

Muestreo en estudios observacionales

Dr. César Gutiérrez Villafuerte
cesar.gutierrez@udep.pe



UNIVERSIDAD
DE PIURA
Facultad de Medicina

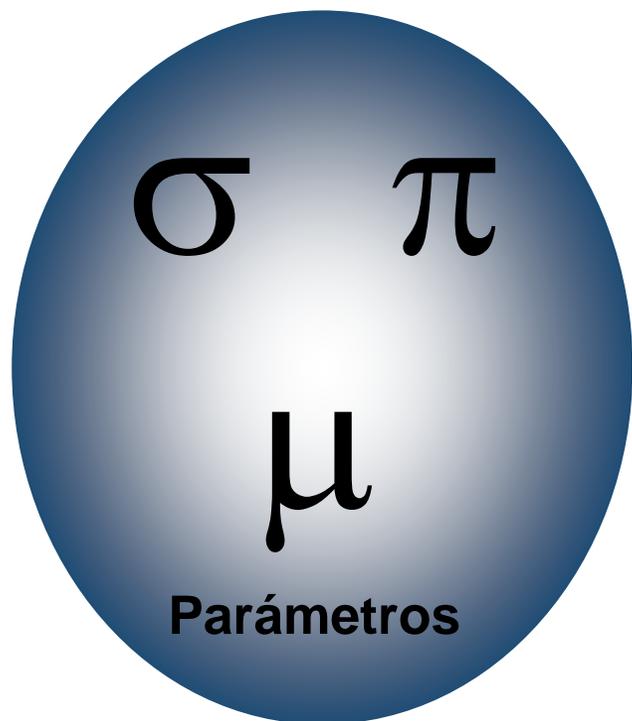
Temas a desarrollar

1. Concepto de muestreo y definiciones básicas.
2. Diseños muestrales probabilísticos y no probabilísticos.
3. Cálculo del tamaño muestral.

Concepto de muestreo

Se denomina muestreo al procedimiento mediante el cual se selecciona un conjunto de elementos de la población, para, a partir de este conjunto de elementos, inferir el valor de una o varias características de la población.

Población (N)



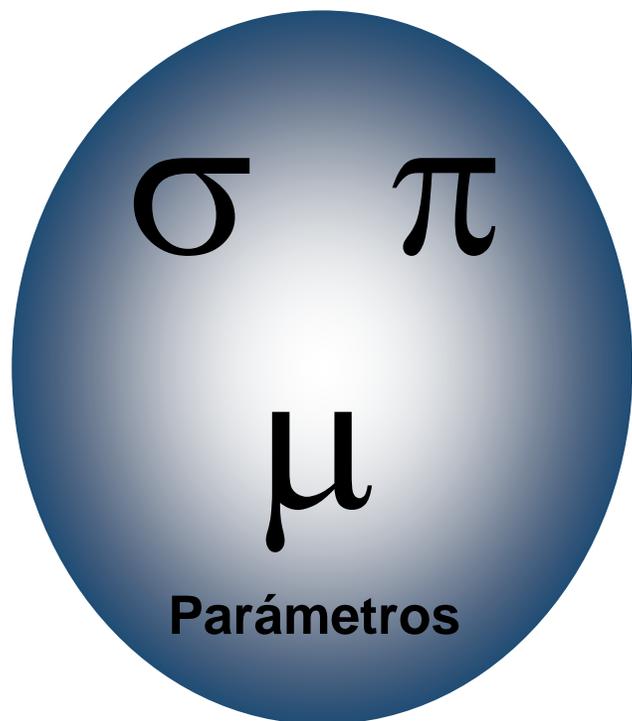
Muestreo

Muestra (n)



Inferencia

Población (N)



Muestreo

Muestra (n)



Inferencia

Población

Conjunto de elementos (objetos, situaciones o sujetos) que comparten características similares definidas por el investigador.

Población

Conjunto de elementos (objetos, situaciones o sujetos) que comparten características similares definidas por el investigador.

El término población ***en estadística*** no siempre se refiere a seres humanos (p.e. historias clínicas en un determinado hospital, muestras de sangre almacenadas en un laboratorio, viviendas en una comunidad, etc.).

Población diana

Población a la que se desea generalizar los resultados

Población diana

Población a la que se desea generalizar los resultados

Criterios de selección



Población diana

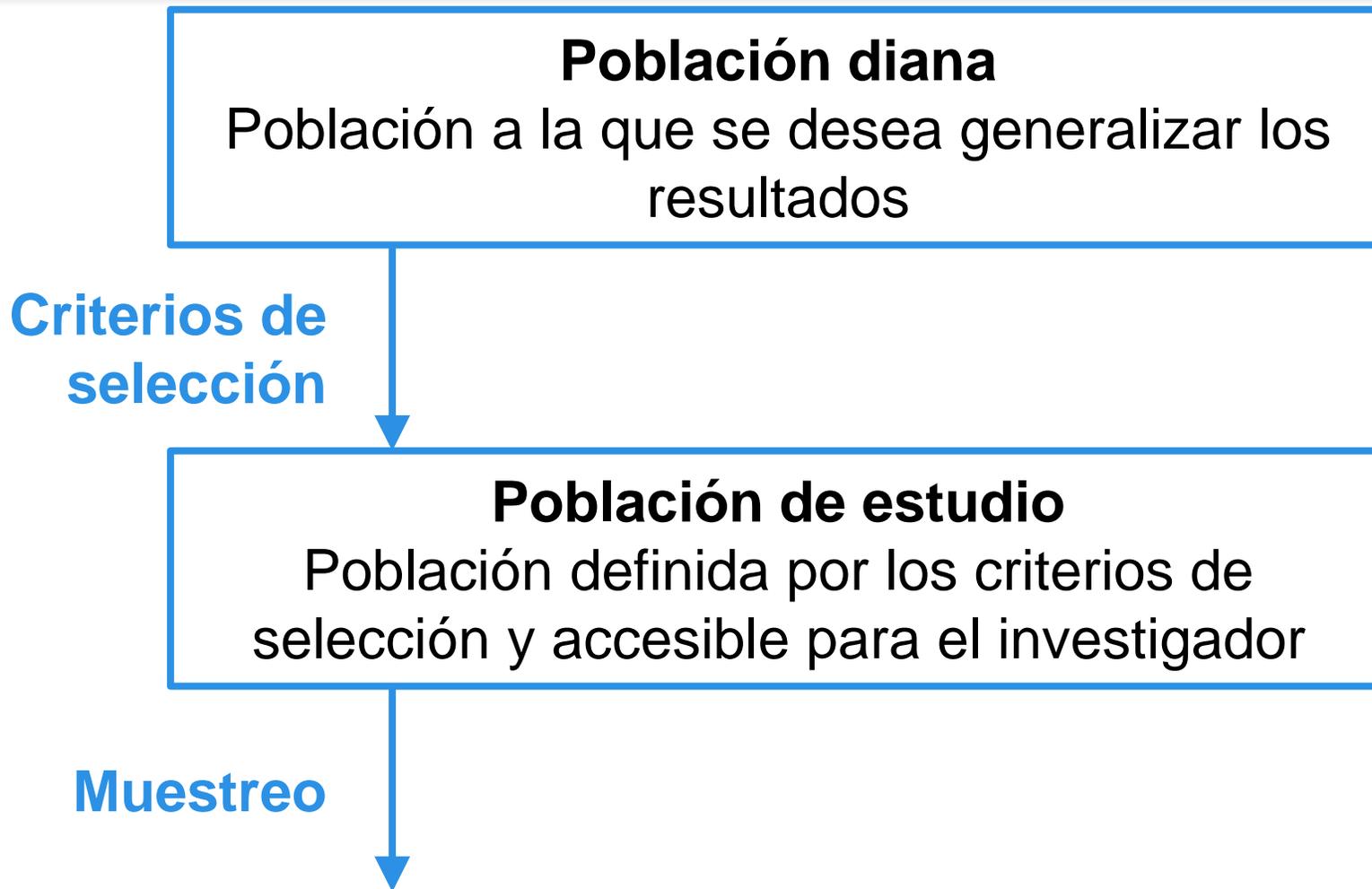
Población a la que se desea generalizar los resultados

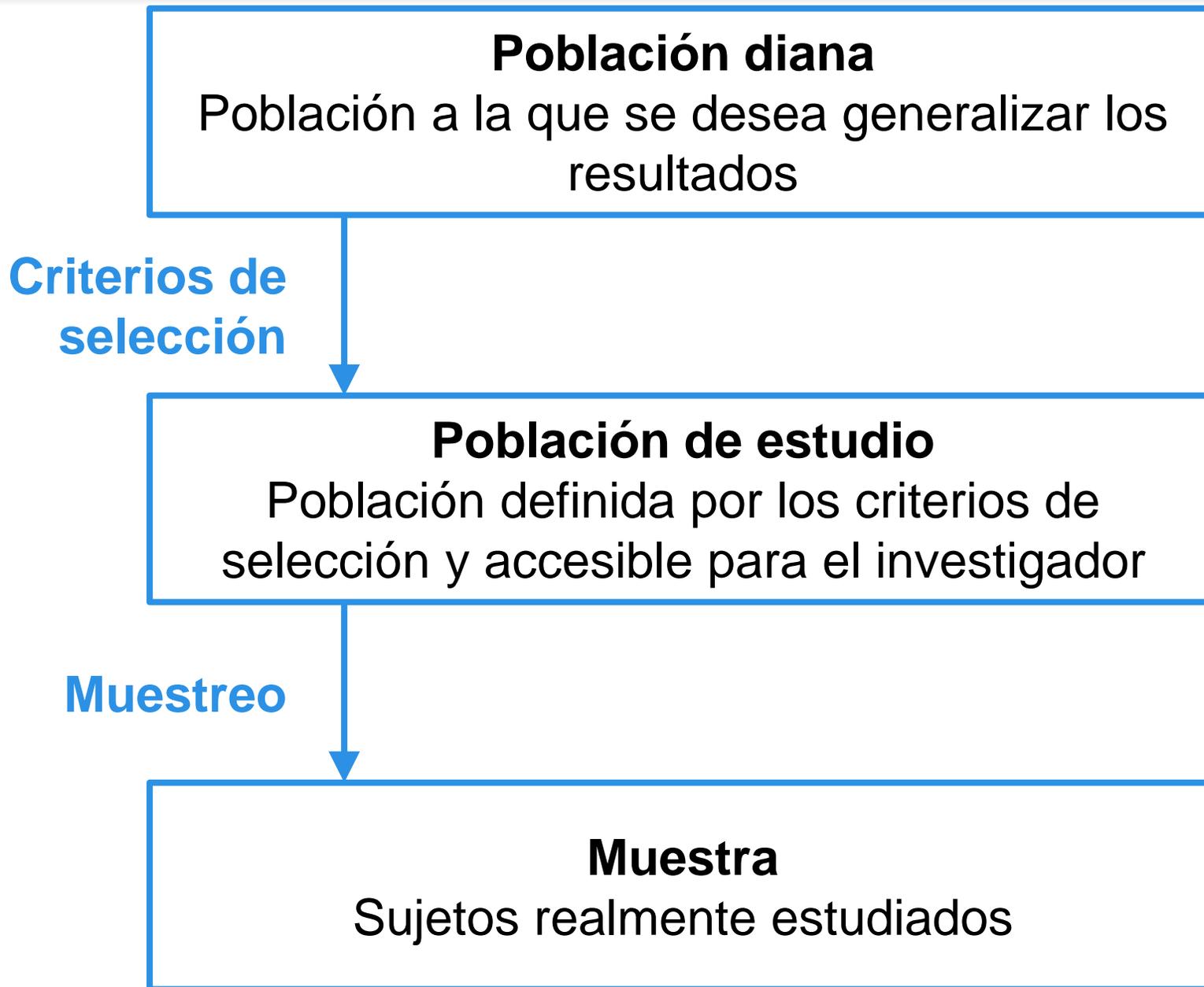
Criterios de selección

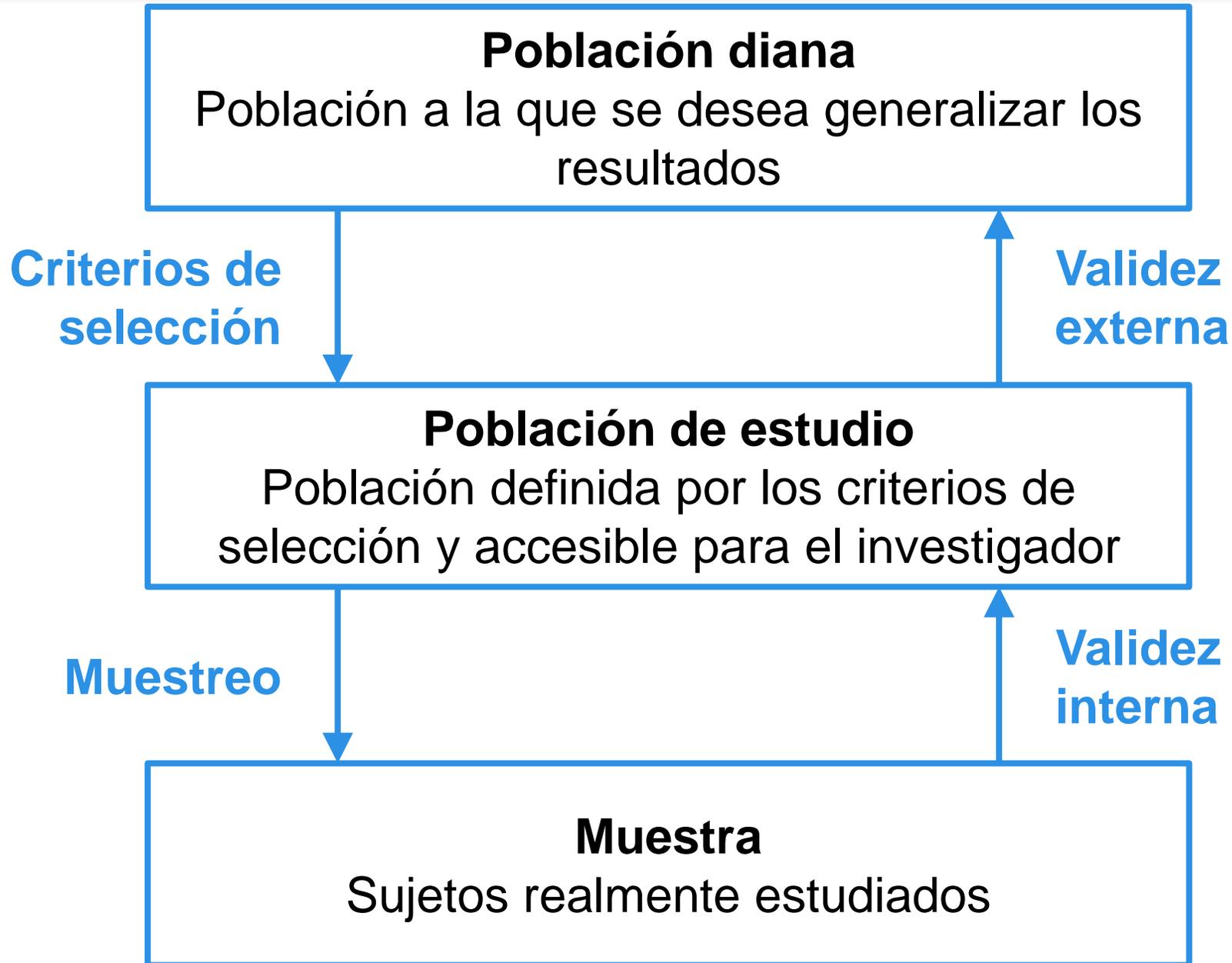


Población de estudio

Población definida por los criterios de selección y accesible para el investigador







TRUTH IN THE UNIVERSE

Target population
(GENERALIZATION FAIRLY SECURE)
Same association exists in all suburban U.S. adults

EXTERNAL VALIDITY INFERENCE #2

Accessible population
Same association exists in all Framingham adults

EXTERNAL VALIDITY INFERENCE #1

TRUTH IN THE STUDY

Intended sample
Same association exists in the designed sample of Framingham adults

INTERNAL VALIDITY INFERENCE

FINDINGS IN THE STUDY

Actual subjects
Association between hypertension and CHD observed in the actual sample of Framingham adults

Consideraciones al definir criterios de selección

1. Características socio-demográficas
2. Características de la enfermedad o exposición
3. Otras características de importancia (p.e. co-morbilidades, limitaciones fisiológicas o cognitivas, factores de riesgo, factores ambientales, entre otras)
4. Características de accesibilidad

Unidad de análisis, de muestreo y marco muestral

- Unidad de análisis:** elementos de la población objetivo de nuestra investigación.
- Unidad de muestreo:** elementos que nos permiten llegar a las unidades de análisis.
- Marco muestral:** listado de cada una de las unidades de muestreo.

Unidad de análisis, de muestreo y marco muestral

- Observación 1: No siempre la unidad de análisis se corresponde con la unidad de muestreo.
- Observación 2: No siempre es posible contar con un marco muestral.

Ejemplo de marco muestral: ENDES

Marco Muestral

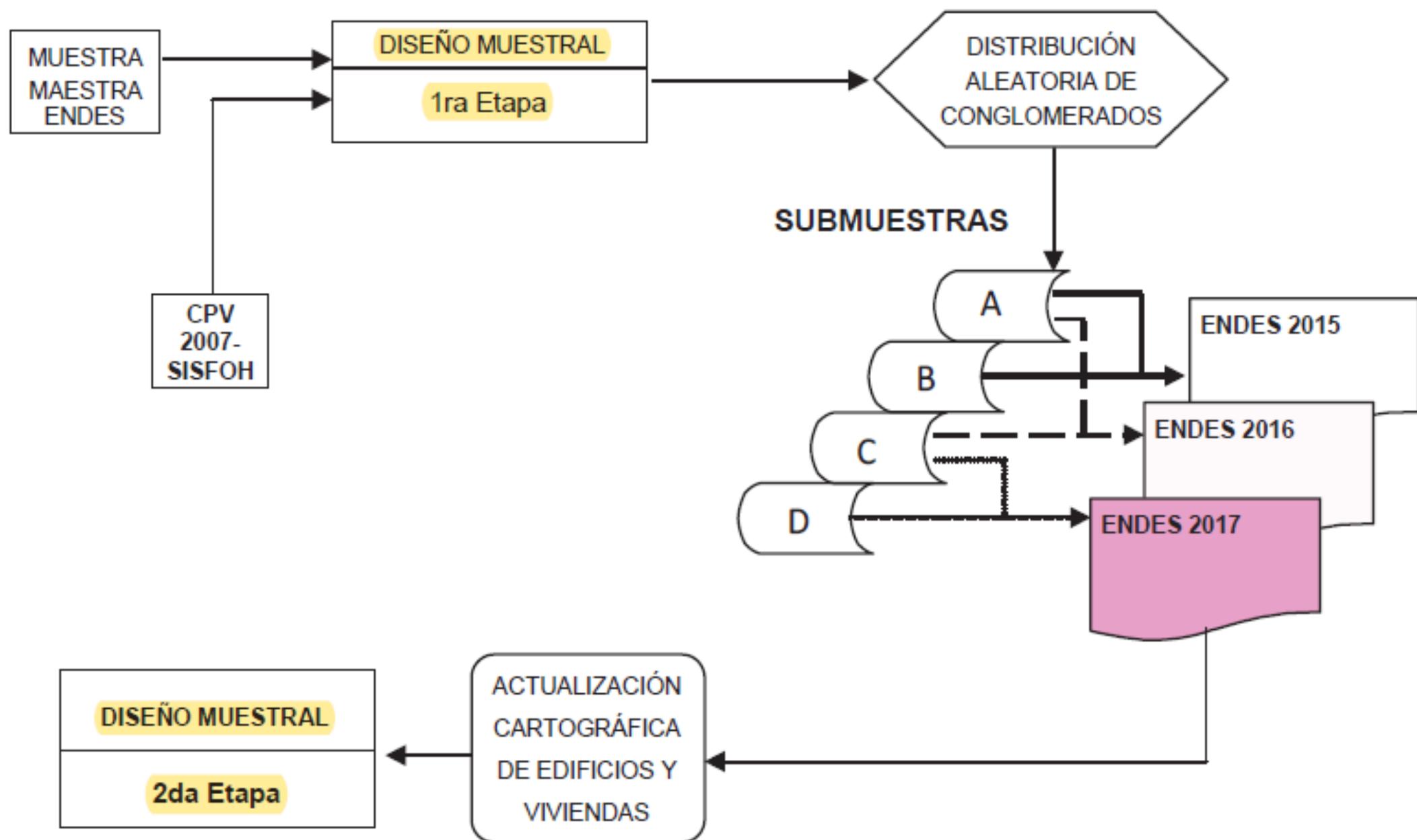
Dado que el diseño para la Encuesta ENDES Continua 2015 - 2017 es bietápica, se utiliza un marco muestral en cada una de las etapas de la selección de las unidades de muestreo.

Para la primera etapa: selección de **UPM** (conglomerados), se utiliza la información del Censo de Población y Vivienda del 2007 y la Actualización del Sistema de Focalización de Hogares -SISFOH 2012-2013.

Ejemplo de marco muestral: ENDES

Para la segunda etapa: selección de **USM** (viviendas), se utiliza el marco muestral proveniente de la actualización cartográfica y registro de edificios y viviendas realizada previamente a las entrevistas, con el objetivo de identificar y registrar cambios en las áreas seleccionadas. Producto de este proceso se obtiene un registro actualizado de viviendas y sus residentes habituales con datos de edad y sexo al momento del registro en el área de trabajo, el cual constituirá el marco de muestreo para la selección de viviendas.

Gráfico A.1 Marco Muestral y Selección de la Muestra ENDES 2015 – 2017



**PROCEDIMIENTO DE SELECCIÓN DE LA MUJER A SER ENTREVISTADA
EN LA SECCIÓN DE "VIOLENCIA DOMÉSTICA"**

1. Verifique en la portada del Cuestionario del Hogar el **último dígito del número de selección de la vivienda** y escríbalo en la siguiente casilla para determinar la **fila** a usar en el cuadro. FILA
2. Verifique **en la columna 9**, el total de mujeres entre 15-49 años que hay en el listado de hogar y escriba ese total en la siguiente casilla para determinar la **columna** a usar COLUMNNA
3. Circule en el cuadro siguiente el número en la intersección de la correspondiente **FILA Y COLUMNA**.

Ultimo dígito del Número de Selección de la Vivienda	Número de mujeres listadas en Cuestionario de Hogar							
	Una mujer	Dos mujeres	Tres mujeres	Cuatro mujeres	Cinco mujeres	Seis mujeres	Siete mujeres	Ocho mujeres
0	1	2	2	4	3	6	5	4
1	1	1	3	1	4	1	6	5
2	1	2	1	2	5	2	7	6
3	1	1	2	3	1	3	1	7
4	1	2	3	4	2	4	2	8
5	1	1	1	1	3	5	3	1
6	1	2	2	2	4	6	4	2
7	1	1	3	3	5	1	5	3
8	1	2	1	4	1	2	6	4
9	1	1	2	1	2	3	7	5

4. El número circulado le indica cuál mujer en el LISTADO DE HOGAR es elegible para las preguntas de Violencia Doméstica, es decir, si es la primera mujer listada, o la segunda, la tercera, etc. Recuerde que el número circulado en el cuadro de arriba no es equivalente al número de orden de la mujer.
5. Revise la **columna 9** del LISTADO DE HOGAR para poder determinar el **número de orden** de la mujer seleccionada, es decir, el número de orden de la primera mujer listada, o la segunda o la tercera, etc. Registre el número de orden en las siguientes casillas y ponga una "X" al lado del número de orden en la **columna 9**.

Ultimo dígito del Número de Selección de la Vivienda	Número de mujeres listadas en Cuestionario de Hogar							
	Una mujer	Dos mujeres	Tres mujeres	Cuatro mujeres	Cinco mujeres	Seis mujeres	Siete mujeres	Ocho mujeres
0	1	2	2	4	3	6	5	4
1	1	1	3	1	4	1	6	5
2	1	2	1	2	5	2	7	6
3	1	1	2	3	1	3	1	7
4	1	2	3	4	2	4	2	8
5	1	1	1	1	3	5	3	1
6	1	2	2	2	4	6	4	2
7	1	1	3	3	5	1	5	3
8	1	2	1	4	1	2	6	4
9	1	1	2	1	2	3	7	5

Selección de muestras: consideraciones

- Forma de selección
- Tamaño de la muestra

Selección de muestras: consideraciones

- Forma de selección
- Tamaño de la muestra

Van a depender de los objetivos, diseño de la investigación y características de la población en estudio.

Tipos de diseños de selección de muestras

Muestreo Probabilístico

Muestreo No Probabilístico

Tipos de diseños de selección de muestras

Muestreo Probabilístico

- Aleatorio Simple
- Sistemático
- Estratificado
- Conglomerados

Muestreo No Probabilístico

Tipos de diseños de selección de muestras

Muestreo Probabilístico

- Aleatorio Simple
- Sistemático
- Estratificado
- Conglomerados

Muestreo No Probabilístico

- Por conveniencia
- Por casos consecutivos
- Por cuota
- Bola de nieve

Muestreo aleatorio simple

Cada elemento del marco muestral tiene la misma probabilidad de ser seleccionado.

Requiere un marco muestral completo.

Al tener el tamaño de la muestra, se seleccionarán los elementos mediante cualquier procedimiento que haga la selección al azar.

Generalmente la selección se hace “sin reemplazo”.

Muestreo aleatorio simple

Ventajas:

- Todos los elementos tienen igual probabilidad de selección: n/N .
- Los cálculos matemáticos son sencillos (los programas estadísticos por lo general asumen que se trabaja con una muestra obtenida mediante este procedimiento).

Muestreo aleatorio simple

Desventajas:

- Se requiere un marco muestral completo y detallado.
- La muestra puede no ser representativa.

Muestreo sistemático

Permite obtener la muestra deseada al seleccionar aleatoriamente un elemento de los primeros k elementos en el marco muestral y después cada k -ésimo elemento hasta completar los n necesarios.

Se incorpora un criterio importante respecto al muestreo aleatorio simple: el ***orden*** establecido previamente por el investigador. Con esto se gana representatividad.

Muestreo sistemático

Ventajas:

- Más fácil de extraer una muestra.
- Se ahorra tiempo.
- La probabilidad de selección es $1/k$.
- La generación de un número aleatorio se da solo una vez.
- Se gana representatividad (respecto al aleatorio simple).

Muestreo sistemático

Desventajas:

- Posibilidad de sesgo:

No debe utilizarse cuando existe repetición cíclica o comportamiento cíclico inherente al marco muestral.

Ejemplo: Necesitamos elegir una muestra sistemática de tamaño 30 de los 120 centros de salud pertenecientes a una Dirección de Salud.

Ejemplo: Necesitamos elegir una muestra sistemática de tamaño 30 de los 120 centros de salud pertenecientes a una Dirección de Salud.

1. Elaboración del marco muestral en función de un **orden**.

No	Centro de Salud
1	La Esperanza
2	Los Girasoles
3	El Progreso
4	Solidaridad
5	San Marcos
6	San Fernando
7	La Felicidad
8	La Amistad
9	Santa María
10	La Sagrada Familia
11	El Remanso
12	Los Patriotas
13	San Pedro
14	San Pablo
15	Nueva Victoria
16	La Planicie
17	La Encalada
18	Santo Tomás
19	5 de Octubre
20	Los Próceres

Ejemplo: Necesitamos elegir una muestra sistemática de tamaño 30 de los 120 centros de salud pertenecientes a una Dirección de Salud.

1. Elaboración del marco muestral en función de un **orden**.
2. Calcular la fracción de muestreo: $n/N=1/k$

$$n/N = 30/120 = 1/4$$

el inverso k es llamado intervalo de muestreo.

No	Centro de Salud
1	La Esperanza
2	Los Girasoles
3	El Progreso
4	Solidaridad
5	San Marcos
6	San Fernando
7	La Felicidad
8	La Amistad
9	Santa María
10	La Sagrada Familia
11	El Remanso
12	Los Patriotas
13	San Pedro
14	San Pablo
15	Nueva Victoria
16	La Planicie
17	La Encalada
18	Santo Tomás
19	5 de Octubre
20	Los Próceres

Ejemplo: Necesitamos elegir una muestra sistemática de tamaño 30 de los 120 centros de salud pertenecientes a una Dirección de Salud.

1. Elaboración del marco muestral en función de un **orden**.
2. Calcular la fracción de muestreo: $n/N=1/k$

$$n/N = 30/120 = 1/4$$

el inverso k es llamado intervalo de muestreo.

3. Seleccionar el arranque aleatorio entre 1 y k . En nuestro ejemplo resultó 3.

No	Centro de Salud
1	La Esperanza
2	Los Girasoles
3	El Progreso
4	Solidaridad
5	San Marcos
6	San Fernando
7	La Felicidad
8	La Amistad
9	Santa María
10	La Sagrada Familia
11	El Remanso
12	Los Patriotas
13	San Pedro
14	San Pablo
15	Nueva Victoria
16	La Planicie
17	La Encalada
18	Santo Tomás
19	5 de Octubre
20	Los Próceres

Ejemplo: Necesitamos elegir una muestra sistemática de tamaño 30 de los 120 centros de salud pertenecientes a una Dirección de Salud.

1. Elaboración del marco muestral en función de un **orden**.
2. Calcular la fracción de muestreo: $n/N=1/k$

$$n/N = 30/120 = 1/4$$

el inverso k es llamado intervalo de muestreo.

3. Seleccionar el arranque aleatorio entre 1 y k . En nuestro ejemplo resultó 3.

No	Centro de Salud
1	La Esperanza
2	Los Girasoles
3	El Progreso
4	Solidaridad
5	San Marcos
6	San Fernando
7	La Felicidad
8	La Amistad
9	Santa María
10	La Sagrada Familia
11	El Remanso
12	Los Patriotas
13	San Pedro
14	San Pablo
15	Nueva Victoria
16	La Planicie
17	La Encalada
18	Santo Tomás
19	5 de Octubre
20	Los Próceres

Ejemplo: Necesitamos elegir una muestra sistemática de tamaño 30 de los 120 centros de salud pertenecientes a una Dirección de Salud.

1. Elaboración del marco muestral en función de un **orden**.
2. Calcular la fracción de muestreo: $n/N=1/k$

$$n/N = 30/120 = 1/4$$

el inverso k es llamado intervalo de muestreo.

3. Seleccionar el arranque aleatorio entre 1 y k . En nuestro ejemplo resultó 3.
4. Seleccionar las unidades a partir del arranque aleatorio, y cada k -ésimo lugar, hasta completar el tamaño de la muestra requerido.

No	Centro de Salud
1	La Esperanza
2	Los Girasoles
3	El Progreso
4	Solidaridad
5	San Marcos
6	San Fernando
7	La Felicidad
8	La Amistad
9	Santa María
10	La Sagrada Familia
11	El Remanso
12	Los Patriotas
13	San Pedro
14	San Pablo
15	Nueva Victoria
16	La Planicie
17	La Encalada
18	Santo Tomás
19	5 de Octubre
20	Los Próceres

Ejemplo: Necesitamos elegir una muestra sistemática de tamaño 30 de los 120 centros de salud pertenecientes a una Dirección de Salud.

1. Elaboración del marco muestral en función de un **orden**.
2. Calcular la fracción de muestreo: $n/N=1/k$

$$n/N = 30/120 = 1/4$$

el inverso k es llamado intervalo de muestreo.

3. Seleccionar el arranque aleatorio entre 1 y k . En nuestro ejemplo resultó 3.
4. Seleccionar las unidades a partir del arranque aleatorio, y cada k -ésimo lugar, hasta completar el tamaño de la muestra requerido.

No	Centro de Salud
1	La Esperanza
2	Los Girasoles
3	El Progreso
4	Solidaridad
5	San Marcos
6	San Fernando
7	La Felicidad
8	La Amistad
9	Santa María
10	La Sagrada Familia
11	El Remanso
12	Los Patriotas
13	San Pedro
14	San Pablo
15	Nueva Victoria
16	La Planicie
17	La Encalada
18	Santo Tomás
19	5 de Octubre
20	Los Próceres

Ejemplo: Necesitamos elegir una muestra sistemática de tamaño 30 de los 120 centros de salud pertenecientes a una Dirección de Salud.

1. Elaboración del marco muestral en función de un **orden**.
2. Calcular la fracción de muestreo: $n/N=1/k$

$$n/N = 30/120 = 1/4$$

el inverso k es llamado intervalo de muestreo.

3. Seleccionar el arranque aleatorio entre 1 y k . En nuestro ejemplo resultó 3.
4. Seleccionar las unidades a partir del arranque aleatorio, y cada k -ésimo lugar, hasta completar el tamaño de la muestra requerido.

No	Centro de Salud
1	La Esperanza
2	Los Girasoles
3	El Progreso
4	Solidaridad
5	San Marcos
6	San Fernando
7	La Felicidad
8	La Amistad
9	Santa María
10	La Sagrada Familia
11	El Remanso
12	Los Patriotas
13	San Pedro
14	San Pablo
15	Nueva Victoria
16	La Planicie
17	La Encalada
18	Santo Tomás
19	5 de Octubre
20	Los Próceres

Ejemplo: Necesitamos elegir una muestra sistemática de tamaño 30 de los 120 centros de salud pertenecientes a una Dirección de Salud.

1. Elaboración del marco muestral en función de un **orden**.
2. Calcular la fracción de muestreo: $n/N=1/k$

$$n/N = 30/120 = 1/4$$

el inverso k es llamado intervalo de muestreo.

3. Seleccionar el arranque aleatorio entre 1 y k . En nuestro ejemplo resultó 3.
4. Seleccionar las unidades a partir del arranque aleatorio, y cada k -ésimo lugar, hasta completar el tamaño de la muestra requerido.

No	Centro de Salud
1	La Esperanza
2	Los Girasoles
3	El Progreso
4	Solidaridad
5	San Marcos
6	San Fernando
7	La Felicidad
8	La Amistad
9	Santa María
10	La Sagrada Familia
11	El Remanso
12	Los Patriotas
13	San Pedro
14	San Pablo
15	Nueva Victoria
16	La Planicie
17	La Encalada
18	Santo Tomás
19	5 de Octubre
20	Los Próceres

Ejemplo: Necesitamos elegir una muestra sistemática de tamaño 30 de los 120 centros de salud pertenecientes a una Dirección de Salud.

1. Elaboración del marco muestral en función de un **orden**.
2. Calcular la fracción de muestreo: $n/N=1/k$

$$n/N = 30/120 = 1/4$$

el inverso k es llamado intervalo de muestreo.

3. Seleccionar el arranque aleatorio entre 1 y k . En nuestro ejemplo resultó 3.
4. Seleccionar las unidades a partir del arranque aleatorio, y cada k -ésimo lugar, hasta completar el tamaño de la muestra requerido.

No	Centro de Salud
1	La Esperanza
2	Los Girasoles
3	El Progreso
4	Solidaridad
5	San Marcos
6	San Fernando
7	La Felicidad
8	La Amistad
9	Santa María
10	La Sagrada Familia
11	El Remanso
12	Los Patriotas
13	San Pedro
14	San Pablo
15	Nueva Victoria
16	La Planicie
17	La Encalada
18	Santo Tomás
19	5 de Octubre
20	Los Próceres

Muestreo estratificado

Primero se clasifica los elementos de la población en subgrupos (estratos).

En cada estrato se selecciona una muestra aleatoria simple, tomando al menos un elemento de cada estrato.

Los estratos deben ser mutuamente excluyentes y en su conjunto se corresponden a toda la población.

Los estratos pueden reflejar distintos grupos de edad, diferente grado de exposición, nivel socio-económico, etc.

Muestreo estratificado

Ventajas:

- La estimación final del parámetro puede tener un error menor al obtenido para una muestra aleatoria simple de similar tamaño.
- Permite estimar los parámetros para cada estrato y para la población total.
- Asegura la participación de todos los estratos.

Muestreo estratificado

Desventajas:

- Mayor complejidad en los cálculos.
- Se necesita un marco muestral detallado para cada estrato.
- Los criterios para conformar los estratos podrían ser operativamente difíciles (trabajo de campo).

Muestreo estratificado: afijación

La **afijación** es la distribución de la muestra en función de los diferentes estratos. Puede ser:

- **Uniforme** (muestra de igual tamaño en cada estrato)
- **Proporcional** (muestra proporcional al tamaño del estrato)
- **De Neyman** (muestra en función a la varianza y tamaño del estrato)
- **Óptima** (muestra en función al costo, varianza y tamaño del estrato)

Approximate allocation that minimizes cost for a fixed value of $V(\hat{p}_{st})$ or minimizes $V(\hat{p}_{st})$ for a fixed cost:

$$\begin{aligned}n_1 &= n \left(\frac{N_i \sqrt{p_i q_i / c_i}}{N_1 \sqrt{p_1 q_1 / c_1} + N_2 \sqrt{p_2 q_2 / c_2} + \cdots + N_L \sqrt{p_L q_L / c_L}} \right) \\ &= n \left(\frac{N_i \sqrt{p_i q_i / c_i}}{\sum_{k=1}^L N_k \sqrt{p_k q_k / c_k}} \right) \tag{5.16}\end{aligned}$$

where N_i denotes the size of the i th stratum, p_i denotes the population proportion for the i th stratum, and c_i denotes the cost of obtaining a single observation from the i th stratum.

Muestreo por conglomerados

Es la selección de grupos de unidades de estudio, o colecciones de elementos llamados conglomerados, en lugar de unidades de estudio individuales (generalmente son unidades geográficas u organizacionales).

Muestreo por conglomerados

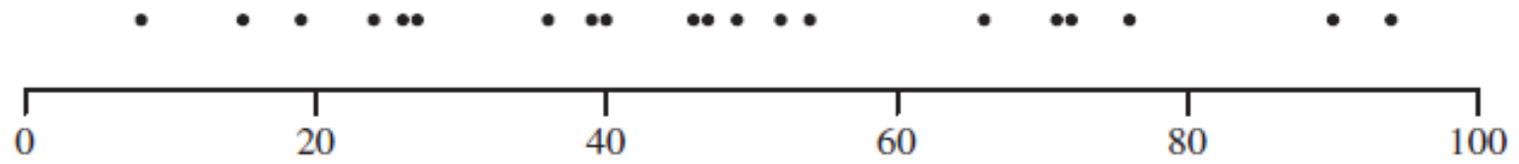
Ventajas:

- Disminución en los costos del trabajo de campo.
- Poder trabajar en ausencia de listados muy complejos.

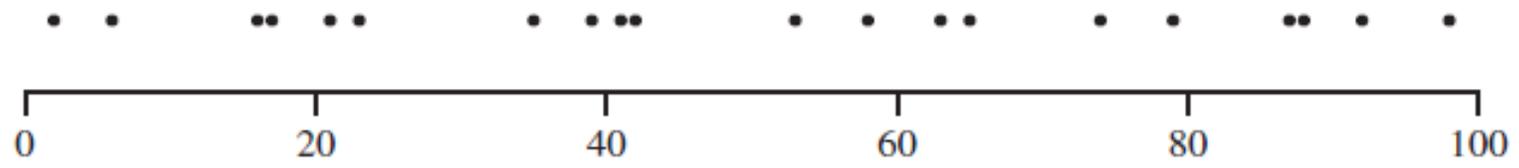
Muestreo por conglomerados

Desventajas:

- Complicación de los cálculos para la estimación de los parámetros.
- Si no se incluyen en el estudio a todos los individuos de cada conglomerado se puede generar sesgo.
- Es un método menos preciso y requiere muestras de mayor tamaño.



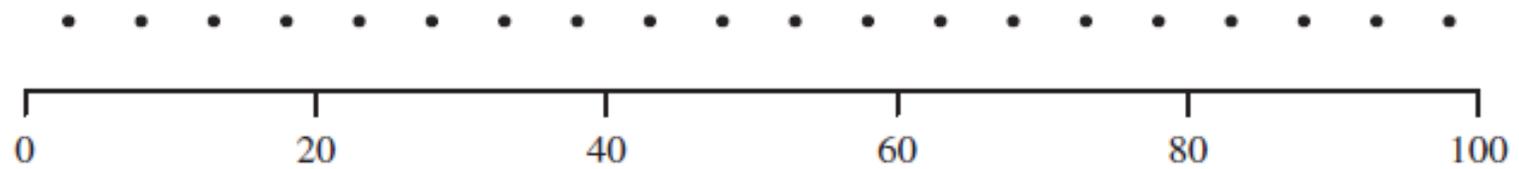
Simple random sample of 20 numbers from population of 100 numbers



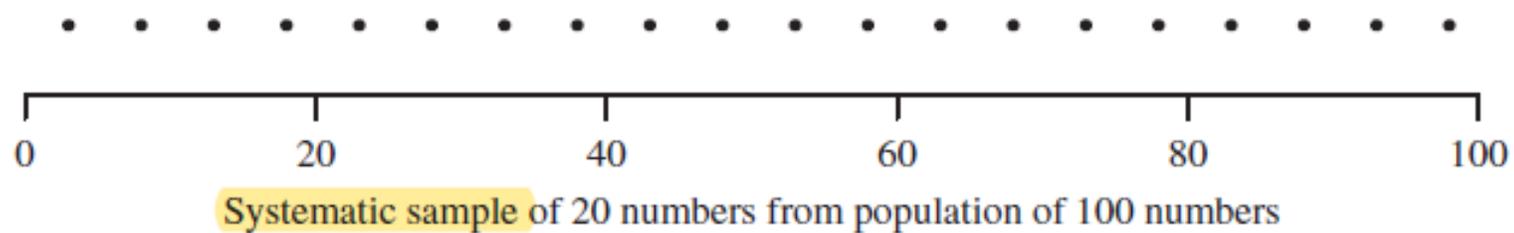
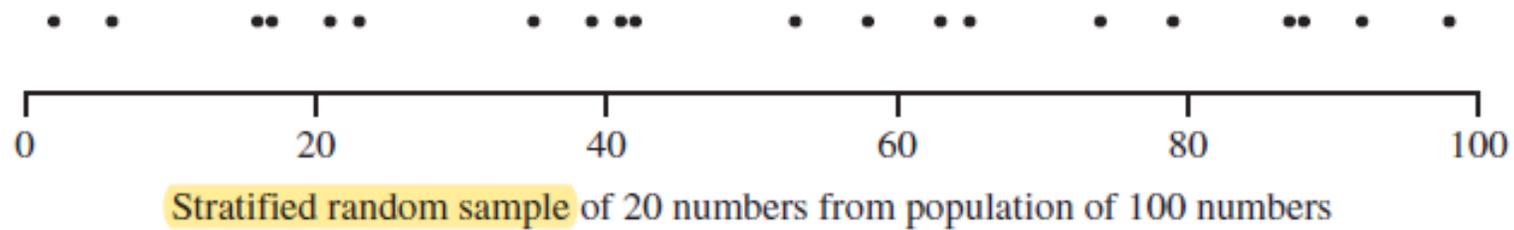
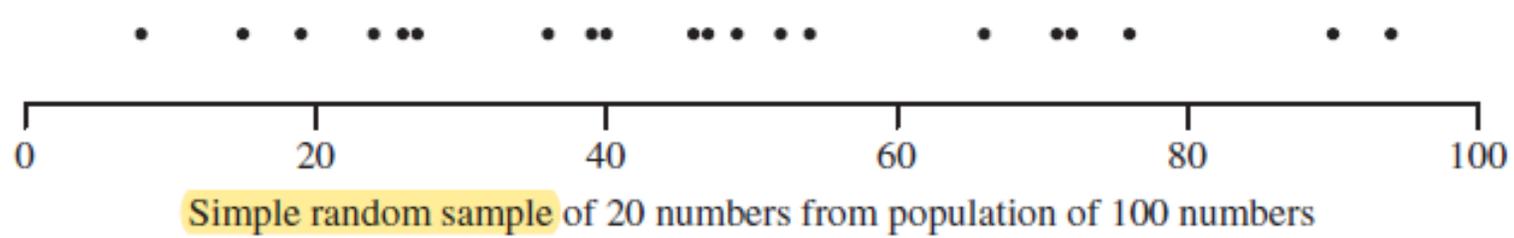
Stratified random sample of 20 numbers from population of 100 numbers

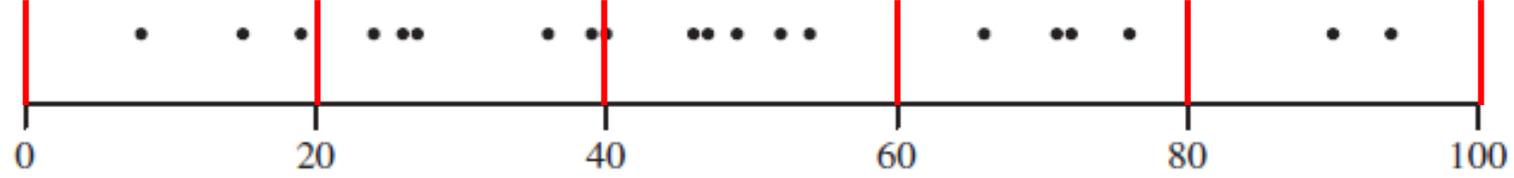


Cluster sample of 20 numbers from population of 100 numbers

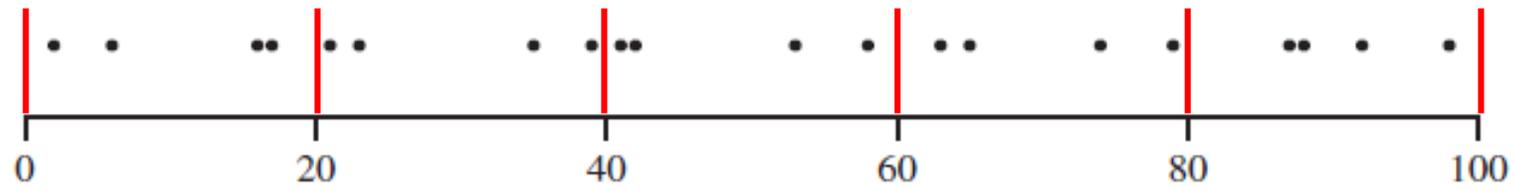


Systematic sample of 20 numbers from population of 100 numbers





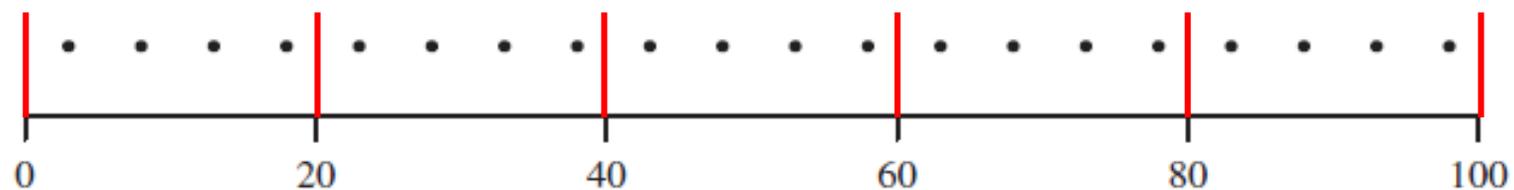
Simple random sample of 20 numbers from population of 100 numbers



Stratified random sample of 20 numbers from population of 100 numbers



Cluster sample of 20 numbers from population of 100 numbers



Systematic sample of 20 numbers from population of 100 numbers

Muestreo multi-etápico o poli-etápico

En la práctica, cuando se hace investigaciones de base poblacional donde se requiere representatividad a nivel de país o grandes divisiones administrativas, el muestreo se realizan combinando los diseños muestrales básicos descritos.

Se hace una primera división en unidades primarias de muestro (**UPM**), luego en cada UPM se divide la población en unidades secundarias de muestreo (**USM**) y de ser necesario así sucesivamente hasta llegar en una última etapa a los elementos o unidades de análisis.

Muestreo no probabilístico

Por conveniencia (a criterio): Se seleccionan a las unidades de estudio que se encuentren disponibles al momento de la recolección de datos

Por casos consecutivos: Consiste en elegir a cada paciente que cumpla con los criterios de selección dentro de un intervalo de tiempo específico o hasta alcanzar un número definido de pacientes.

Muestreo no probabilístico

Por cuota: Se seleccionan unidades de estudio de cada uno de los subgrupos que componen la población en una cuota predeterminada.

Por bola de nieve: Se selecciona un grupo inicial de entrevistados por lo general en forma aleatoria, después de la entrevista se pide a los participantes que identifiquen a otros que pertenecen a la población objetivo.

Cálculo del tamaño de la muestra

Nivel de confianza

Potencia (cuando se comparan grupos)

Varianza de la variable

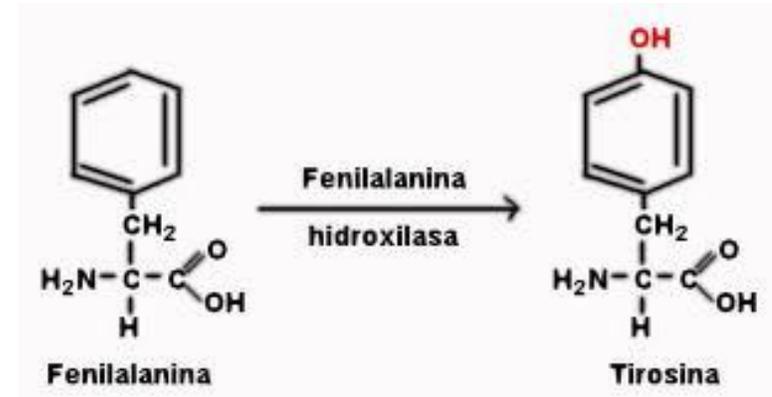
Margen de Error (en algunos textos figura ***Precisión***)

$$n = \frac{(\text{Nivel de Confianza} + \text{Potencia}) \times \text{Varianza}}{\text{Margen de Error}}$$

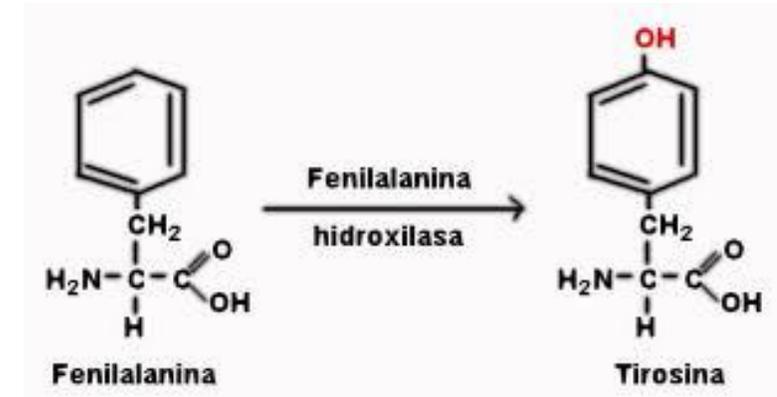


Si usted va de compras al mercado y quiere comprar uvas, ¿cuántas uvas toma para probar?

Si desea determinar la frecuencia de fenilcetonuria en recién nacidos, ¿necesitará una muestra muy grande o no muy grande?



Si desea determinar la frecuencia de fenilcetonuria en recién nacidos, ¿necesitará una muestra muy grande o no muy grande?



Si desea determinar la frecuencia de bajo peso para la edad gestacional en recién nacidos, ¿necesitará una muestra muy grande o no muy grande?



La lógica perplejidad en este momento surge de la pregunta sobre cómo obtener p ($p = 0,25$ en el ejemplo) si el estudio se inicia precisamente porque se desea estimar tal proporción, que es desconocida. Hay tres modos de hacerlo:

1. Buscar bibliografía y consultar estudios similares.
2. Realizar un pequeño estudio piloto con pocos sujetos para tener una idea sobre p .
3. Elegir el caso que necesita mayor tamaño muestral n , que es $p = q = 0,5$, ya que maximiza el producto pq . Si se supone que $p = 0,50$, a no ser que se esté en el peor de los casos, seguro que *sobra* tamaño muestral.

La lógica perplejidad en este momento surge de la pregunta sobre cómo obtener p ($p = 0,25$ en el ejemplo) si el estudio se inicia precisamente porque se desea estimar tal proporción, que es desconocida. Hay tres modos de hacerlo:

1. Buscar bibliografía y consultar estudios similares.
2. Realizar un pequeño estudio piloto con pocos sujetos para tener una idea sobre p .
3. Elegir el caso que necesita mayor tamaño muestral n , que es $p = q = 0,5$, ya que maximiza el producto pq . Si se supone que $p = 0,50$, a no ser que se esté en el peor de los casos, seguro que *sobra* tamaño muestral.

La lógica perplejidad en este momento surge de la pregunta sobre cómo obtener p ($p = 0,25$ en el ejemplo) si el estudio se inicia precisamente porque se desea estimar tal proporción, que es desconocida. Hay tres modos de hacerlo:

1. Buscar bibliografía y consultar estudios similares.
2. Realizar un pequeño estudio piloto con pocos sujetos para tener una idea sobre p .
3. Elegir el caso que necesita mayor tamaño muestral n , que es $p = q = 0,5$, ya que maximiza el producto pq . Si se supone que $p = 0,50$, a no ser que se esté en el peor de los casos, seguro que *sobra* tamaño muestral.

OJO: Situación a analizar críticamente

Cálculo del tamaño de muestras para la estimación de parámetros

Estimación de una proporción

$$n = \frac{Z^2 pq}{d^2}$$

Z = Valor Z para el nivel de confianza establecido.

p = proporción del evento de interés.

q = 1 - p

d = error absoluto.

Estimación de una media

$$n = \frac{Z^2 \sigma^2}{d^2}$$

Z = Valor Z para el nivel de confianza establecido.

σ^2 = varianza de la variable de interés.

d = error absoluto.

Ajuste del tamaño muestral en función al tamaño de la población

Sample size required to estimate μ with a bound on the error of estimation B :

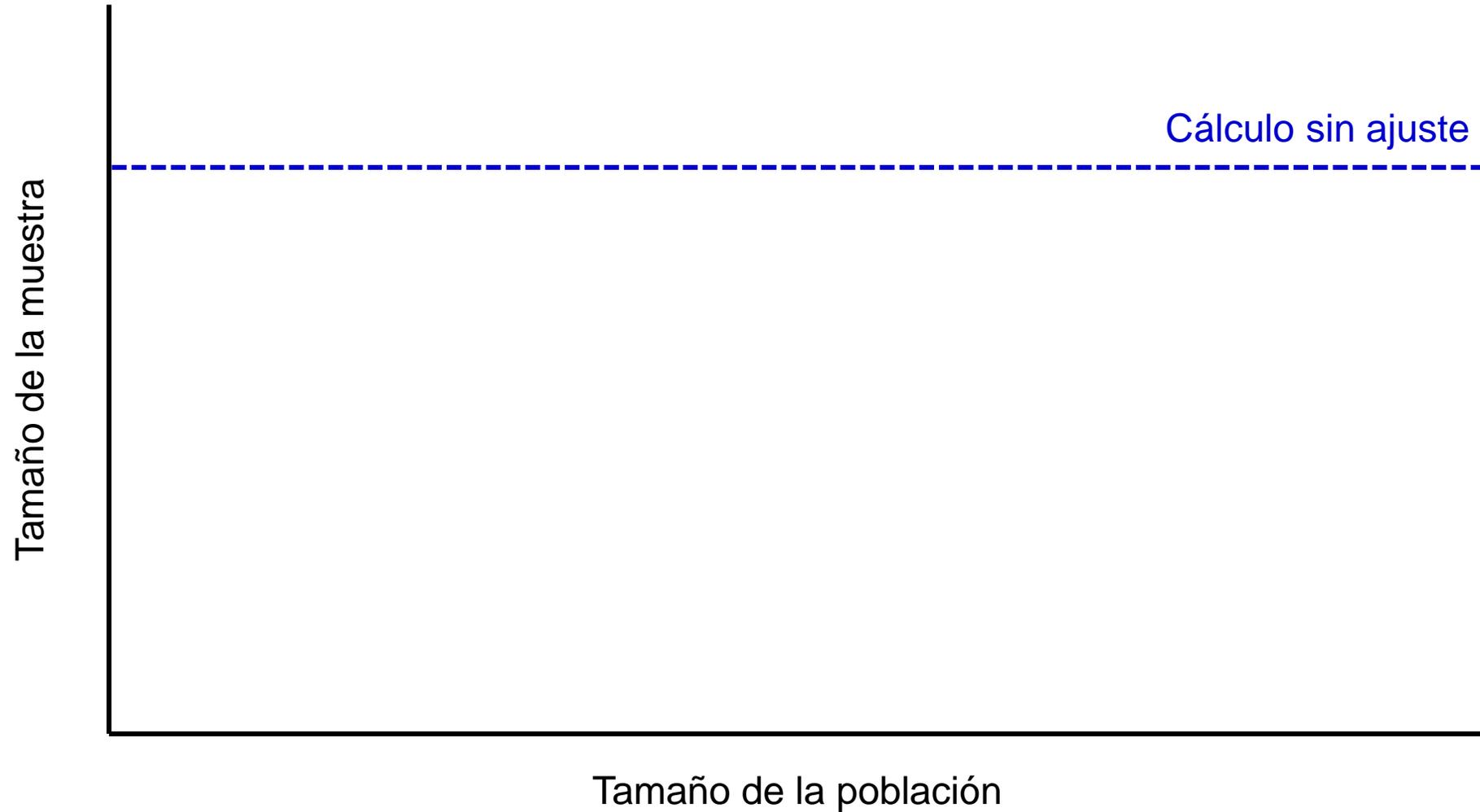
$$n = \frac{N\sigma^2}{(N - 1)D + \sigma^2} \quad (4.11)$$

where

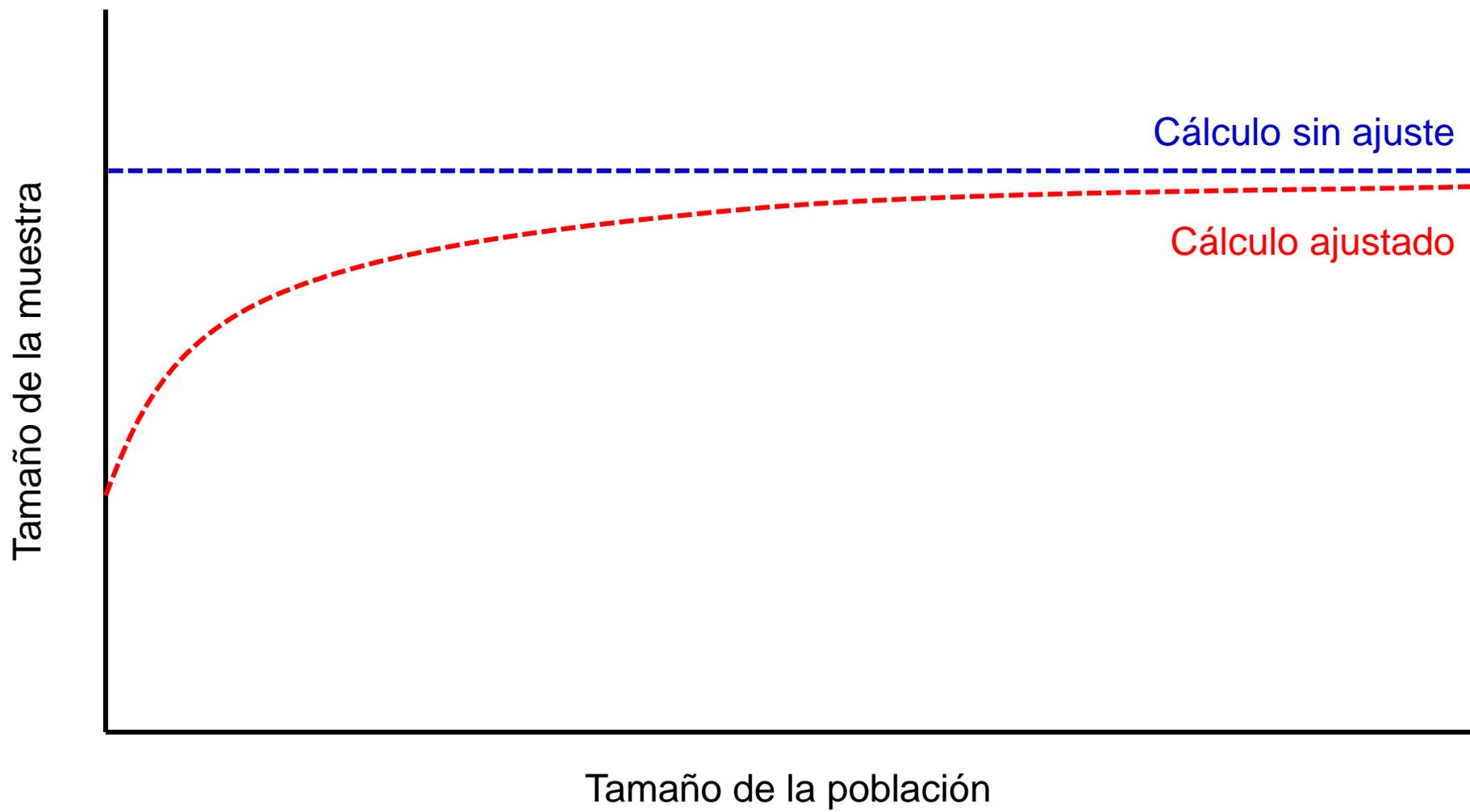
$$D = \frac{B^2}{4}$$

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

Ajuste del tamaño muestral en función al tamaño de la población



Ajuste del tamaño muestral en función al tamaño de la población



Cálculo del tamaño de muestras para pruebas de hipótesis

Comparación de proporciones (el tamaño calculado es por grupo)

$$n = \frac{(Z_{\alpha} + Z_{\beta})^2 \cdot 2[\bar{p}(1 - \bar{p})]}{\bar{d}^2}$$

Z_{α} = Valor Z para el nivel de confianza establecido.

Z_{β} = Valor Z para el nivel de poder establecido.

\bar{p} = proporción promedio entre ambos grupos.

\bar{d} = diferencia entre ambos grupos a ser detectada.

Cálculo del tamaño de muestras para pruebas de hipótesis

Comparación de medias (el tamaño calculado es por grupo)

$$n = \frac{(Z_{\alpha} + Z_{\beta})^2 \cdot 2s^2}{\bar{d}^2}$$

Z_{α} = Valor Z para el nivel de confianza establecido.

Z_{β} = Valor Z para el nivel de poder establecido.

s^2 = varianza de la variable de interés.

\bar{d} = diferencia entre ambos grupos a ser detectada.

Sample Sizes for One Group, n_A ($n_B = n_A$), in a Parallel Group Study for Different Standardised Differences and Allocation Ratios for 90% Power and a Two-Sided Type I Error of 5%

δ	Allocation Ratios			
	1	2	3	4
0.05	8,407	6,306	5,605	5,255
0.10	2,103	1,577	1,402	1,314
0.15	935	702	624	585
0.20	527	395	351	329
0.25	338	253	225	211
0.30	235	176	157	147
0.35	173	130	115	108
0.40	133	100	89	83
0.45	105	79	70	66
0.50	86	64	57	53
0.55	71	53	47	44
0.60	60	45	40	37
0.65	51	38	34	32
0.70	44	33	30	28
0.75	39	29	26	24
0.80	34	26	23	21
0.85	31	23	20	19
0.90	27	21	18	17
0.95	25	19	17	15
1.00	23	17	15	14

Ventajas del muestreo (frente a un censo)

- Reducción de costos y mayor rapidez
- Mayor flexibilidad y mayores posibilidades de estudio
- Mayor control de calidad
- El muestreo *puede* ser más exacto

Principios fundamentales que debe cumplir toda muestra

Simplicidad. Los elementos que pertenecen a la población deben estar definidos en forma clara, simple y precisa.

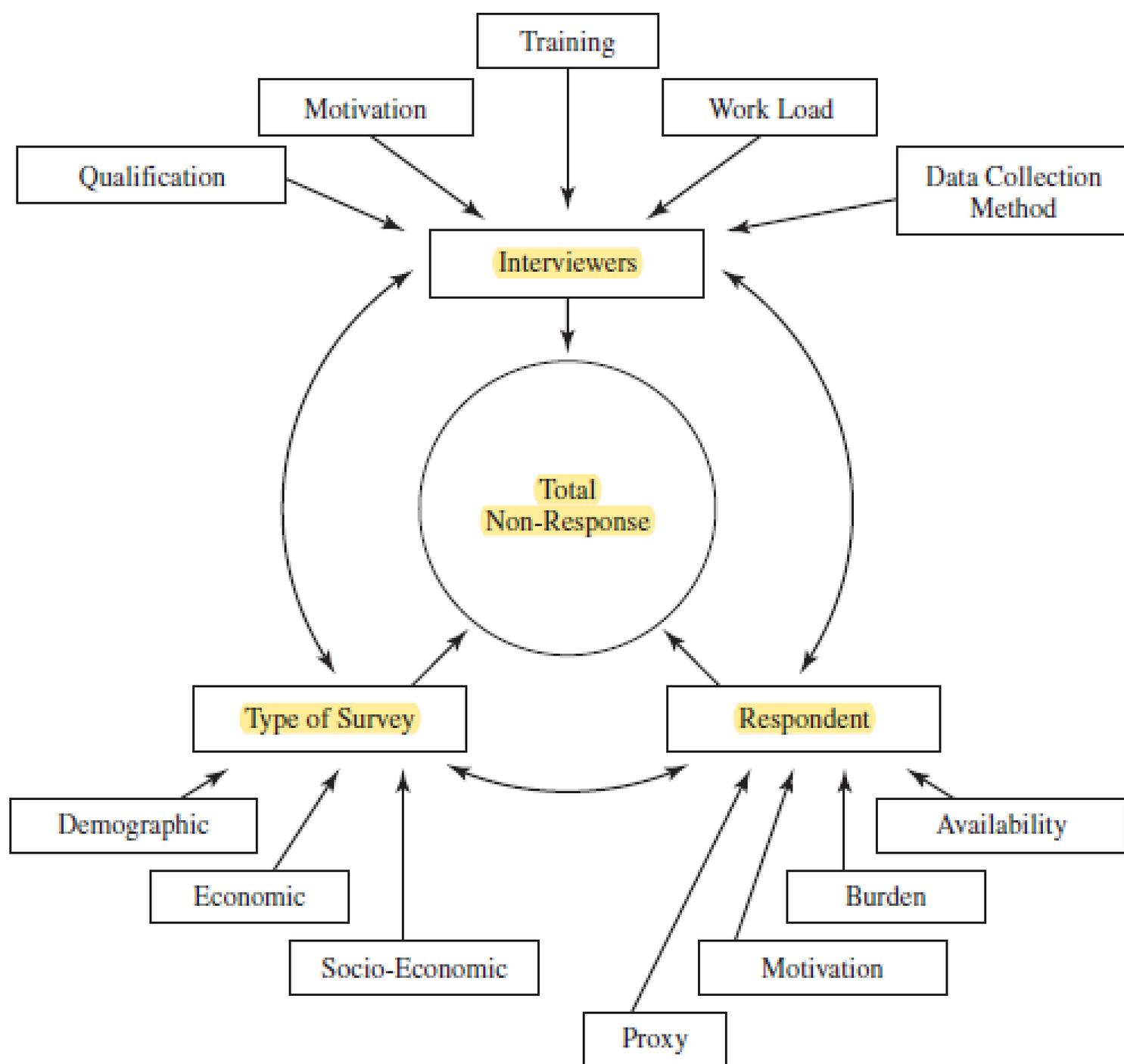
Representatividad. Las características relevantes de la población deben encontrarse en la muestra.

Comparabilidad. Cuando en un estudio es necesario contar con dos muestras entre las cuales se desea establecer alguna diferencia o asociación, es necesario que ambas no conduzcan a hallazgos sesgados (sesgo de selección).

Fuentes de sesgo en el muestreo

- No respuesta.
- Estudio solamente de voluntarios.
- Muestreo sólo de los registrados.
- Pérdida de casos de corta duración.
- Sesgo estacional.
- Accesibilidad de áreas seleccionadas.

FIGURE 8.2
Factors affecting nonresponse



POINT COUNTERPOINT

Why representativeness should be avoided

Kenneth J Rothman,^{1,2} John EJ Gallacher³ and Elizabeth E Hatch¹

¹Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA, ²RTI Health Solutions, RTI International, Research Triangle Park, NC, USA and ³Institute of Primary Care and Public Health, Cardiff University, Cardiff, UK

Accepted 21 November 2012

The essence of knowledge is generalisation. That rubbing wood in a certain way can produce fire is a knowledge derived by generalisation from individual experiences; the statement means that rubbing wood in this way will always produce fire. The art of discovery is therefore the art of correct generalisation. What is irrelevant, such as the particular shape or size of the piece of wood used, is to be excluded from the generalisation; what is relevant, for ex-

characteristics and environment of the hamsters, inferences can be drawn that may generalize to people.

A simple view of generalization casts it as a process of constructing a correct statement about the way nature works. That process is uncertain, along with everything else in empirical science, but it is not an extrapolation from sample to target population. When Pasteur created the experiment that refuted the



- Análisis descriptivo ▶
- Muestreo ▶**
- Inferencia sobre parámetros ▶
- Concordancia y consistencia ▶
- Ajuste de tasas ▶
- Demografía ▶
- Estimación de la mortalidad atribuida ▶
- Regresión logística ▶
- Distribuciones de probabilidad ▶
- Análisis bayesiano ▶
- Índices de desarrollo o privación ▶
- Medición de desigualdades en salud ▶

- Cálculo de tamaños de muestra ▶
- Selección de muestras ▶**
- Asignación de sujetos a tratamientos ▶
- Estimación con muestras complejas ▶

- Intervalos de confianza ▶
- Contraste de hipótesis ▶**

Índice de cálculos



- Análisis descriptivo ▶
- Muestreo** ▶
- Inferencia sobre parámetros ▶
- Concordancia y consistencia ▶
- Ajuste de tasas ▶
- Demografía ▶
- Estimación de la mortalidad atribuida ▶
- Regresión logística ▶
- Distribuciones de probabilidad ▶
- Análisis bayesiano ▶
- Índices de desarrollo o privación ▶
- Medición de desigualdades en salud ▶

- Cálculo de tamaños de muestra ▶
- Selección de muestras ▶
- Asignación de sujetos a tratamientos ▶
- Estimación con muestras complejas ▶

- Intervalos de confianza ▶
- Contraste de hipótesis ▶

- Media
- Proporción
- Odds ratio
- Riesgo relativo
- Concordancia
- Pruebas diagnósticas

Índice de cálculos



Índice de cálculos

- Análisis descriptivo ▶
- Muestreo ▶
 - Cálculo de tamaños de muestra ▶
 - Intervalos de confianza ▶
 - Contraste de hipótesis ▶
 - Comparación de medias ▶
 - Comparación de proporciones ▶
 - Estudios de casos y controles ▶
 - Estudios de cohorte
 - Estudios de equivalencia ▶
 - Pruebas diagnósticas ▶
 - Calidad de lotes
 - Supervivencia
 - Coefficiente de correlación
 - Selección de muestras ▶
 - Asignación de sujetos a tratamientos
 - Estimación con muestras complejas
- Inferencia sobre parámetros ▶
- Concordancia y consistencia ▶
- Ajuste de tasas ▶
- Demografía ▶
- Estimación de la mortalidad atribuida ▶
- Regresión logística ▶
- Distribuciones de probabilidad ▶
- Análisis bayesiano ▶
- Índices de desarrollo o privación ▶
- Medición de desigualdades en salud ▶

Muestreo en estudios observacionales

Dr. César Gutiérrez Villafuerte
cesar.gutierrez@udep.pe



UNIVERSIDAD
DE PIURA
Facultad de Medicina