



Buenas cifras,  
mejores vidas

# Muestreo de Encuestas de Hogares Con R

• . . .  
• . . . INEC – Diseño Muestral  
• . . .

Diciembre - 2025

[www.ecuadorencifras.gob.ec](http://www.ecuadorencifras.gob.ec)





## Enlaces importantes



psuR



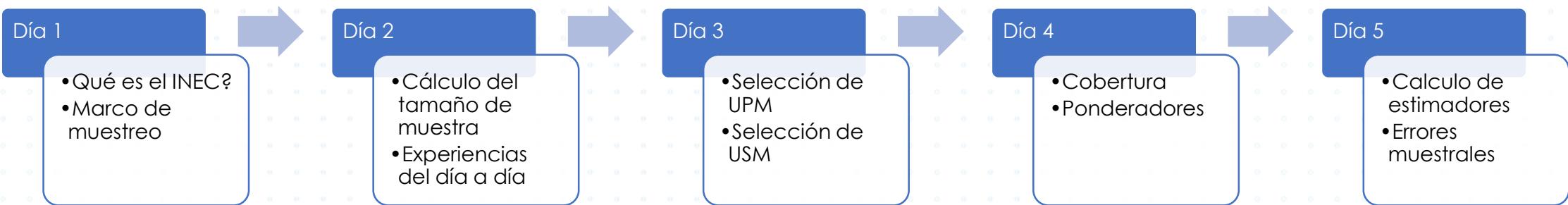
Base ejercicio tamaño de  
muestra



- 01 ► Ejercicio psuR
- 02 ► Cálculo tamaño de muestra
- 03 ► Parámetros muestrales
- 04 ► Ejercicios
- 05 ► Discusión

# Contenido del curso

# Contenido del curso



# Ejercicio psuR

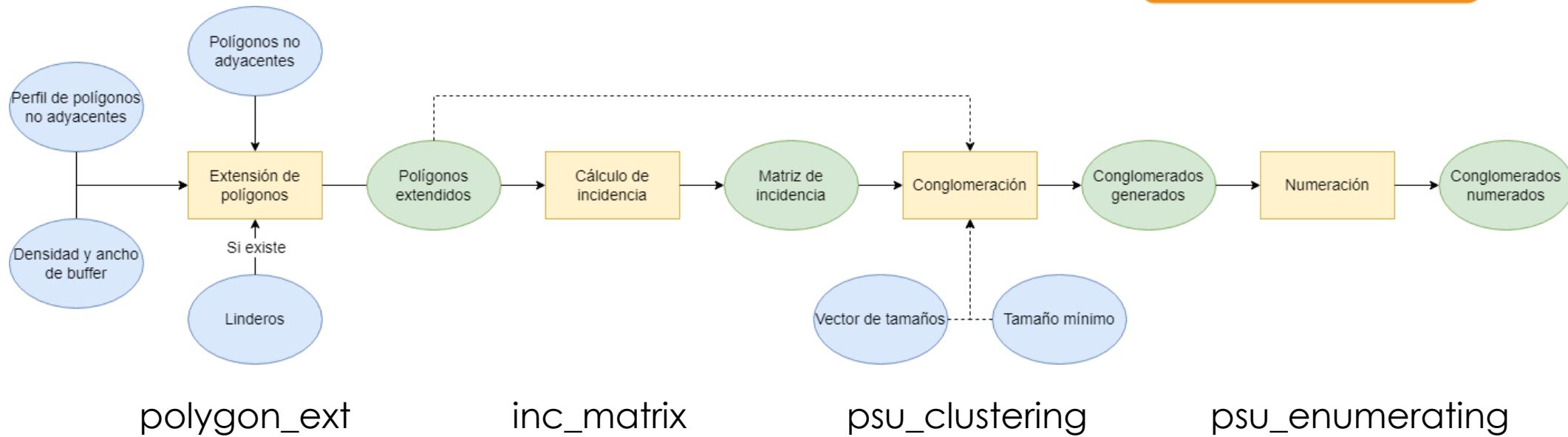
# Ejercicio 2

Instalación:

1. Instalar y cargar la librería devtools.
2. Instalar la librería psuR.



psuR



# Tamaño de muestra



# Tamaño de muestra

Su definición toma en cuenta dos aspectos importantes:

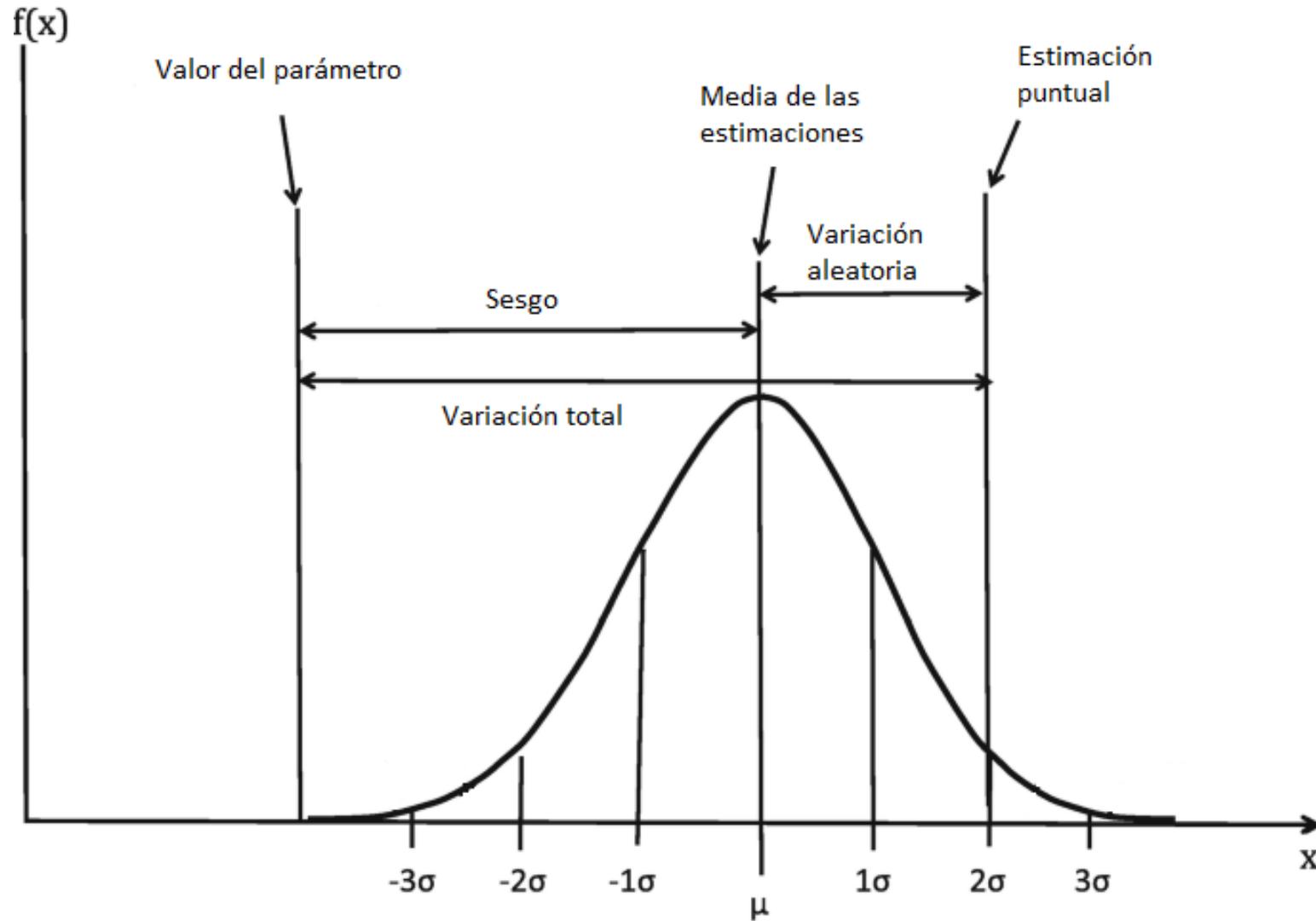
- Proceso de selección: reglas y operaciones bajo las cuales algunos elementos de la población son incluidos en la muestra (Probabilidad de selección).
- Proceso de estimación: fórmulas para calcular las estimaciones de una muestra probabilística (estimador y su varianza).

En su desarrollo se considera:

- Definir las variables de muestreo: tipo de variables a medir y población objetivo.
- La precisión deseada: tamaño que se traduce en costo.



# Función de distribución





Mide la **dispersión o variabilidad** de la característica que se está estudiando en la población.

En muestreo, el término  $p(1 - p)$  es la varianza clave, donde  $p$  es la proporción estimada.

Si se desconoce el valor de  $p$ , se utiliza el valor 0.5 para maximizar la varianza  $p(1 - p) = 0.25$ .

Una **mayor varianza** requiere un **mayor tamaño muestral** para capturar la diversidad de la población con precisión.



# Nivel de confianza

Probabilidad de que el Intervalo de Confianza calculado contenga el **verdadero valor poblacional** (el parámetro que buscamos).

**Valores Comunes:** 90%, 95% o 99%.

**Valor z crítico:** El NC se traduce en un valor  $z$  de la distribución normal (p. ej., 1.96 para 95%).

**Impacto:** Para lograr un mayor Nivel de Confianza (ej. 99%), el valor  $z$  es mayor, lo que obliga a tener un mayor tamaño muestral para aumentar la seguridad de la estimación.



# Error absoluto

Es la **máxima diferencia** que el investigador está dispuesto a aceptar entre la estimación muestral y el valor poblacional real.

**Unidades:** Se expresa en las mismas unidades (ej.  $\pm 5\%$  o 0.05 para una proporción).

**Impacto Crucial en el Tamaño Muestral:** El error  $e$  está al cuadrado en el denominador de la fórmula de  $n$ .

**Reducir el error a la mitad** (buscando el doble de precisión) requiere **cuadruplicar** el tamaño de la muestra, haciendo este el factor más costoso en el diseño muestral.



# Tamaño de muestra

$$n \geq \left( \frac{z \cdot S}{e} \right)^2$$

Donde:

- $z$  : nivel de confianza.
- $S$  : varianza asociada al diseño:
- $e$  : error absoluto requerido.



# Muestreo estratificado

Es un método de muestreo probabilístico diseñado para **aumentar la representatividad y precisión** de la muestra, donde la población se divide en subgrupos mutuamente excluyentes y exhaustivos llamados **estratos**. Se toma una muestra aleatoria de todos y cada uno de los estratos definidos.

Diseño de Estratos:

- **Criterio:** La división se basa en una característica poblacional relevante para el estudio (ej. edad, sexo, nivel socioeconómico).
- **Homogeneidad Interna:** Los elementos dentro de un mismo estrato deben ser muy similares entre sí respecto a la variable de estratificación.
- **Heterogeneidad Externa:** Los diferentes estratos entre sí deben ser significativamente diferentes.

Al garantizar que cada estrato esté adecuadamente representado en la muestra, reduce el error muestral y aumenta la precisión de las estimaciones poblacionales.



# Muestreo por conglomerados

Es un método de **muestreo probabilístico** donde la población se divide primero en grupos o unidades llamados **conglomerados**.

A diferencia del muestreo estratificado, donde se muestran todos los estratos, en el muestreo por conglomerados se selecciona una **muestra aleatoria de los conglomerados** y todos los elementos dentro de los conglomerados seleccionados (o una submuestra) se incluyen en la muestra final.

Es muy útil cuando la población está dispersa geográficamente.





# Características de los conglomerados

**Homogeneidad externa (entre conglomerados):** Idealmente, cada conglomerado debería ser muy similar a los demás conglomerados en términos de las características de la población.

**Heterogeneidad interna (dentro de los conglomerados):** Idealmente, los elementos dentro de un mismo conglomerado deberían ser lo más diversos o representativos de toda la población posible.

*Un buen conglomerado es como una "mini-versión" de la población total.*

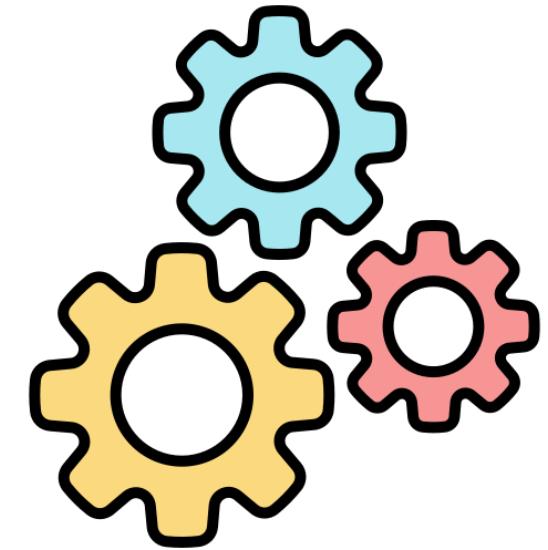
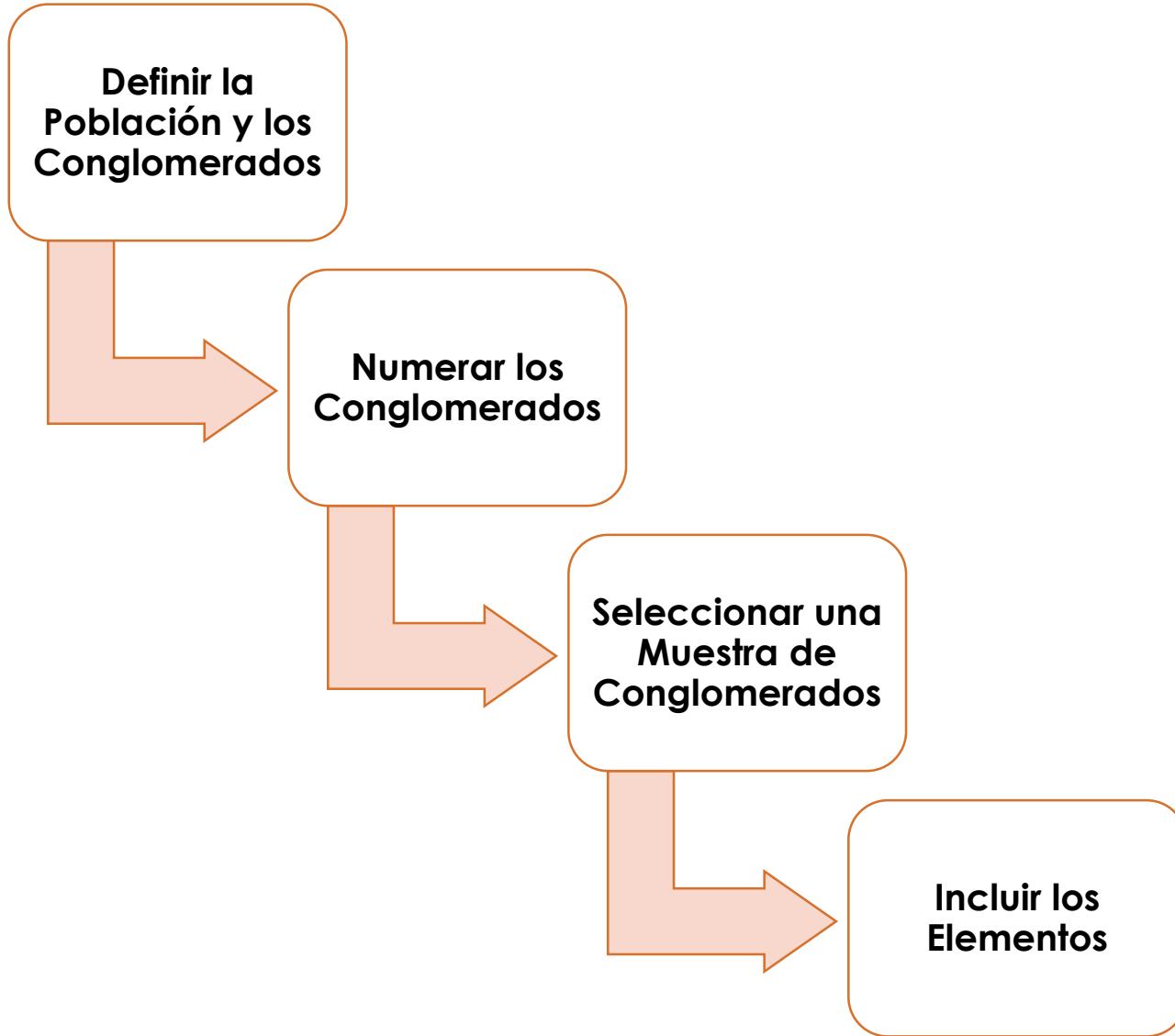
## Ejemplos de conglomerados:

- Áreas geográficas (ciudades, barrios, manzanas).
- Escuelas, hospitales, o fábricas (cada institución es un conglomerado).





# Pasos para realizar el muestreo



**Eficiencia de Costos:** Se reducen significativamente los costos de viaje y logísticos, ya que la recopilación de datos se concentra en áreas limitadas.

**Facilidad Operacional:** El marco muestral (lista completa de todas las unidades) solo se necesita para los conglomerados seleccionados, no para toda la población.

**Viabilidad:** A menudo es el único método práctico cuando la población es muy grande o está muy dispersa geográficamente.



**Mayor Error Muestral:** Si los conglomerados no son lo suficientemente heterogéneos internamente (es decir, los elementos dentro de un conglomerado son demasiado parecidos entre sí), la muestra puede no ser tan representativa, lo que resulta en una mayor varianza.

**Complejidad del Análisis:** Los métodos estadísticos para el análisis de datos suelen ser más complejos que los utilizados para el muestreo aleatorio simple.

**Muestra más grande:** Para obtener la misma precisión que un MAS se necesita visitar más unidades de observación.





# Muestreo complejo

- Se denomina muestreo complejo a la estrategia de muestreo que define mas de una etapa de selección en las cuales se consideran estratos, conglomerados o ambos.
- El 99,9% de las encuestas dirigidas a hogares del INEC tienen una estrategia de muestreo compleja bietápica, estratificada de elementos.
- Como regla general:
  - Al estratificarla la muestra su error asociado disminuye.
  - Al utilizar conglomerados, su eficiencia se reduce.
- Toda las consideraciones del diseño complejo se resumen en lo que se denomina **efecto de diseño**.



# Efecto de diseño

Mide el impacto de usar un **diseño de muestreo complejo** (como el muestreo por conglomerados) frente a un **Muestreo Aleatorio Simple (MAS)** del mismo tamaño.

$$deff = \frac{\text{Varianza del diseño muestral complejo}}{\text{Varianza del muestreo aleatorio simple}}$$

En muestreo por conglomerados, DEFF es casi siempre  $> 1$ . Indica que se necesita un tamaño de muestra  $n$  mayor para mantener la misma precisión.



# Correlación intraclass (ICC)

Medida de la **homogeneidad** de las unidades de análisis dentro de un mismo conglomerado.

Mide qué tan parecidos son los individuos agrupados.

Rango entre 0 (máxima diversidad interna) y 1 (máxima similitud interna).

La ICC es el factor clave para el Efecto de Diseño:

$$deff \approx 1 + \rho \cdot (m - 1)$$

A mayor  $\rho$  (más parecidos son los elementos dentro de un conglomerado), mayor es el *deff*, lo que reduce la precisión y obliga a aumentar el tamaño muestral.



# Tasa de no respuesta

Porcentaje de unidades seleccionadas que no participan en el estudio (rechazo, no contacto, etc.).

1. Reduce el tamaño efectivo de la muestra.
2. Puede introducir sesgo si los no respondedores difieren de los respondedores.

Para compensar la pérdida de datos, el tamaño de muestra inicial debe ser **inflado** (aumentado):

$$n \geq n_0 \cdot \frac{1}{1 - tnr}$$

El número total de unidades que componen la población objetivo.

Es crucial solo cuando la muestra  $n$  es una fracción significativa (típicamente  $> 5\%$ ) de la población total ( $N$ ).

Cuando  $n$  es grande en relación con  $N$ , se aplica un factor de corrección para reducir el tamaño de muestra requerido, ya que el muestreo es más eficiente.

$$n \geq \frac{n_0}{1 + \frac{n_0}{N}}$$

Si  $N$  es muy grande (población infinita) o si  $n/N < 5\%$ , el factor de corrección es insignificante y se puede ignorar  $N$  en el cálculo.



# Tamaño de muestra

$$n \geq \frac{N \cdot S^2 \cdot deff}{N \cdot \left(\frac{e}{z}\right)^2 + S^2 \cdot deff} \cdot \frac{1}{1 - tnr}$$

Donde:

- $z$  : nivel de confianza.
- $S$  : varianza asociada al diseño:
- $e$  : error requerido.
- $N$  : total poblacional.
- $deff$  : efecto de diseño.
- $tnr$  : tasa de no respuesta

# Ejercicios



# Ejercicio 3

Cálculo del tamaño de muestra para la ENIGHUR

1. Cargar el archivo “bbd\_ej3\_enighur.xlsx”
2. Calcular la ICC.
3. Estimar tamaños muestrales variando la submuestra por UPM, TNR, el nivel de confianza del y el error (uno a la vez). Con un número promedio de viviendas por UPM de 10.





# Experiencia día a día



## Ejercicio 4

Desde el Ministerio de Finanzas se nos informa que por varias restricciones presupuestarias, el tamaño de muestra de las provincias de Azuay, El Oro, Guayas y Zamora Chinchipe no puede superar las 700 viviendas.

- En base al ejercicio anterior, proponer 3 escenarios muestrales para cada provincia.



Buenas cifras,  
mejores vidas

