

Cada uno da lo que recibe,
Y luego recibe lo que da,
Nada es más simple,
No hay otra norma:
Nada se pierde,
Todo se transforma.

Todo se transforma (Jorge Drexler, cantautor uruguayo)

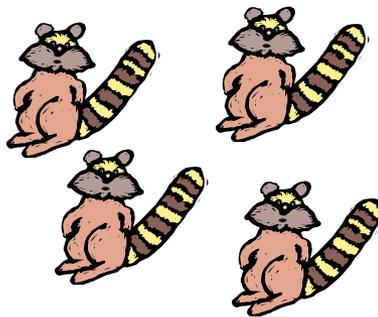
Diseño de Muestreo

Diseño de Muestreo

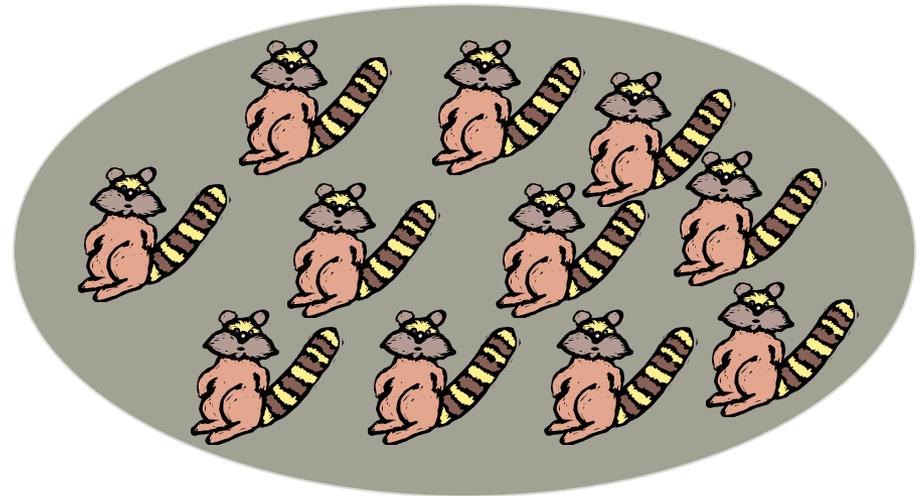
Objetivo: hacer inferencias sobre parámetros de una población basado en muestras

Las muestras obtenidas para estimar los parámetros son representativas de la población objetivo (diana)

Muestra



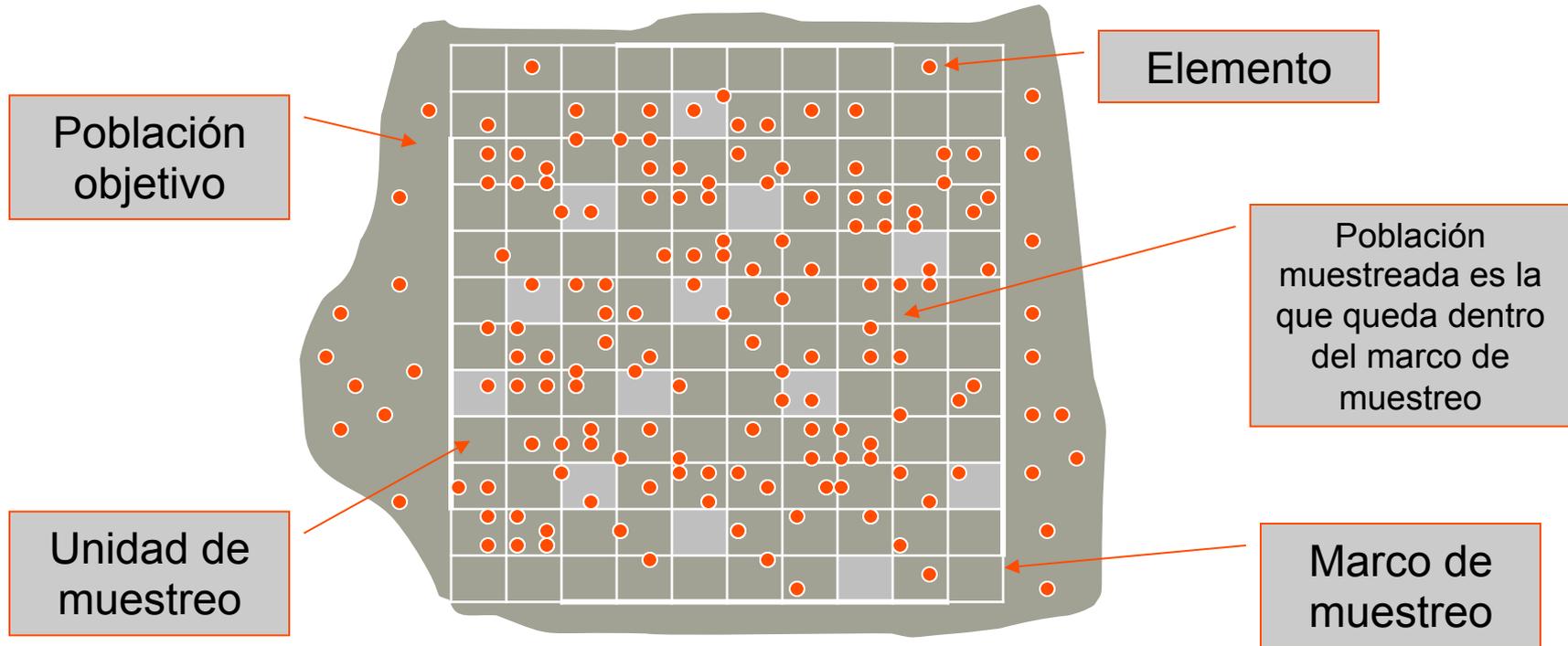
Población objetivo



2 elementos claves para obtener una muestra representativa a partir de la cual se pueden hacer inferencias:

1. Aleatorización (selección de la muestra)
2. Replicación (tamaño de muestra)

Diseño de Muestreo



Elementos

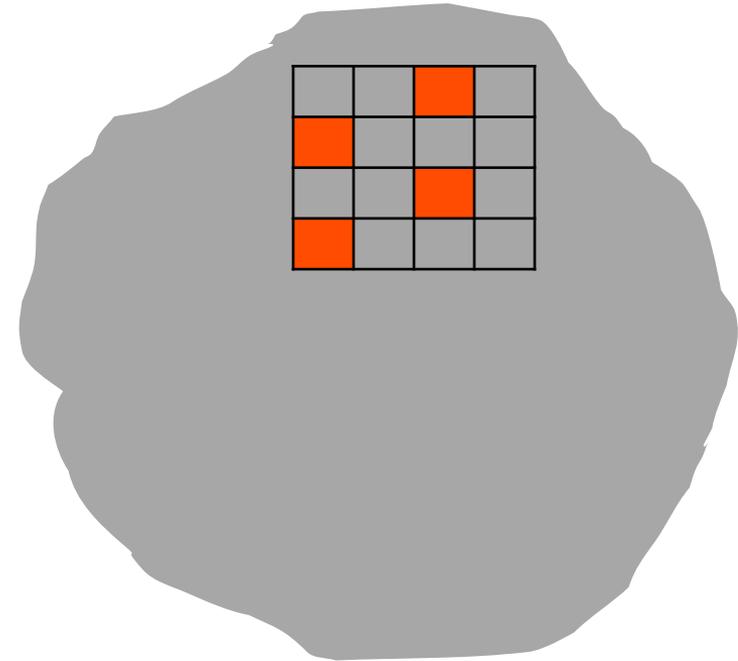
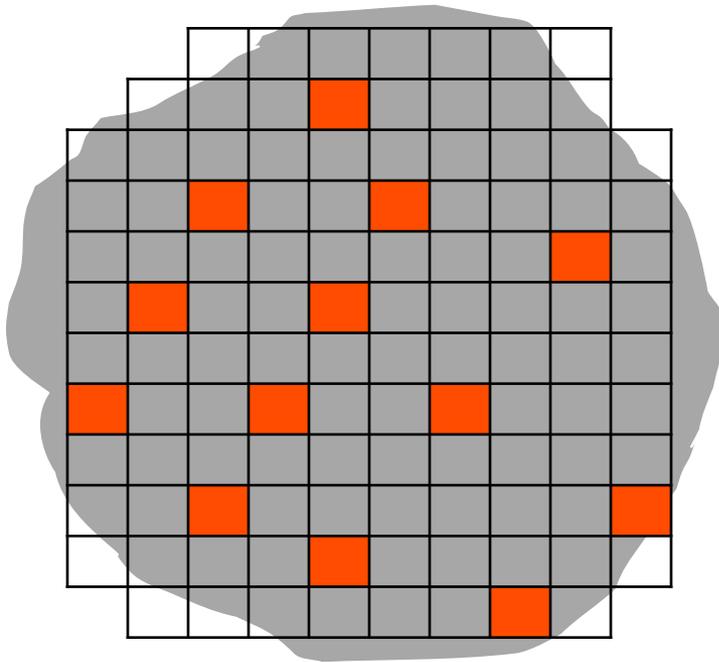
Unidad de muestreo

Marco de muestreo

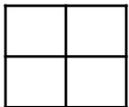
Población muestreada

Población objetivo

Población Objetivo y Población Muestreada



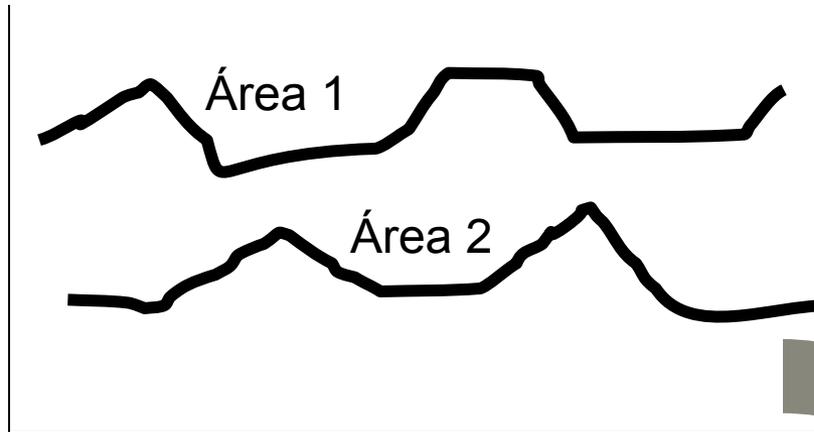
Población objetivo



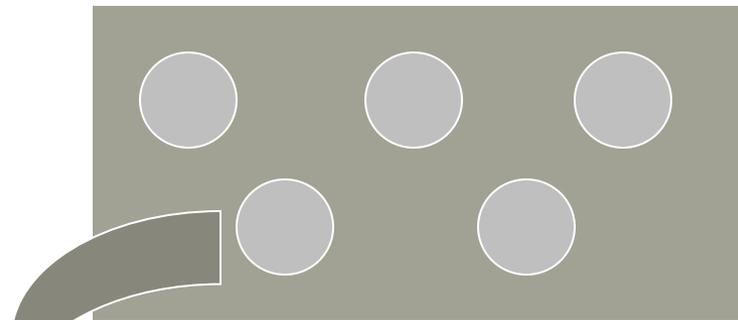
Población muestreada

Fuentes de Variación

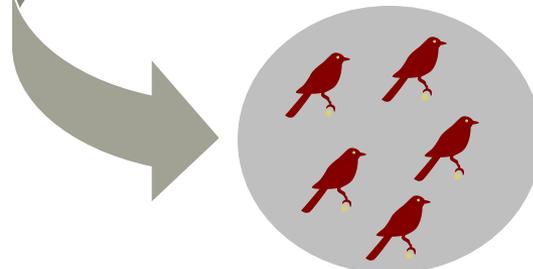
Variación espacial y temporal



Variación de muestreo



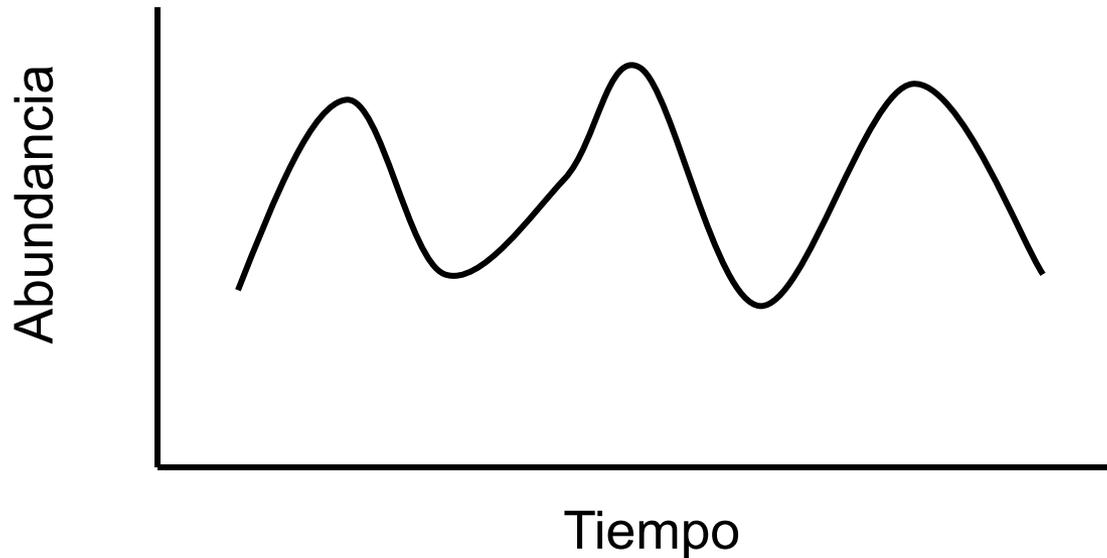
Variación entre unidades de muestreo



Variación por enumeración

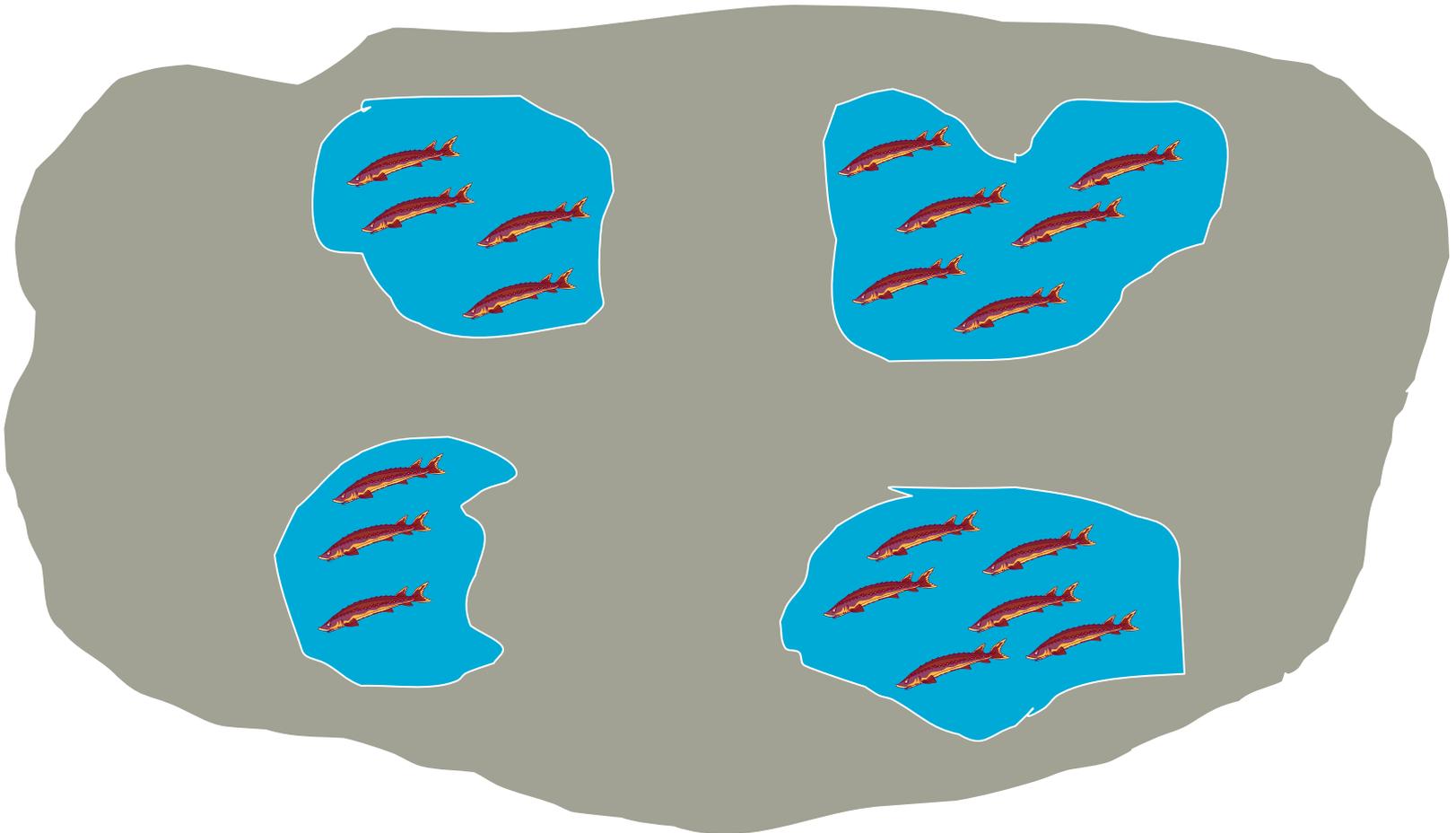
Fuentes de Variación

1. **Heterogeneidad temporal:** variabilidad que está asociada al período de tiempo sobre el cual el muestreo fue hecho.



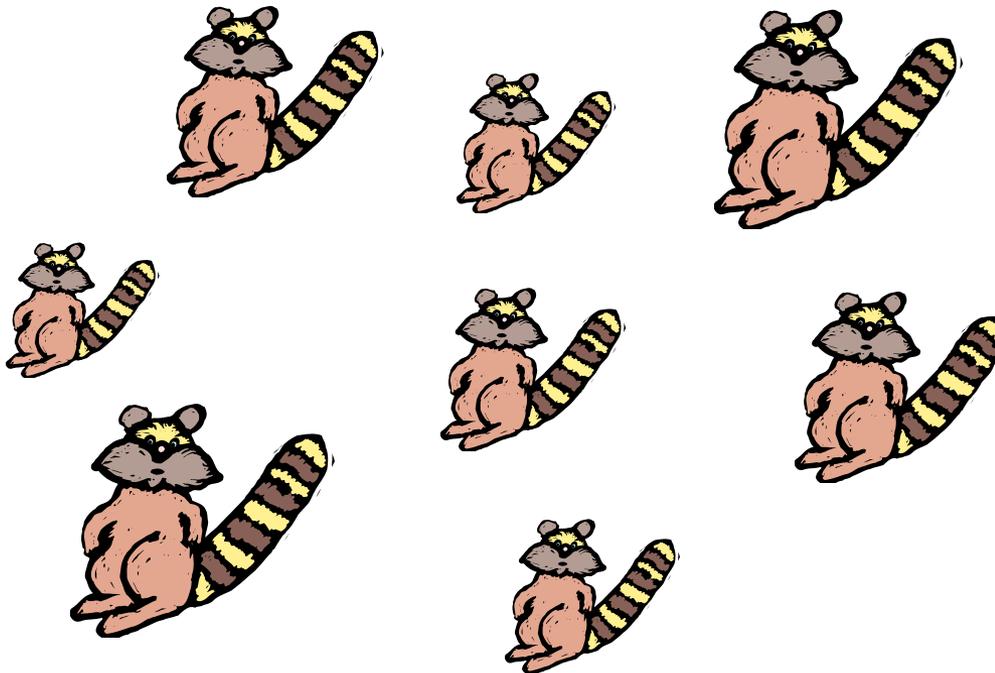
Fuentes de Variación

2. **Heterogeneidad espacial:** variabilidad que está asociada a diferencias entre áreas dentro de la población (parchosidad)



Fuentes de Variación

3. **Variabilidad de muestreo:** variabilidad que se debe al proceso de seleccionar individuos al azar.



Elementos de un Diseño de Muestreo

Replicación: la selección de múltiples muestras permite la estimación de la variación en la estimación de parámetros.

Aleatorización: la selección al azar de las unidades de muestreo de una población nos protege de una influencia sistemática de una fuente de variación no reconocida.

Control de la variación: el reducir la variación produce los siguientes beneficios:

- ✓ Aumenta la precisión de los estimados de los parámetros
- ✓ Aumenta el poder de la prueba estadística

Elementos de un Diseño de Muestreo

Control de la variación

¿Cómo puede ser reducida la variación en un diseño de muestreo?

- ❖ Estratificar para eliminar la variación sistemática
- ❖ Usar co-variables para eliminar variación no deseada entre unidades muestreo (características de la vegetación, temperatura, altitud, etc.)
- ❖ Aumentar el tamaño de muestra para aumentar la precisión del estimado.

Muestreo probabilístico vs muestreo no probabilístico

Muestreos probabilísticos

- ❖ Cada unidad de muestreo tiene una probabilidad conocida de estar en la muestra.
- ❖ La muestra es colectada con algún método consistente con esas probabilidades.
- ❖ Las probabilidades de selección son usadas cuando se realizan estimaciones a partir de la muestra.
- ❖ Se puede determinar sesgo de nuestro diseño de muestreo
- ❖ Se pueden obtener errores estándares y medidas de precisión

Ventajas

- ❖ Se puede determinar sesgo de nuestro protocolo de muestreo
- ❖ Se pueden obtener errores estándares y medidas de precisión

Muestreo probabilístico vs muestreo no probabilístico

Muestreos no probabilísticos

- ❖ Muestreo de cuotas: seleccionar 50 machos y 50 hembras.
- ❖ Muestreo de juicio: seleccionar el promedio o valores típicos. Método rápido que puede ser útil si hay unos pocos valores extremos.
- ❖ Muestreo por conveniencia: selecciona lo que está disponible o accesible.
- ❖ Muestreo sin criterio (haphazard): Esto es útil si el material de muestreo es homogéneo y diseminado a través de la población (químicos en agua potable).

Desventajas

- ❖ No es posible evaluar sesgo de manera racional.
- ❖ No se pueden estimar medida de precisión. El uso de las fórmulas de los muestreos probabilísticos es EQUIVOCADO.
- ❖ Los expertos pueden estar en desacuerdo cual es la “mejor” muestra.

Métodos de muestreo

- ❖ Muestreo Aleatorio simple
- ❖ Muestreo Aleatorio estratificado
- ❖ Muestreo Sistemático
- ❖ Muestreo por Conglomerados
- ❖ Muestreo Adaptativo por Conglomerados

A estos métodos se les llaman muestreos o métodos probabilísticos

Muestreo Aleatorio Simple

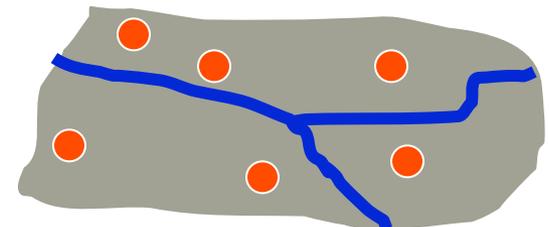
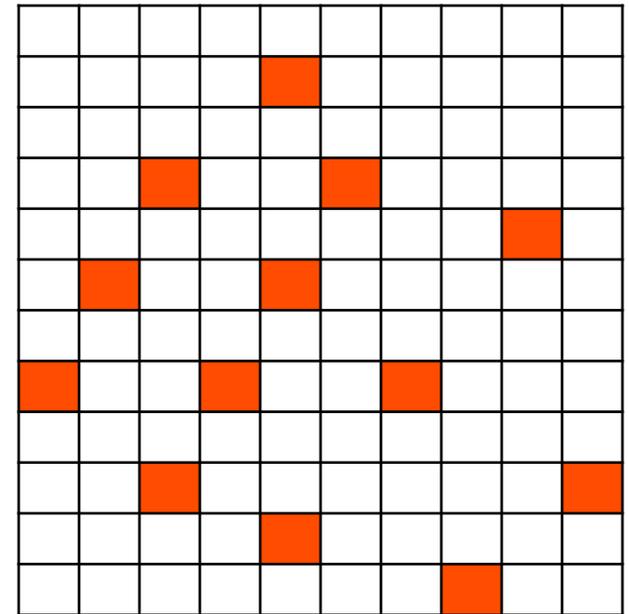
- Cuando se muestrea una población de tamaño finito, n unidades muestrales son seleccionadas al azar de una lista total N de unidades en la población muestreada.
- Con el Muestreo Aleatorio Simple cada unidad muestral tiene igual probabilidad de ser seleccionada

Ventajas

1. Estimado del parámetro y error estándar sin sesgo.
2. Extrapolar los resultados a la población es justificado.

Desventajas

1. El método es a menudo lógicamente difícil y costoso.
2. Puede que falte información sobre la población para muestrearla al azar (¿Cómo muestrear al azar una población de tapires?)



Muestreo Aleatorio Simple

Para una población con N unidades de muestreo, cada una con un atributo y_i , un estimador no sesgado de la media poblacional es la media muestral:

$$\bar{y} = \sum_{i=1}^n y_i / n$$

Si el muestreo es con reemplazo, un estimador no sesgado de la varianza de la media muestral es:

$$\hat{\text{var}}(\bar{y}) = s^2 / n \quad \text{con} \quad s^2 = \sum_{i=1}^n (y_i - \bar{y})^2 / (n - 1)$$

Si el muestreo es sin reemplazo:

$$\hat{\text{var}}(\bar{y}) = \frac{s^2}{n} \left(1 - \frac{n}{N} \right)$$

Donde $(1-n/N)$ es el factor de corrección de la población finita lo que resulta en una mayor precisión de \hat{Y}

Muestreo Aleatorio Simple

La media muestral puede ser usada para estimar el total de la población Y:

$$\hat{Y} = N\bar{y}$$

Con varianza:

$$\text{var } \hat{Y} = N^2 \text{var}(\bar{y})$$

Muestreo Aleatorio Simple

Ejemplo

Se realizaron conteos en unidades de muestreo cuadradas para estimar la abundancia de nidos de una especie de ave costera en un área de estudio de 10 ha.

El área fue dividida en 100 parcelas 0.1-ha, y 20 cuadros fueron seleccionados al azar.

El resultado de las parcelas muestreadas fue de una media muestral = 6 nidos y una varianza = 12.

$$\text{var } \hat{Y} = N^2 \text{var}(\bar{y})$$

$$\text{var } \hat{Y} = N^2 \frac{s^2}{n} \left(1 - \frac{n}{N}\right)$$

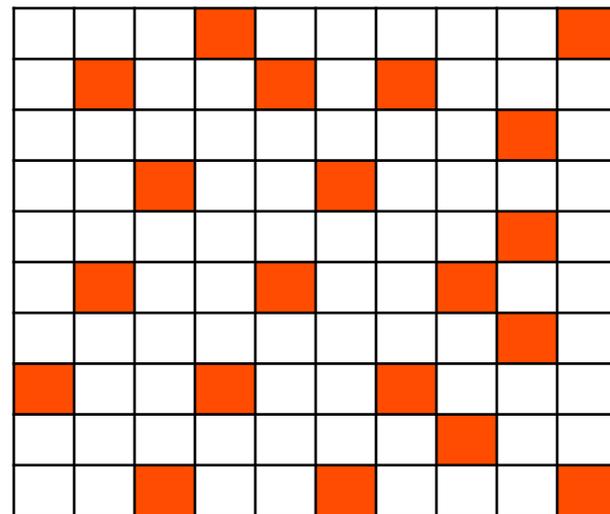
$$\text{var } \hat{Y} = 100^2 \times \frac{12}{20} \times 1 - \frac{20}{100}$$

$$\text{var } \hat{Y} = 4800$$

$$\hat{Y} = N\bar{y}$$

$$\hat{Y} = 100(6)$$

$$\hat{Y} = 600$$



Muestreo Aleatorio Simple

Determinación del tamaño de muestra

El tamaño de muestra necesario para asegurar que la media muestral este dentro de r unidades de la verdadera media, para un nivel de confianza de $1-\alpha$ (alfa) es:

$$n = \left(\frac{z_{\alpha/2} CV}{r} \right)^2 / \left(1 + \frac{n}{N} \right)$$

Fórmula con el ajuste de población finita

Objetivo: estimar el número medio de plantas en parcelas de 0.1 ha en un área de estudio de 91 ha. Hay 910 parcelas de donde se pueden tomar al azar un número n de parcelas. El objetivo es coleccionar un número suficiente de parcelas para estimar la media dentro del 10% del verdadero valor (alfa=0.05). Estudios previos indican a coeficiente de variación (CV) de =0.486.

$$n = \left(\frac{1.96}{0.10} 0.486 \right)^2 / \left(1 + \frac{91}{910} \right) \quad n \approx 83$$

Muestreo Aleatorio Estratificado

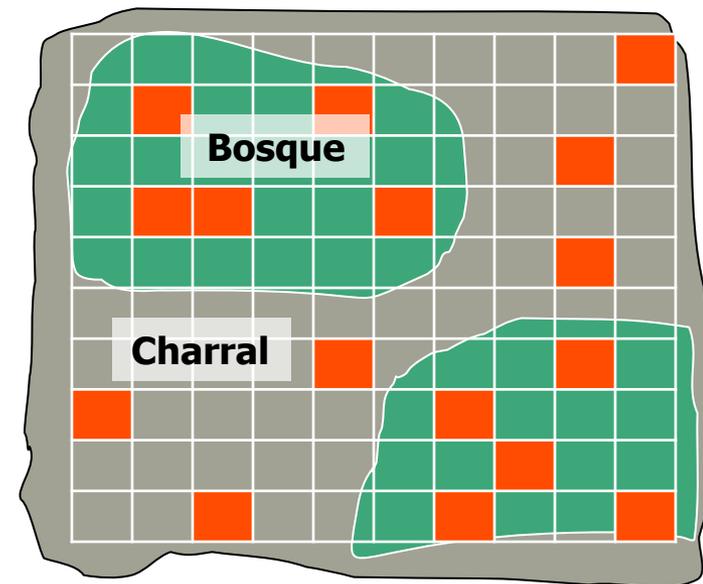
- ❖ La población puede ser dividida en subpoblaciones más o menos homogéneas.
- ❖ La población se divide en grupos o estratos de acuerdo a algún principio (ejemplo, edad, sexo, localización geográfica, morfología, tipo de hábitat) así la variación dentro del grupo es relativamente baja, y la variabilidad de la población refleja primariamente la diferencia entre grupos.

Muestreo en 2 estados

Estado 1: se divide la población en estratos (grupos)

Estado 2: se toman muestras al azar de cada estrato

Ventajas: si los estratos difieren en variabilidad, los estimados tienen mayor precisión (errores estándares menores) que los obtenidos mediante el muestreo al azar simple.



Muestreo Aleatorio Estratificado

Objetivo: identificar las unidades de muestreo dentro de la población que son internamente homogéneas con respecto al parámetro de interés.

Idealmente: uno podría identificar estratos en el cual todos sus miembros fueran idénticos. Entonces, una muestra de $n=1$ de cada estrato sería suficiente para estimar el parámetro de cada estrato (porque la varianza de cada estrato = 0).

Realísticamente: los estratos deberían ser escogidos para reflejar diferencias que se presumen existen en el parámetro de interés. A menudo, esto se realiza basado en información preliminar o información previa del sistema.

Muestreo Aleatorio Estratificado

Estimación de Parámetros

Media poblacional es: $\bar{y}_{st} = \sum_{i=1}^I W_i \bar{y}_i$ Media ponderada de las medias muestrales de cada estrato

$W_i = N_i/N$ (para cada estrato)

Varianza de la media muestral es: $\hat{\text{var}}(\bar{y}_{st}) = \frac{1}{N^2} \sum_i N_i (N_i - n_i) \frac{s_i^2}{n_i}$

Donde la varianza dentro de cada estratos es: $s_i^2 = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 / (n_i - 1)$

Total de población es: $\hat{Y} = N \bar{y}_{st}$

Muestreo Aleatorio Estratificado

Ejemplo

| Estrato (i) | N_i (ha) | N_i/N | n_i | y_i | s^2 |
|-------------------|------------|---------|-------|-------|-------|
| Bosque primario | 90 | 0.10 | 21 | 20.50 | 16.0 |
| Bosque secundario | 100 | 0.11 | 21 | 15.00 | 9.0 |
| Charral | 400 | 0.44 | 21 | 30.00 | 9.0 |
| Pastizal | 320 | 0.35 | 20 | 21.00 | 4.0 |
| Total | 910 | 1.00 | 83 | | |

$$\bar{y}_{st} = 0.10(20.5) + 0.11(15) + 0.44(30) + 0.35(21) = 24.25$$

$$\hat{\text{var}}(\bar{y}_{st}) = \frac{1}{910^2} (4731.4 + 3385.7 + 6497.4 + 19136.0) = 0.11$$

$$\hat{Y} = 910 \times 24.5 = 22295$$

Muestreo Aleatorio Estratificado

Distribución del esfuerzo de muestreo entre estratos

1. Asignación igualitaria (ejemplo anterior).
2. Asignación proporcional

n unidades de muestreo son distribuidas de acuerdo al tamaño relativo de cada estrato

$$n_i = n(N_i / N)$$

3. Asignación óptima

Tiene en consideración la varianza de cada estrato, el costo del muestreo y el tamaño del estrato.

$$n_i = n \left(\frac{N_i \hat{S}_i / \sqrt{C_i}}{\sum_{i=1}^I N_i \hat{S}_i / \sqrt{C_i}} \right)$$

Si los costos son constantes para cada estrato,

\hat{S}_i Desvío Estándar en estrato i

$$n_i = n \left(\frac{N_i \hat{S}_i}{\sum_{i=1}^I N_i \hat{S}_i} \right)$$

Muestreo Aleatorio Estratificado

Ejemplo

Área de estudio de 91 ha, compuesta de 4 estratos conteniendo 90, 100, 400, y 320 cuadros de 0.1 ha, respectivamente.

Tamaño de muestra = 83 unidades muestrales (cuadros)

Distribución Proporcional

| Estrato | N_i | N_i/N | $n_i=(N_i/N)$ |
|---------|-------|---------|---------------|
| 1 | 90 | 0.10 | 8 |
| 2 | 100 | 0.11 | 9 |
| 3 | 400 | 0.44 | 37 |
| 4 | 320 | 0.35 | 29 |
| Total | 910 | 1.00 | 83 |

Distribución óptima

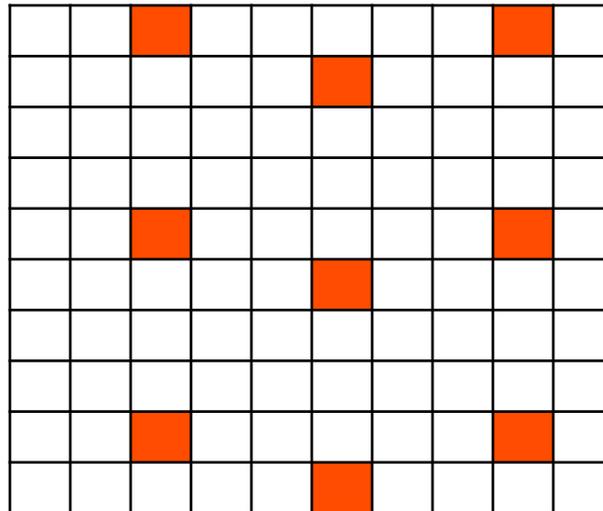
| Estrato | N_i | S_i | $S_i N_i$ | n_i |
|---------|-------|-------|-----------|-------|
| 1 | 90 | 4.00 | 360.00 | 12 |
| 2 | 100 | 3.00 | 300.00 | 10 |
| 3 | 400 | 3.00 | 1200.00 | 40 |
| 4 | 320 | 2.00 | 640.00 | 21 |
| Total | 910 | | | 83 |

Distribución basad en tamaño y desvío estándar de los estratos, los costos de los muestreo de la unidades de muestreo se suponen cosntantes

Muestreo Sistemático

Después de seleccionar aleatoriamente la primera unidad de muestreo, todas las unidades de muestreos restantes se colocan a intervalos regulares.

Usa los mismos procedimientos de estimación del muestreo aleatorio simple



Ventajas:

1. Alternativa práctica cuando el muestreo aleatorio simple es difícil de implementar
2. No requiere un marco de muestreo bien definido

Desventajas:

1. El estimado de punto es insesgado, pero el error estándar es sobre-estimado
2. El estimado de punto puede ser sesgado si existe una periodicidad o gradiente en la población

Muestreo por Conglomerados

- ❖ Es un muestreo aleatorio simple en el cual cada unidad de muestreo es un conglomerado o colección de observaciones.
- ❖ Resulta en una alta variación dentro del conglomerado en relación a la variación entre conglomerados (opuesto al muestreo aleatorio estratificado), así la varianza refleja la variación dentro de los conglomerados.
- ❖ 1) se especifican los conglomerados y se hace una lista de todos los conglomerados, 2) se toma una muestra aleatoria de la lista de conglomerados y 3) se miden todos los elementos de interés dentro de cada conglomerado.
- ❖ Los conglomerados son las unidades de muestreo primarias y los elementos dentro de conglomerados son las unidades de muestreo secundarias, pero a diferencia de un muestreo multi-estado, cada unidad de muestreo secundaria dentro de la unidad de muestreo primaria (conglomerado) debe ser incluida en la muestra.

¿Cuándo se debe usar?

Si no hay una forma efectiva de obtener una lista de unidades de muestreo secundarias.

Si el costo o tiempo para viajar de una unidad de muestreo a otro es prohibitivo.

Muestreo por Conglomerados

$$\hat{Y} = \frac{M}{m} \sum_{i=1}^m y_i = \hat{Y} = M\bar{y}_c \leftarrow \begin{array}{l} \text{Media por} \\ \text{conglomerado} \end{array}$$

M = número de conglomerados en la población

m = número de conglomerados muestreados

Cada conglomerado contiene un número N'_i de unidades secundarias dentro de cada conglomerado i , $i = 1, \dots, M$.

Si un conglomerado es seleccionado para ser muestreado, todas las unidades secundarias dentro son incluidas en la muestra.

$$\text{vâr}(\hat{Y}) = M(M - m)(s_c^2 / m)$$

$$s_c^2 = \frac{1}{m - 1} \sum_{i=1}^m (y_i - \bar{y}_c)^2$$

Muestreo por Conglomerados

Objetivo: estimar el total de huevos producidos por una especie de pato en un área que contiene $M=50$ lagunas.

Se selecciona una muestra al azar de $m=5$ lagunas y se cuenta el número de huevos en cada nido alrededor de cada laguna.

| Laguna (i) | Número de nidos (N'_i) | Número de huevos (y_i) |
|------------|----------------------------|----------------------------|
| 1 | 10 | 60 |
| 2 | 12 | 72 |
| 3 | 15 | 90 |
| 4 | 8 | 48 |
| 5 | 14 | 64 |
| Total | 59 | 334 |

$$\hat{Y} = \frac{50}{5} \sum_{i=1}^5 y_i$$

$$\hat{Y} = (50/5)(334) = 3340$$

$$\text{vâr}(\hat{Y}) = 50(50 - 5)(243.2 / 5)$$

$$\text{vâr}(\hat{Y}) = 109,440$$

$$EE(\hat{Y}) = \sqrt{109,440}$$

Muestreo Adaptativo por Conglomerados

- ❖ Como los animales frecuentemente están distribuidos de manera heterogénea (de manera parchosa), los métodos de muestreos convencionales tales como, el muestreo simple aleatorio y el muestreo aleatorio estratificado, resultan en una distribución del esfuerzo de muestreo ineficiente y estimados altamente variables (baja precisión).

Ejemplo: conteo aéreo de patos resultan en muchas unidades de muestreo con pocos o ningún pato, y algunas unidades de muestreo con miles de patos:

$\{1, 0, 0, 12000, 20, 0, 0\}$

Estimados basados en este tipo de datos tienen muy baja precisión

- ❖ Una solución a este problema es el muestreo adaptativo (hay muchos diseños adaptativos).
- ❖ Los muestreos adaptativos adjudican una probabilidad a cada posible muestra, sin embargo la selección de probabilidad en cada punto depende del valor de la unidad de muestreo seleccionada previamente.

Muestreo Adaptativo por Conglomerados

- ❖ Un muestra inicial de tamaño n es tomada al azar de una población y para cada unidad seleccionada se aplica el criterio de selección C .
- ❖ La colección de unidades incluidas en la muestra es un conglomerado, estos conglomerados contienen unidades que pueden o no cumplir con el criterio de selección C .
- ❖ El grupo de unidades de muestreo dentro de un conglomerado que reúne la condición C es llamado una red.
- ❖ Las unidades de muestreo del conglomerado que no satisface el criterio son llamados unidades de borde

| | | | | | | | | | |
|---|----|----|----|---|---|---|---|---|---|
| | | | | | | | | | |
| | 10 | | 22 | 6 | | 6 | | | |
| | 20 | | 4 | | | | | 4 | |
| | | 15 | | | 8 | | | | |
| | | | | | | | | | |
| | | | 1 | | | | 6 | | 8 |
| | | | | | | | | | 8 |
| 3 | | | | 2 | 5 | | | | |
| | 30 | | | | | | 7 | | |
| | | | 23 | | 4 | | | | |

| | | | | | | | | | |
|---|----|----|----|---|---|---|---|---|---|
| | | | | | | | | | |
| | 10 | | 22 | 6 | | 6 | | | |
| | 20 | | 4 | | | | | 4 | |
| | | 15 | | | 8 | | | | |
| | | | | | | | | | |
| | | | 1 | | | | 6 | | 8 |
| | | | | | | | | | 8 |
| 3 | | | | 2 | 5 | | | | |
| | 30 | | | | | | 7 | | |
| | | | 23 | | 4 | | | | |

Conglomerado (5 unidades muestrales)

Red de tamaño 2

Muestreo Adaptativo por Conglomerados

Población Total

$$\hat{Y}_a = \sum_{k=1}^K \frac{y_k^* \times z_k}{p_k}$$

Total de valores de y observados en la red k

Es igual a 1

N_k es el número de unidades muestrales en la red k (tamaño de la red)

Probabilidad de Inclusión para cada red

$$p_k = 1 - \frac{\binom{N-N_k}{n}}{\binom{N}{n}}$$

N es el número de unidades muestrales consideradas en el muestreo (total de cuadros)

Media poblacional por unidad de muestreo

$$\bar{y}_a = \hat{Y}_a / N$$

Número de unidades muestrales consideradas inicialmente

Muestreo Adaptativo por Conglomerados

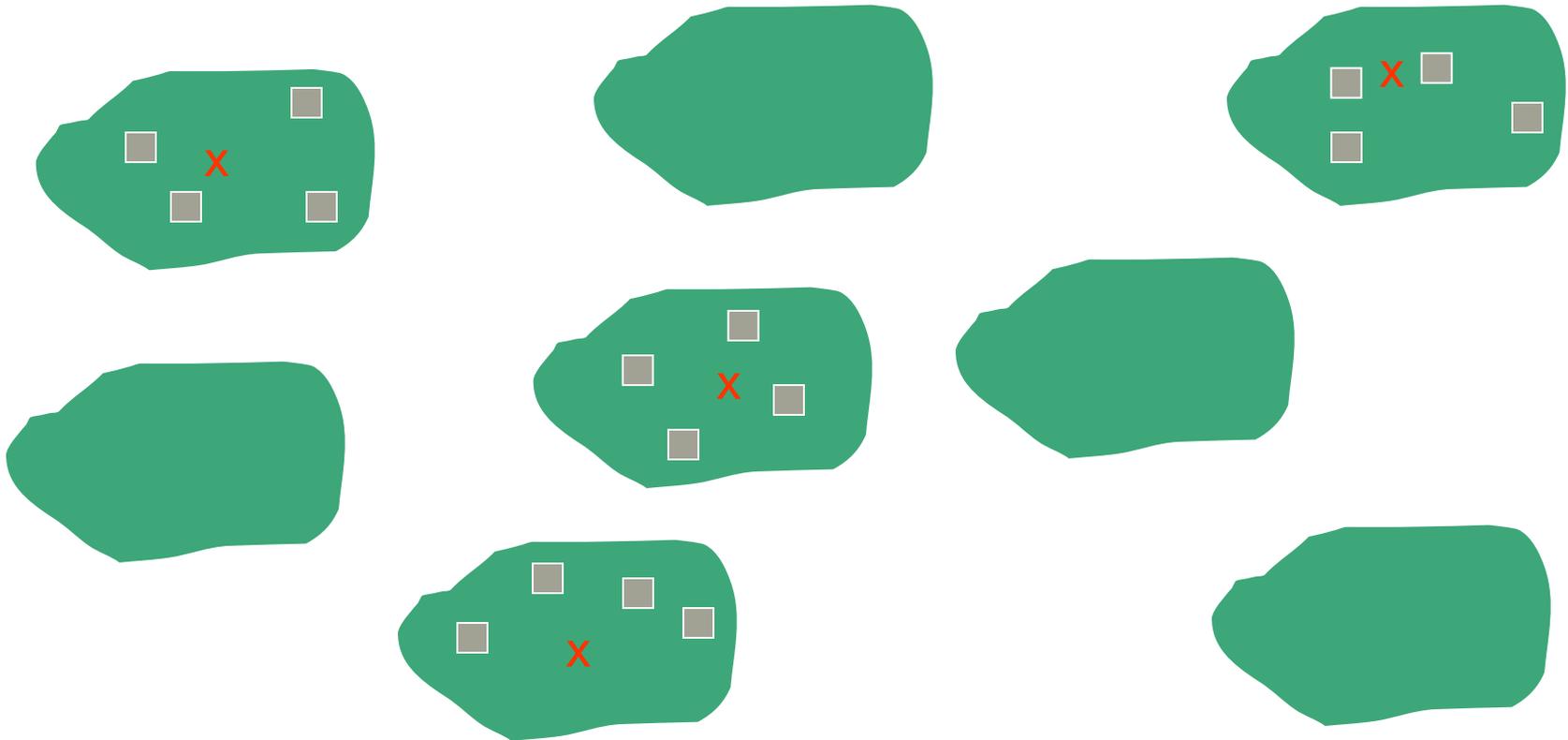
| | | | | | | | | | |
|---|----|----|----|---|---|---|---|---|---|
| | | | | | | | | | |
| | 10 | | 22 | 6 | | 6 | | | |
| | 20 | | 4 | | | | | 4 | |
| | | 15 | | | | 8 | | | |
| | | | | | | | | | |
| | | | 1 | | | | 6 | | 8 |
| | | | | | | | | | 8 |
| 3 | | | | 2 | 5 | | | | |
| | 30 | | | | | | 7 | | |
| | | | 23 | | 4 | | | | |

| | | | | | | | | | |
|---|----|----|----|---|---|---|---|---|---|
| | | | | | | | | | |
| | 10 | | 22 | 6 | | 6 | | | |
| | 20 | | 4 | | | | | 4 | |
| | | 15 | | | | 8 | | | |
| | | | | | | | | | |
| | | | 1 | | | | 6 | | 8 |
| | | | | | | | | | 8 |
| 3 | | | | 2 | 5 | | | | |
| | 30 | | | | | | 7 | | |
| | | | 23 | | 4 | | | | |

| Unidad (k) | Tamaño de red (N_k) | Conteo (y_k^*) | Probabilidad de inclusión (P_k) | $y_k^*z_k/P_k$ |
|------------------|-------------------------|--------------------|-------------------------------------|----------------|
| 1 | 1 | 0 | 0.100 | 0.000 |
| 2 | 1 | 0 | 0.100 | 0.000 |
| 3 | 3 | 32 | 0.273 | 117.015 |
| 4 | 2 | 7 | 0.191 | 36.667 |
| 5 | 1 | 0 | 0.100 | 0.000 |
| 6 | 1 | 4 | 0.100 | 40.000 |
| 7 | 1 | 0 | 0.100 | 0.000 |
| Población | | | | 193.681 |
| Media por cuadro | | | | 1.937 |

Muestreo en múltiples estados

En muchos estudios relacionados con recursos naturales se seleccionan aleatoriamente algunas unidades de muestreo (unidades primarias) y después se realiza una nueva selección aleatoria de unidades de muestreo (unidades secundarias) dentro de cada unidad primaria seleccionada.



Resumen de los diferentes métodos de muestreo

| Método | Usos recomendados | Ventajas | Desventajas |
|--|---|---|---|
| Muestreo Aleatorio Simple | Áreas geográficas pequeñas | Fórmulas simples | Difícil de ubicar las unidades cuando el área es grande |
| Muestreo Aleatorio Estratificado | Cuando hay heterogeneidad | Mejor estimación que MAS | Igual que MAS |
| Muestreo Sistemático | Las unidades deben estar relativamente lejos para alcanzar independencia | Fácil para estudios de campo. Mejor entremezcla de las unidades de muestreo que en MAS. | Si el número de unidades de muestreo es bajo (25-30) el resultado es cuestionable |
| Muestre por Conglomerados | Cuando es difícil tomar una muestra aleatoria de los elementos de interés | Menos costoso que tomar el mismo número de elementos de la población al azar | Cada elemento en el conglomerado debe ser medido |
| Muestreo Adaptativo por Conglomerados | Elementos distribuidos de manera parchosa | Distribución eficiente del esfuerzo de muestreo y mejor estimación que MAS | ? |

Problemas en los Diseños de Muestreo

1. **Falla en definir la población objetivo y la población muestreada**
 - Idealmente la población muestreada = población objetivo.
 - Diferencias moderadas están bien, pero diferencias extremas hace que la inferencias no puedan aplicarse a la población objetivo.
 - Soluciones: a) Rediseñar el muestreo; b) Usar información auxiliar (caso especial de muestreo doble); c) Redefinir la población objetivo.
2. **Falta de Replicación y cometer “Pseudoreplicación”**
 - Se debe lograr un apropiado nivel de replicación.
 - Idealmente las 3 Is: Individualidad: la muestra está compuesta de unidades de muestreo diferenciables (unidades individuales); Identidad: las unidades comparten una idéntica distribución y parámetros (o podrían ser estratificadas para que sean idénticas); Independencia: independencia estadística.
3. **Mal interpretación de patrón como Causa-Efecto**
 - Covariación entre la variable respuesta y la variable explicatoria no explica necesariamente una relación de causa-efecto (necesidad de una aproximación experimental).