La variable de interés (ingreso) es continua y se estima que la media oscila alrededor de $\bar{y}_U = 1458$ dólares con una varianza de $S^2_{y_U} = 2191^2$. Nótese que la población objetivo son todas las personas empleadas, cuya proporción se estima en $r = 46\%$. La correlación intraclase de la característica de interés es $\rho_y = 0.038$.

Contexto del Ejemplo

- Estimar el ingreso promedio en personas empleadas con un margen de error relativo máximo del 2%.
- La variable de interés (ingreso) es continua, con una media estimada de $\bar{y}_U = 1458$ dólares y una varianza de $S^2_{y_U} = 2191^2$.
- La población objetivo son todas las personas empleadas, estimándose que representan el 46% de la población.
- La correlación intraclase de la variable de interés es $\rho_y = 0.038$.
- Se obtiene un efecto de diseño $DEFF = 1.8$ al seleccionar $\bar{n} = 23$ personas de la población de interés por UPM.

Resultados para el escenario seleccionado

- Se seleccionan y enlistan en promedio 50 personas por UPM, lo que da como resultado $n = 28029$ personas empleadas en la muestra distribuidas en $n_I = 28029/23 \cong 1219$ UPM.
- El operativo de campo abarcaría la selección y enlistamiento de 60933 personas, de las cuales se esperaría que 28029 fueran de la población de interés (personas empleadas).

Personas seleccionadas por UPM $(\bar{n}/r)$	Personas empleadas por UPM $\bar{n}$	DEFF	Muestra de UPM (n_I)	Muestra de personas empleadas (n)	Muestra de personas (n/r)
50	23	1.8	1219	28029	60933

Escenarios adicionales

Tabla 12: Tabla de muestreo para la estimación del ingreso promedio en personas empleadas en el ejemplo.

Personas seleccionadas por UPM $(\bar{n}/r)$	Personas empleadas por UPM $\bar{n}$	DEFF	Muestra de UPM (n_I)	Muestra de personas empleadas (n)	Muestra de personas (n/r)
25	12	1.4	1857	21360	46435
50	23	1.8	1219	28029	60933
75	34	2.3	1006	34695	75424
100	46	2.7	899	41360	89913
125	58	3.1	835	48023	104398

Ejemplo: proporción de personas analfabetas pobres

Suponga que se quiere estimar la proporción de incidencia de la pobreza sobre la población analfabeta con un margen de error relativo máximo admisible del 15%. Se ha estimado que alrededor del $r = 14\%$ de las personas del país no saben leer ni escribir. Por otro lado, tal como se vio en un ejemplo anterior, el fenómeno de la pobreza está estimado en $P = 4\%$. y la correlación intraclase de la característica de interés es $\rho_y = 0.0454$.

Contexto del Ejemplo

- Estimar la proporción de personas analfabetas pobres con un margen de error relativo máximo del 15%.
- Se busca estimar la proporción de personas analfabetas en situación de pobreza.
- La proporción de personas analfabetas en el país es estimada en $r = 14\%$.
- La incidencia de la pobreza en la población general es $P = 4\%$.
- La correlación intraclase de la variable de interés es $\rho_y = 0.0454$.

Metodología de muestreo

- Se selecciona un promedio de 14 personas analfabetas por UPM.
- Esto implica seleccionar y enlistar 100 personas por UPM $(14/0.14)$.
- El efecto de diseño resultante es $DEFF = 1.6$.
- La muestra total sería de 4574 personas analfabetas, distribuidas en 327 UPM.
- En total, se enlistarían 32671 personas en la muestra.

Resultados de otros escenarios

Tabla 13: Tabla de muestreo para la estimación de la proporción de personas analfabetas pobres en el ejemplo.

Personas seleccionadas por UPM $(\bar{n}/r)$	Personas analfabetas por UPM $(\bar{n})$	DEFF	Muestra de UPM (n_I)	Muestra de personas analfabetas (n)	Muestra de personas (n/r)
25	3.5	1.1	917	3211	22936
50	7.0	1.3	524	3665	26179
75	10.5	1.4	392	4120	29429
100	14.0	1.6	327	4574	32671
125	17.5	1.7	287	5029	35921

Tamaño de muestra para otros parámetros de interés

Introducción

- En las encuestas de hogares también surgen escenarios particulares que llevan a sugerir distintos caminos para la adopción de un determinado tamaño de muestra.
- Se parte de una población U dividida en dos subpoblaciones U_1 y U_2, con tamaños respectivos N_1 y N_2. Se busca entender la diferencia de proporciones entre estas subpoblaciones, como el caso de los niños con desnutrición por sexo.

Parámetro de interés

La diferencia entre las proporciones se define como

$$\theta = P_1 - P_2 = \frac{N_{d1}}{N_1} - \frac{N_{d2}}{N_2},$$

donde P_i representa la proporción de individuos desnutridos en la subpoblación U_i, $N_{di} = \sum_{k \in U_i} z_{dik}$ $(i = 1, 2)$ y z_{dik} es una característica dicotómica que indica si el individuo k-ésimo de la subpoblación U_i está en estado de desnutrición.

Estimador insesgado

Bajo un muestreo aleatorio simple, el estimador insesgado para θ es

$$\hat{\theta} = \hat{P}_1 - \hat{P}_2 = \frac{\hat{N}_{d1}}{N_1} - \frac{\hat{N}_{d2}}{N_2},$$

donde $\hat{P}_i$ es la proporción estimada de desnutridos en la muestra asociada con la subpoblación U_i, $\hat{N}_{di} = \frac{N_i}{n_i} \sum_{k \in s_i} z_{dik}$ y s_i es la muestra asociada con la población U_i.

La varianza de $\hat{\theta}$ se calcula como

$$Var(\hat{\theta}) = Var\left(\hat{P}_1\right) + Var\left(\hat{P}_2\right) - 2Cov\left(\hat{P}_1, \hat{P}_2\right)$$

Relaciones entre subpoblaciones:

- Sea $|U_i|$ la cardinalidad del conjunto U_i, se definen las siguientes relaciones:

$$T_i = \frac{|U_1 \cap U_2|}{|U_i|} \qquad i = 1, 2.$$

De esta forma, T_1 y T_2 corresponde al porcentaje de traslape de las subpoblaciones.

- Sea $R_{1,2}$ como la correlación de Pearson entre los datos observados de ambas subpoblaciones.

Covarianza entre subpoblaciones

- La covarianza entre este par de estimadores estaría determinada por la siguiente relación (Kish 2004):

$$Cov(\hat{P}_1, \hat{P}_2) = \sqrt{Var(\hat{P}_1)}\sqrt{Var(\hat{P}_2)}\sqrt{T_1}\sqrt{T_2}R_{1,2}$$

- Si las poblaciones U_1 y U_2 son estratos que inducen conjuntos dijuntos y la selección de la muestra en cada uno es independiente por diseño, entonces $Cov(\hat{P}_1, \hat{P}_2) = 0$.
- Si, no existe independencia en el muestreo de ambas poblaciones, entonces $R_{1,2} \neq 0$ necesariamente.
- La correlación se debe evaluar a través de las UPM.

Supuestos y tamaño de muestra óptimo

1. Asumir que las subpoblaciones son grandes y por ende $N_1 = N_2 = N$.
2. Por lo anterior, asumir que los tamaños de muestra pueden ser iguales, tales que $n_1 = n_2 = n$.

Consideraciones sobre el diseño de muestreo

Si el levantamiento de las observaciones fue mediante un diseño de muestreo complejo con un efecto de diseño[1] $(DEFF)$ no ignorable y mayor a uno, entonces la varianza tomaría la siguiente forma

$$Var(\hat{\theta}) = \frac{DEFF}{n}\left(1 - \frac{n}{N}\right) S_{\theta}^{2}$$

En donde, definiendo a $Q_i = 1 - P_i$, se tiene que:

$$S_{\theta}^{2} = P_1 Q_1 + P_2 Q_2 - 2\sqrt{T_1}\sqrt{T_2} R_{1,2} \sqrt{P_1 Q_1}\sqrt{P_2 Q_2}$$

[1] Recuerde que si el muestreo es aleatorio simple, el efecto de diseño es $DEFF = 1$.

Intervalo de Confianza y Margen de Error (ME)

- Un intervalo de confianza del 95% para la diferencia de proporciones está dado por

$$IC(95\%)_\theta = \hat{\theta} \pm z_{1-\alpha/2}\sqrt{\frac{DEFF}{n}\left(1-\frac{n}{N}\right)S_\theta^2}$$

- El margen de error (ME) de la encuesta debe ser tal que:

$$ME < z_{1-\alpha/2}\sqrt{\frac{DEFF}{n}\left(1-\frac{n}{N}\right)S_\theta^2}$$

Tamaño de las Muestra

Se tiene que la muestra en cada subgrupo debe ser mayor que:

$$n > \frac{DEFF\ S_\theta^2}{\frac{ME^2}{z_{1-\alpha/2}^2} + \frac{DEFF\ S_\theta^2}{N}}$$

Consideraciones

Dependiendo de los pocentajes de traslape $\sqrt{T_1}$, $\sqrt{T_2}$ y de la correlación de la característica de interés en ambas subpoblaciones $R_{1,2}$, la varianza S_θ^2 tomará diferentes formas:

1. Si no hay traslape, $T_1 = T_2 = 0$, y $S_\theta^2 = P_1Q_1 + P_2Q_2$.
2. Si hay traslape completo, $T_1 = T_2 = 1$ y

$$S_\theta^2 = P_1Q_1 + P_2Q_2 - 2R_{1,2}\sqrt{P_1Q_1}\sqrt{P_2Q_2}$$

.

3. Si hay traslape parcial y balanceo, $T_1 = T_2 = T$ y si además se considera que las varianzas en cada subgrupo o periodo son similares $P_1Q_1 = P_2Q_2 = PQ$, entonces $S_\theta^2 = 2PQ(1 - TR_{1,2})$.

Covarianza en comparaciones mensuales

- Comparación de la tasa de desempleo nacional entre dos meses consecutivos.
- Supuesto de independencia en el muestreo de los dos meses.
- El porcentaje de traslape de muestra entre los dos meses es nulo ($T_1 = T_2 = 0$).
- La covarianza entre los meses se anula.
- Varianza del estimador en este caso:

$$Var(\hat{P}_1 - \hat{P}_2) = Var(\hat{P}_1) + Var(\hat{P}_2)$$

.

Covarianza en comparaciones trimestrales o anuales

- Comparación de la tasa de desempleo nacional entre trimestres consecutivos o entre el mismo mes de dos años consecutivos.
- Esquema rotativo 2(2)2.
- No hay independencia en el muestreo de los dos trimestres consecutivos debido al traslape del 50% ($T_1 = T_2 \approx 0.5$).
- Existe correlación natural entre las viviendas comunes en el panel.
- La correlación se calcula sobre los individuos comunes en el panel y sobre la variable dicotómica que induce la tasa de desempleo.
- Término de covarianza en este caso:

$$Cov(\hat{P}_1, \hat{P}_2) = \frac{1}{2}\sqrt{Var(\hat{P}_1)}\sqrt{Var(\hat{P}_2)}R_{1,2}$$

.

Covarianza en comparaciones de un mismo mes

- Comparación de la tasa de desempleo entre hombres y mujeres en un mismo mes.
- No hay independencia en el muestreo de hombres y mujeres.
- T_1 es la proporción de hombres y T_2 es la proporción de mujeres ($T_1 \neq T_2$).
- Existe correlación natural entre las UPM que contienen tanto a hombres como a mujeres ($R_{12} \neq 0$).
- Término de covarianza:

$$Cov(\hat{P}_1, \hat{P}_2) = \sqrt{Var(\hat{P}_1)}\sqrt{Var(\hat{P}_2)}\sqrt{T_1}\sqrt{T_2}R_{1,2}$$

Covarianza en comparaciones de dos regiones de un mismo mes

- Comparación de la tasa de desempleo entre dos regiones del mismo país en un mismo mes.
- Existe independencia en el muestreo de las dos regiones.
- T_1 es la proporción de personas de la primera ciudad y T_2 es la proporción de personas de la segunda ciudad.
- No hay correlación entre las UPM seleccionadas en estas regiones ($R_{12} = 0$).
- La varianza del estimador sería igual a

$$Var(\hat{d}) = Var(\hat{P}_1) + Var(\hat{P}_2)$$

Tamaño de muestra para la estimación del impacto en dos mediciones longitudinales

Definición del efecto de intervención

- El efecto de intervención se define como la diferencia en diferencias de las proporciones entre dos grupos en dos momentos de tiempo distintos.
- El efecto se define como:

$$\theta = (P_{1,1} - P_{2,1}) - (P_{1,2} - P_{2,2})$$

En donde $P_{i,j}$ $(i, j = 1, 2.)$ corresponden a las proporciones del grupo i en la oleada j.

Tamaño de Muestra Mínimo

- El tamaño de muestra[2] mínimo necesario para lograr una estimación confiable de esta diferencia, con menos del $ME \times 100\%$ de margen de error, es:

$$n \geq \frac{DEFF\ S_\theta^2}{\frac{ME^2}{z_{1-\alpha/2}^2} + \frac{DEFF\ S_\theta^2}{N}}$$

`En donde`

$$S_\theta^2 = (P_{1,1}Q_{1,1} + P_{1,2}Q_{1,2} + P_{2,1}Q_{2,1} + P_{2,2}Q_{2,2})(1 - TR)$$

En donde T corresponde a la tasa de traslape y R se define como la correlación entre las dos oleadas.

[2] El tamaño de muestra de toda la encuesta es $4n$, en las dos oleadas, puesto que se debe seleccionar n elementos en cada grupo y en cada oleada.

Tamaño de muestra para el contraste de hipótesis en la diferencia de proporciones

Contraste de hipótesis en la diferencia de proporciones

- El investigador desea contrastar una hipótesis sobre la diferencia entre dos grupos de interés en la población.
- Se plantea un sistema de hipótesis

$$H_o : P_1 - P_2 = 0 \qquad vs. \qquad H_a : P_1 - P_2 = D > 0$$

Regla de decisión

Se utiliza la distribución normal de los estimadores de las proporciones para tomar decisiones sobre el rechazo de la hipótesis nula. La regla de decisión es:

$$\frac{\hat{P}_1 - \hat{P}_2}{\sqrt{Var(\hat{P}_1 - \hat{P}_2)}} > z_{1-\alpha}$$

Consideraciones en muestreo complejo

Si el diseño de muestreo es complejo, se ajusta la regla de decisión considerando el efecto de diseño y el tamaño de muestra. La nueva regla de decisión es:

$$\frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\frac{DEFF}{n}\left(1 - \frac{n}{N}\right)\left(P_1 Q_1 + P_2 Q_2\right)}} > z_{1-\alpha}$$

Control de la probabilidad de error tipo II:

Se introduce el concepto de potencia del contraste para controlar la probabilidad de no rechazar la hipótesis nula cuando es falsa. La potencia se calcula como:

$$1-\beta \geq \Phi\left(z_{1-\alpha}-\frac{D}{\sqrt{\frac{DEFF}{n}\left(1-\frac{n}{N}\right)\left(P_1Q_1+P_2Q_2\right)}}\right)$$

Entonces, dado que la función $\Phi()$ es creciente, se tiene que

$$z_{1-\beta} \geq z_{1-\alpha}-\frac{D}{\sqrt{\frac{DEFF}{n}\left(1-\frac{n}{N}\right)\left(P_1Q_1+P_2Q_2\right)}}$$

Fórmula del tamaño de muestra

Se presenta la fórmula para calcular el tamaño de muestra necesario para el contraste.

$$n \geq \frac{DEFF(P_1Q_1 + P_2Q_2)}{\frac{D^2}{(z_{1-\alpha}+z_\beta)^2} + \frac{DEFF(P_1Q_1+P_2Q_2)}{N}}$$

Algunas relaciones de interés para proporciones

- El estado laboral de los individuos es crucial para comprender la dinámica del mercado laboral.
- Se buscan indicadores como la tasa de desempleo actual, la variación neta entre períodos y los flujos entre categorías de empleo.

Algunas relaciones de interés para proporciones

Encuestas de hogares para abordar adecuadamente los estudios de fuerza laboral

- Encuestas repetidas: mediciones similares en diferentes momentos a diferentes personas.
- Encuestas de panel: mediciones en diferentes momentos a las mismas personas.
- Encuestas rotativas: incluyen y siguen elementos en la muestra durante un período específico.

Algunas relaciones de interés para proporciones

Impacto de la probabilidad de éxito:

- La varianza de la variable de diseño (dicotómica) es máxima cuando la probabilidad de éxito es 0.5.
- Las intervenciones gubernamentales pueden cambiar esta probabilidad y afectar el tamaño de la muestra.

Control del margen de error:

- A medida que la proporción disminuye, el tamaño de la muestra aumenta.
- La función de varianza es simétrica alrededor de 0.5, por lo que el tamaño de muestra necesario es similar para una proporción y su complemento aditivo.

Ejemplos prácticos

Estimación de proporciones

Primer escenario:

- Tasa de desempleo baja: $P = 0.05$, margen de error $ME = 0.0025$.
- Intervalo de confianza:$IC = 0.05 \pm 0.0025 = (0.0475, 0.0525)$.
- Tamaño de muestra requerido: alrededor de 55169.

Segundo escenario:

- Tasa de desempleo alta: $P = 0.2$, margen de error $ME = 0.01$.
- Intervalo de confianza: $IC = 0.2 \pm 0.01 = (0.19, 0.21)$.
- Tamaño de muestra requerido: 12144.

Observaciones

- Ambos escenarios tienen el mismo margen de error relativo (MER), definido como $MER = \frac{ME}{P}$, siendo este igual al 5%.
- La muestra debe ser mayor cuando la incidencia del fenómeno es baja en la población.

Estimación de proporciones

Tercer escenario:

- Tamaño de muestra: $n = 10000$, proporción $P = 0.2$.
- Coeficiente de variación: 2.8%, margen de error: 1.1%.
- Todas las proporciones estimadas tienen un margen de error inferior al 1.4%.

Cuarto escenario:

- Tamaño de muestra: $n = 40000$, proporción $P = 0.05$.
- Coeficiente de variación: 3%, margen de error: 0.2%.
- Todas las proporciones estimadas tienen un margen de error inferior al 0.7%.

Observaciones

- El tamaño de muestra necesario para alcanzar un margen de error particular es el mismo para una proporción y su complemento aditivo.
- El coeficiente de variación puede variar entre una proporción y su complemento aditivo, lo que afecta el tamaño de muestra requerido.
- Anticipamos un mayor tamaño de muestra para proporciones bajas y un tamaño de muestra más pequeño para proporciones altas, manteniendo el MER constante.

Estimación de cambios netos

Nuestra atención a los cambios netos en la tasa de desempleo durante dos períodos, $\Delta = |P_1 - P_2|$.

Quinto escenario:

- Cambio neto pequeño esperado: $\Delta \approx |0.22 - 0.20| = 0.02$, margen de error $ME = 0.001$.
- Intervalo de confianza: $IC = 0.02 \pm 0.001 = (0.019, 0.021)$.
- Tamaño de muestra requerido: alrededor de 96224.

Estimación de cambios netos

Sexto escenario:

- Cambio neto pequeño esperado: $\Delta \approx |0.05 - 0.03| = 0.02$, margen de error $ME = 0.001$.
- Intervalo de confianza: $IC = 0.02 \pm 0.001 = (0.019, 0.021)$.
- Tamaño de muestra requerido: 59536.

Séptimo escenario:

- Cambio neto significativo esperado: $\Delta \approx |0.05 - 0.20| = 0.15$, margen de error $ME = 0.0075$.
- Intervalo de confianza: $IC = 0.15 \pm 0.0075 = (0.1425, 0.1575)$
- Tamaño de muestra requerido: alrededor de 22083.

Observación sobre el margen de error relativo (MER)

- Los escenarios quinto, sexto y séptimo dan como resultado el mismo MER, definido como $MER = \frac{ME}{\overline{ME}}$ y tienen un valor del 5%.
- A pesar de tener el mismo MER, el tamaño de muestra varía según la configuración de las proporciones.

Observaciones

- La configuración de las proporciones y la esperanza de cambio neto afectan el tamaño de muestra necesario.
- Es crucial ajustar el tamaño de muestra según las expectativas de cambio y la configuración de las proporciones.
- Se necesita un mayor tamaño de muestra si no se esperan cambios significativos en las tasas de desempleo.
- Si las tasas son similares pero las proporciones son altas, se requerirá un gran tamaño de muestra para estimar cambios netos con precisión.

Algunas consideraciones adicionales sobre el tamaño de muestra

Asignación del tamaño de muestra en los estratos de muestreo

Se supone que el tamaño de la muestra general es n y que hay H estratos fijos; por ende, se quiere determinar los tamaños de muestra n_h para cada estrato $(h = 1, ..., H)$.

Asignación proporcional:

- Cada estrato tiene una muestra proporcional a su tamaño poblacional.
- La fracción de muestreo es constante en todos los estratos.
- Tamaño de muestra en cada estrato:

$$n_h = f \times N_h$$

.

Asignación de Neyman

- Busca maximizar la eficiencia estadística de la estrategia de muestreo.
- El tamaño de muestra en cada estrato minimiza la varianza de la estrategia de muestreo.
- Tamaño de muestra en cada estrato:

$$n_h = n \times \frac{N_h S_{yU_h}}{\sum_{h=1}^{H} N_h S_{yU_h}}$$

.

Asignación de Kish:

- Proporciona un equilibrio entre la asignación proporcional y la uniforme.
- Ajusta el tamaño de muestra considerando el tamaño relativo del estrato y un índice de importancia relativa.
- Tamaño de muestra en cada estrato:

$$n_h = n \times \frac{\sqrt{\frac{1}{H^2} + I\ W_h^2}}{\sum_{h=1}^{H} \sqrt{\frac{1}{H^2} + I\ W_h^2}}$$

donde $W_h = N_h/N$, e $I \geq 0$ es el índice de asignación de Kish, que denota la importancia relativa entre las estimaciones nacionales y las de cada estrato.

Ajustes por subcobertura

- La ausencia de respuesta es una realidad en encuestas de hogares, lo que puede afectar la precisión y generar sesgos en las estimaciones.
- En encuestas longitudinales, la atrición de los hogares en los paneles también debe ser considerada para evitar problemas de sesgo y baja confiabilidad.

Ajustes por subcobertura

- Se sugiere ajustar el tamaño de muestra inicial para compensar la ausencia de respuesta, dividiendo el tamaño de muestra por la probabilidad estimada de respuesta:

$$n_{\text{final}} = \frac{n_{\text{inicial}}}{\phi}$$

.

- Los institutos nacionales de estadística pueden estimar esta probabilidad y ajustar los tamaños de muestra de manera diferenciada por estratos si la ausencia de respuesta afecta de manera desigual a diferentes poblaciones.

Sustituciones y reemplazos

- En las encuestas de hogares en Latinoamérica, es común sustituir Unidades Primarias de Muestreo (UPM) y viviendas cuando no se obtiene respuesta debido a diversas razones, como problemas de seguridad o falta de consentimiento.
- La metodología de *estratificación implícita* se utiliza para seleccionar automáticamente reemplazos adecuados basados en subgrupos poblacionales similares.
- La estratificación implícita se implementa en la primera etapa del diseño de muestreo, donde las UPM se seleccionan sistemáticamente con probabilidades desiguales.

Sustituciones y reemplazos

- Los reemplazos de UPM se eligen dentro del mismo municipio, departamento y zona, manteniendo un número similar de viviendas para preservar la representatividad.
- Aunque la estratificación implícita ayuda a reducir sesgos por ausencia de respuesta, puede introducir sesgos en las estimaciones debido a diferencias entre áreas accesibles y de difícil acceso.
- Se recomienda evaluar y monitorear los efectos de las sustituciones en la precisión de los estimadores resultantes para mitigar posibles sesgos.

¡Gracias!

Email: andres.gutierrez@cepal.org

Referencias

Kish, Leslie. 2004. *Statistical Design for Research*. Wiley classic biblioteca edición. Wiley. https://www.wiley.com/en-us/Statistical+Design+for+Research-p-9780471691204.

UN. 2008. *Designing household survey samples: practical guidelines*. Studies en methods / United Nations, Department of Economic y Social Affairs, Statistics Division Series F. United Nations.