



Muestreo Estadístico

Muestreo Estratificado

Stratified Sampling (SS)

Andy Domínguez
adominguez@utb.edu.co

Jul 2025

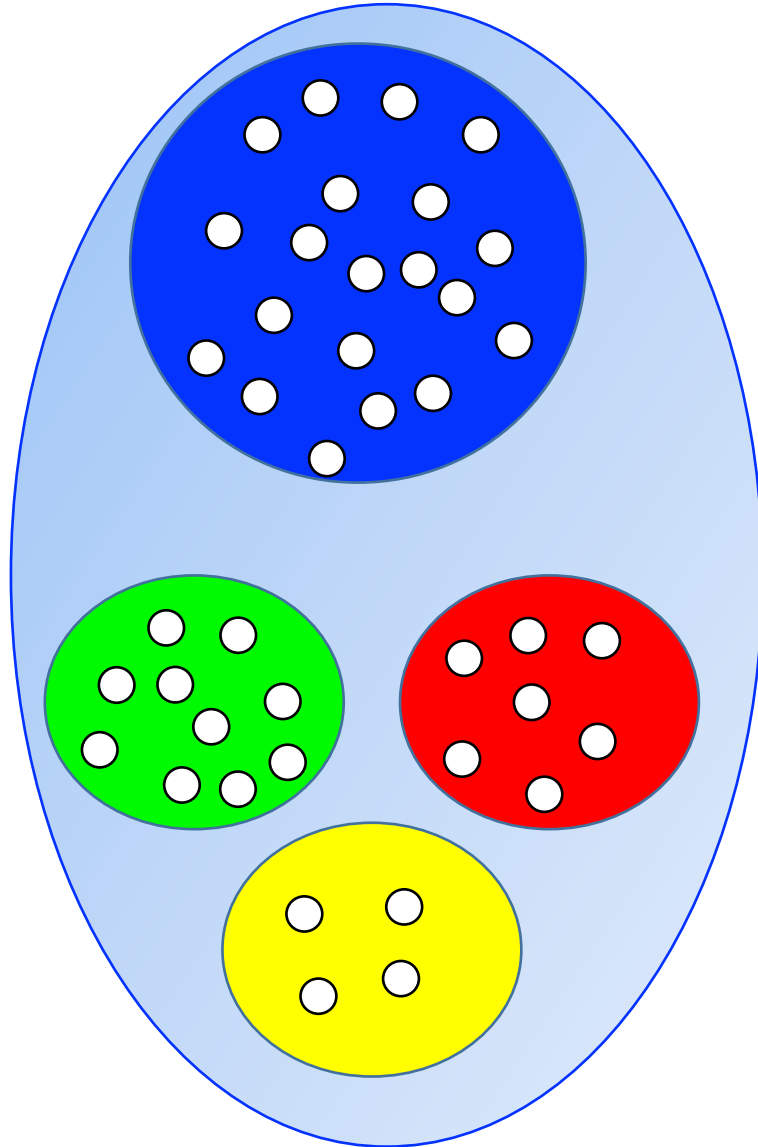


Especialización en Estadística Aplicada y Ciencia de Datos

Contenido

- Introducción
- Muestreo Estratificado -ME
- Ventajas/Limitantes del ME
- Estimación de parámetros bajo ME
- Tamaño de muestra para ME
- Práctica en Python

Muestreo Estratificado(ME)



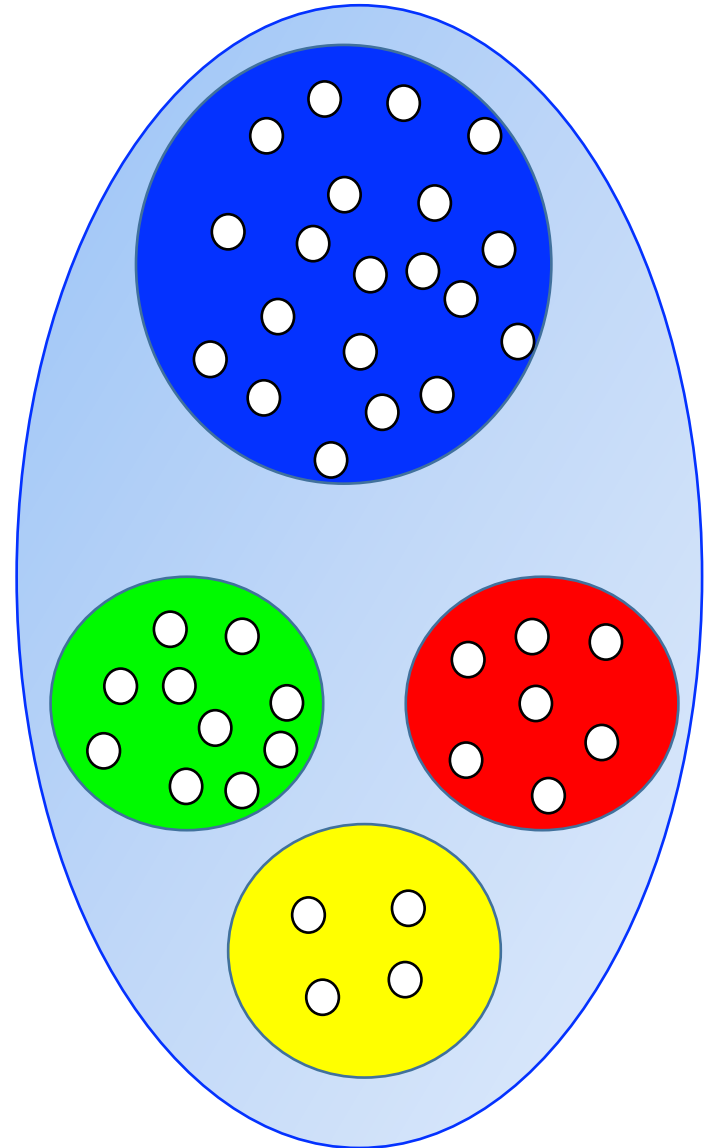
Muestreo aleatorio estratificado

A la población de N individuos se divide en E subpoblaciones o estratos que no se traslapan, con respecto a criterios que puedan ser importantes en el estudio y tratando en la medida posible que exista homogeneidad dentro de cada estrato. Los estratos contienen N_1, \dots, N_E unidades muestrales, de manera que $N = \sum_i^E N_i$ y en cada uno de estos estratos o subpoblaciones se realiza un muestreo aleatorio simple con muestras respectivas de tamaño n_i . As la muestra estratificada de tamaño n es igual a la suma de todas las muestras de cada estrato, es decir, $n = \sum_i^E n_i$.

El ME se utiliza cuando la población es muy heterogénea y las consideraciones de costo limitan el tamaño de la muestra.

Muestreo Estratificado

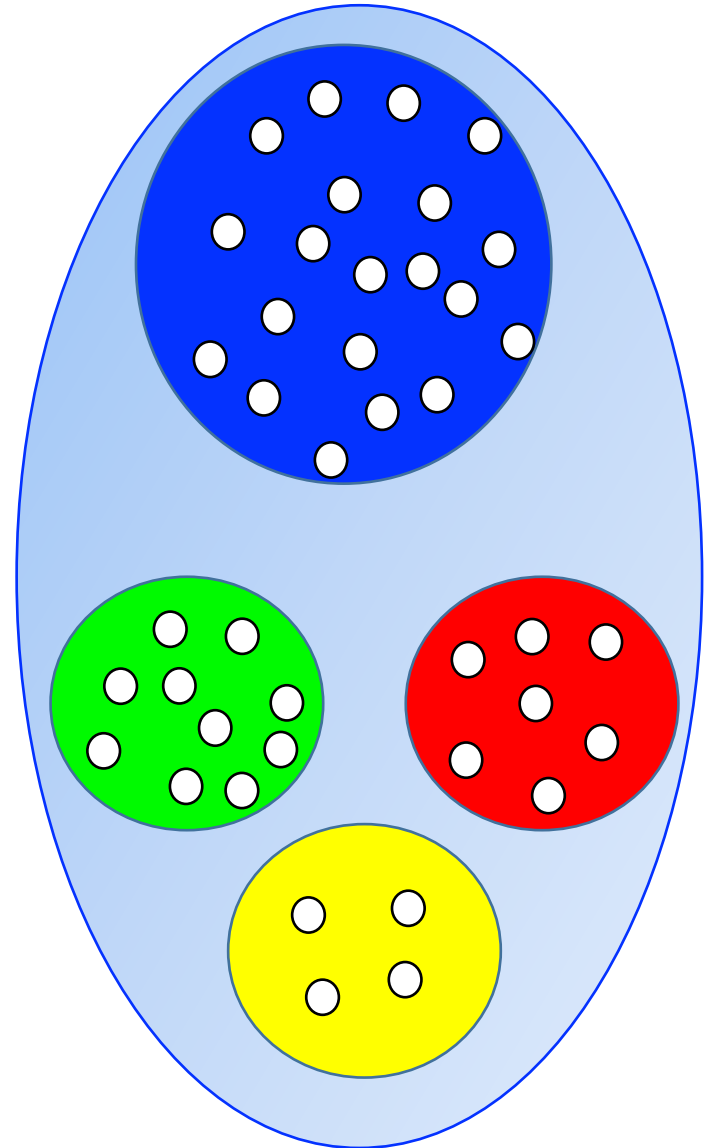
Ventajas Muestreo Estratificado



- Produce estimaciones más precisas que las que se obtienen a partir del segundo método.
- El costo por observación puede ser reducido mediante la estratificación de la población.
- Se puede obtener información de parámetros poblacionales para cada estrato de la población.
- Se simplifica el trabajo administrativo y el de control, ya que se puede usar personal específico para cada estrato.
- El tamaño de muestra será menor si la estratificación está bien definida, en comparación con el MAS.

Muestreo Estratificado

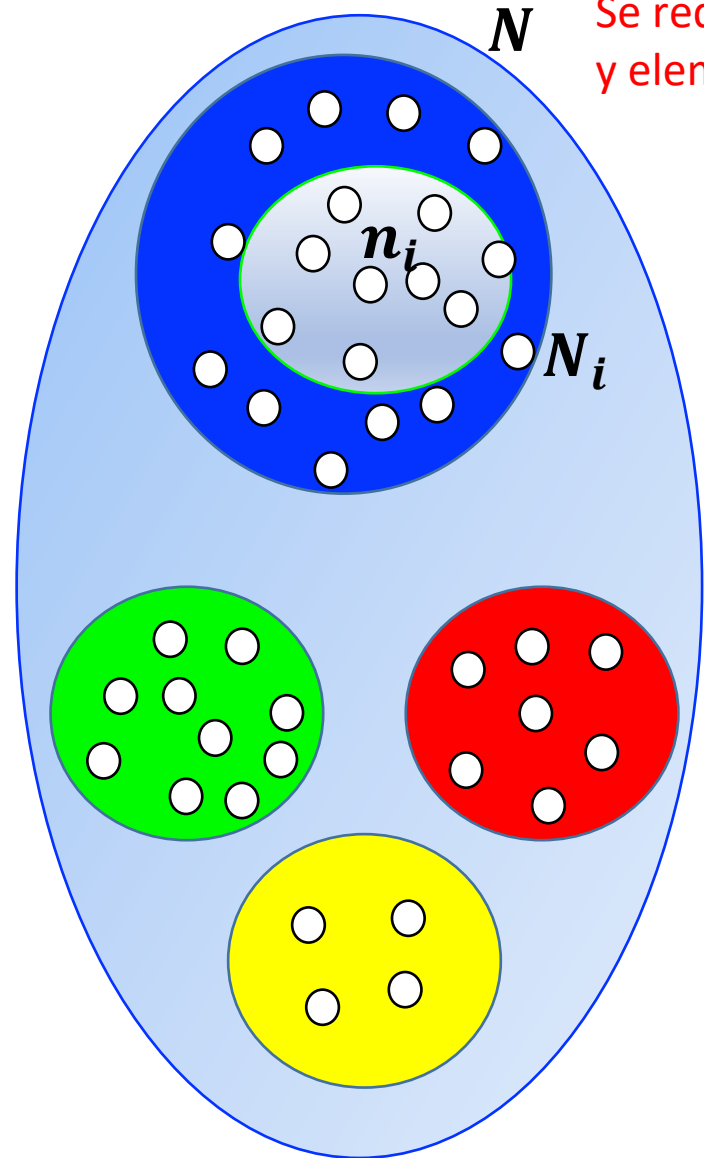
Limitantes Muestreo Estratificado



- En ocasiones las dificultades que plantean los estratos son demasiado grandes, pues exigen un conocimiento detallado de la población, como tamaño geográfico, género, edades, niveles de estudio, etcétera
- Condiciona a que la población sea heterogénea y las consideraciones de costo limitan el tamaño de la muestra. Si no se toman en cuenta tanto la variabilidad como los costos diferenciados y utilizamos el MAS, las estimaciones podrían ser menos precisas o el costo sería demasiado elevado
- Para la población estratificada habrá que determinar dos tamaños de muestra: para la población y para cada estrato.
- Generalmente consume más tiempo y es dispendioso.

Muestreo Estratificado

Notación



Se requiere de una notación adicional que distinga los elementos de la población, los estratos, y elementos en ellos, como la siguiente:

N : el número total de unidades muestrales en la población.

E : el número de estratos en la población.

i : un estrato.

N_i : el número total de unidades en el estrato i .

n_i : el número de unidades en la muestra en el estrato i .

j : alguna unidad muestral que siempre pertenece a algún estrato i .

y_{ij} : el valor obtenido en la j -ésima unidad dentro del estrato i .

$W_i = N_i/N$: la ponderación, peso o tamaño relativo del estrato i .

$f_i = n_i/N_i$: la fracción de muestreo para el estrato i .

$\bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$: la media muestral del estrato i .

$S_i^2 = \frac{\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{n_i - 1} = \frac{\sum_{j=1}^{n_i} y_{ij}^2 - n_i \bar{y}_i^2}{n_i - 1}$: la varianza en el estrato i .

Muestreo Estratificado

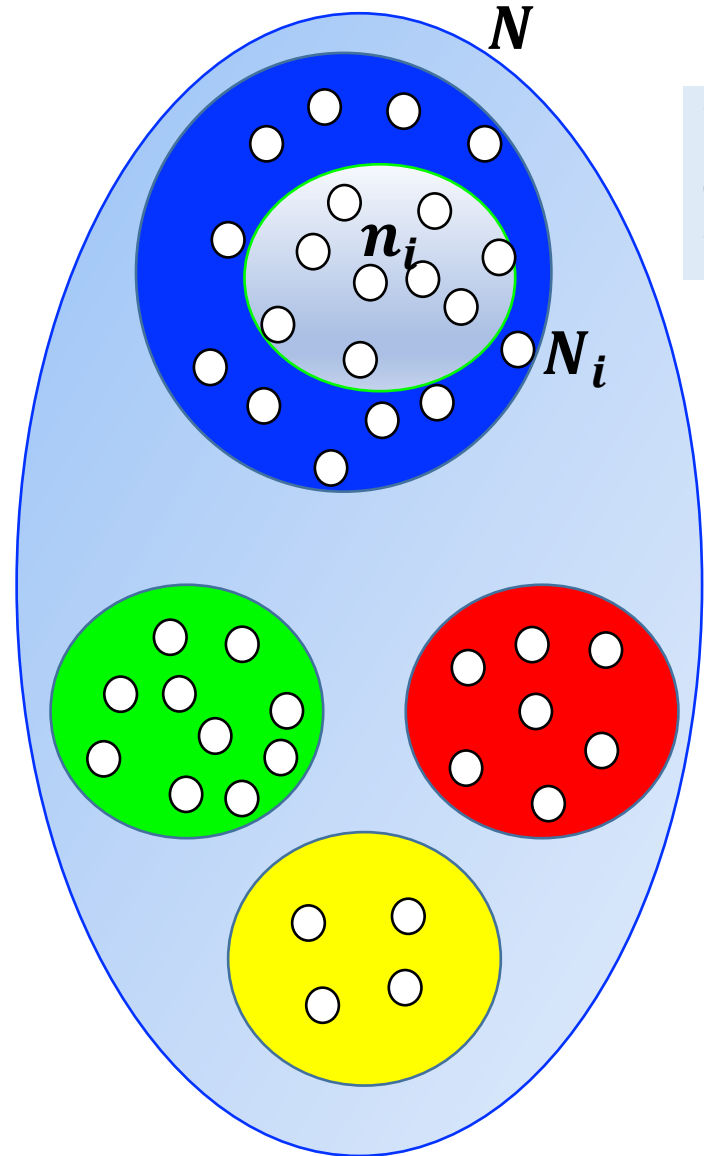
¿Cómo seleccionar una muestra aleatoria estratificada?

La selección de la muestra de cada estrato es diferente, ya que cada uno tiene características y costos de medición distintos, por lo que el número de unidades también será diferente.

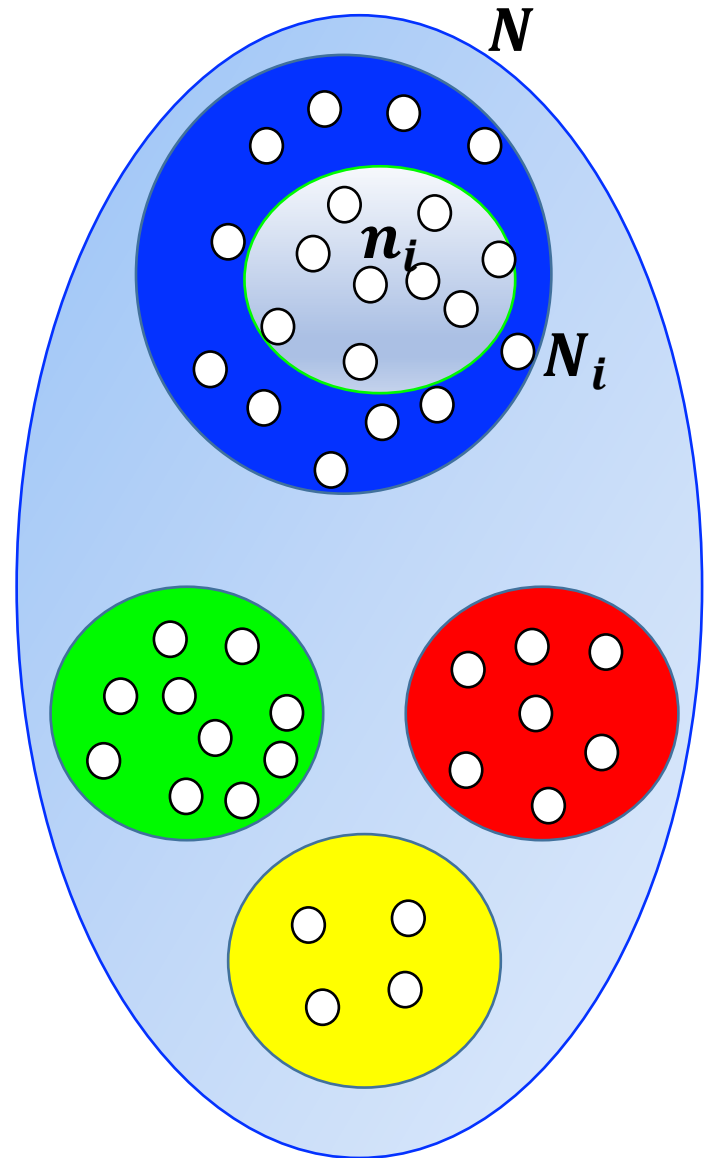
- El tamaño de la muestra del estrato debe ser mayor si es muy variable o si contiene más unidades.
- El tamaño de la muestra del estrato será menor si el costo de la medición es elevado.
- Considerar qué tan grande debe ser la precisión de estimación y de acuerdo con esto, seleccionar el tamaño de la muestra

En resumen, de un estrato dado se toma una muestra más grande si:

- El estrato es más grande.
- Los elementos del estrato tienen alta variabilidad.
- El muestreo es más barato en el estrato.



Muestreo Estratificado



PRINCIPIOS PARA FORMAR ESTRATOS:

1. Homogeneidad Interna
2. Heterogeneidad Externa
3. Tamaños Manejables

Muestreo Estratificado

N ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

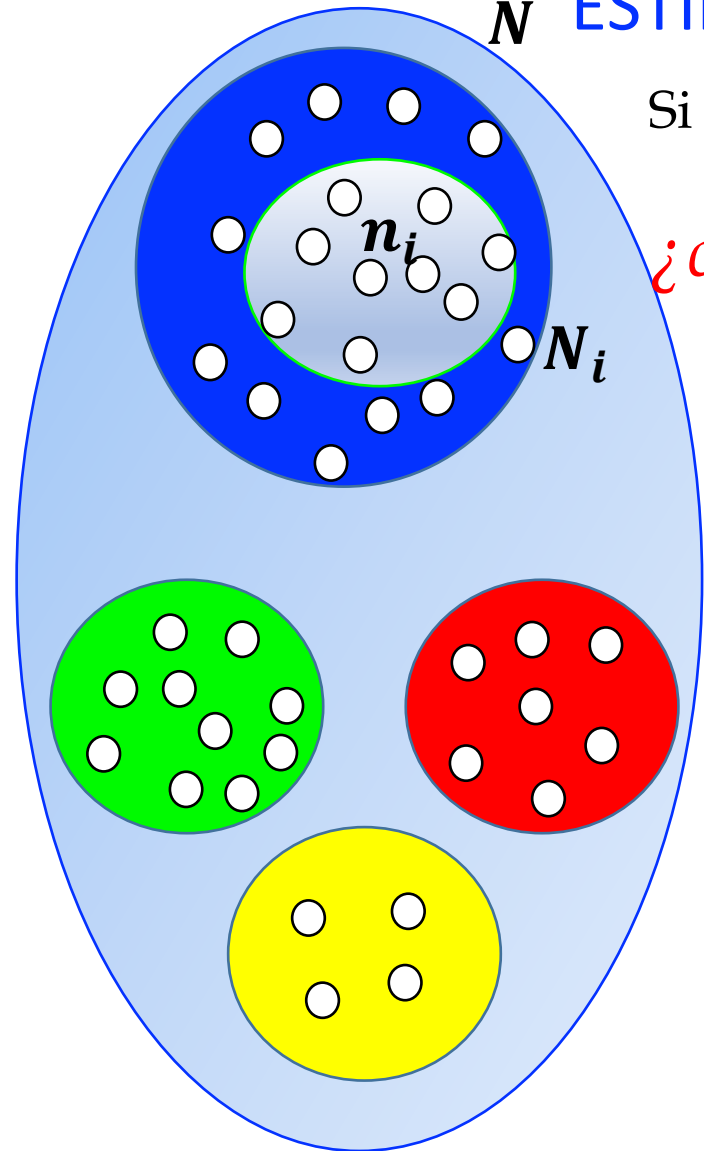
Si hemos tomado nuestra muestra estratificada,

¿cómo debemos usarla para estimar los principales parámetros?

- El tamaño de la muestra del estrato será menor si el costo de la medición es elevado.
- Considerar qué tan grande debe ser la precisión de estimación y de acuerdo con esto, seleccionar el tamaño de la muestra

En resumen, de un estrato dado se toma una muestra más grande si:

- El estrato es más grande.
- Los elementos del estrato tienen alta variabilidad.
- El muestreo es más barato en el estrato.



ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

Estimación de la MEDIA estratificada

Una vez obtenida la muestra estratificada.

Estimador para la Media Poblacional:

$$\bar{y}_{estr} = \frac{\sum_{i=1}^E N_i \bar{y}_i}{N}$$

\bar{y}_{estr} : La media a partir de una muestra estratificada

\bar{y}_i : media en cada estrato de la muestra

Estimador de la varianza de la Media estratificada:

La varianza de \bar{y}_{estr} es la suma de las varianzas de las medias de cada estrato. Este estimador es insesgado.

$$S_{\bar{y}_{estr}}^2 = V(\bar{y}_{estr}) = \frac{1}{N^2} [N_1^2 \hat{V}(\bar{y}_1) + N_2^2 \hat{V}(\bar{y}_2) + \dots + N_E^2 \hat{V}(\bar{y}_E)]$$

$$= \frac{1}{N^2} \left[N_1^2 \left(\frac{N_1 - n_1}{N_1} \right) \left(\frac{S_1^2}{n_1} \right) + \dots + N_E^2 \left(\frac{N_E - n_E}{N_E} \right) \left(\frac{S_E^2}{n_E} \right) \right]$$

$$= \frac{1}{N^2} \sum_{i=1}^E N_i^2 \left(\frac{N_i - n_i}{N_i} \right) \left(\frac{S_i^2}{n_i} \right)$$

$$= \sum_{i=1}^E \frac{N_i^2}{N^2} \left(\frac{N_i - n_i}{N_i} \right) \left(\frac{S_i^2}{n_i} \right)$$

$$= \sum_{i=1}^E W_i^2 \left(\frac{N_i - n_i}{N_i} \right) \left(\frac{S_i^2}{n_i} \right)$$

$$= \sum_{i=1}^E W_i^2 S_{\bar{y}_i}^2$$

$$S_{\bar{y}_{estr}}^2 = \sum_{i=1}^E \left(\frac{N_i}{N} \right)^2 \left(\frac{N_i - n_i}{N_i} \right) \left(\frac{S_i^2}{n_i} \right)$$

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

Estimación Intervalo de Confianza de la MEDIA estratificada

Una vez obtenida la muestra estratificada.

Cuando hay pocos grados de libertad en cada estrato se usa el Estadístico t de t -student y cuando es mayor de 30 se usa z de la distribución normal estándar.

Error de Muestreo: $(t\sqrt{\bar{V}(\bar{y}_{estr})})$

Intervalo de Confianza para MEDIA Estratificada:

$$\bar{y}_{estr} \pm t_{(n-1, 1-\alpha/2)} \sqrt{\frac{1}{N^2} \sum_{i=1}^E N_i^2 \left(\frac{N_i - n_i}{N_i} \right) \left(\frac{S_i^2}{n_i} \right)}$$
$$\bar{y}_{estr} \pm t_{(n-1, 1-\alpha/2)} \sqrt{\sum_{i=1}^E W_i^2 S_{\bar{y}_i}^2}$$

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

Estimación del TOTAL estratificado

Una vez obtenida la muestra estratificada.

Estimador para el TOTAL Poblacional:

$$\hat{t}_{estr} = N\bar{y}_{estr} = N_1\bar{y}_1 + N_2\bar{y}_2 + \dots + N_E\bar{y}_E = \sum_{i=1}^E N_i\bar{y}_i$$

\bar{y}_{estr} : La media a partir de una muestra estratificada

\bar{y}_i : media en cada estrato de la muestra

Estimador de la varianza del TOTAL :

La varianza se deduce de la varianza de la media y hereda sus propiedades. Nota: debe haber al menos dos observaciones en cada estrato.

$$\bar{V}(N\bar{y}_{estr}) = N^2 V(\bar{y}_{estr}) = N^2 \sum_{i=1}^E W_i^2 \left(\frac{N_i - n_i}{N_i} \right) \left(\frac{S_i^2}{n_i} \right)$$

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

Estimación Intervalo de Confianza del TOTAL

Una vez obtenida la muestra estratificada.

Error de Muestreo: $t_{(n-1, 1-\alpha/2)} \sqrt{\sum_{i=1}^E N_i^2 \left(\frac{N_i - n_i}{N_i}\right) \left(\frac{S_i^2}{n_i}\right)}$ Cuando hay pocos grados de libertad en cada estrato se usa el Estadístico t de t -student y cuando es mayor de 30 se usa z de la distribución normal estándar.

Intervalo de Confianza para el TOTAL :

$$N\bar{y}_{estr} \pm t_{(n-1, 1-\alpha/2)} \sqrt{\sum_{i=1}^E N_i^2 \left(\frac{N_i - n_i}{N_i}\right) \left(\frac{S_i^2}{n_i}\right)}$$

$$\hat{t}_{estr} \pm N t_{(n-1, 1-\alpha/2)} S_{\bar{y}_{estr}}$$

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

Tamaño de la Muestra

Para determinar el tamaño de muestra se deben tomar en cuenta varios factores:

- el parámetro a estimar
- la precisión admisible
- la varianza poblacional y
- el nivel de confianza de la inferencia

Para la asignación de la muestra a cada estrato también se requiere información sobre:

- El número total de elementos del estrato.
- La variabilidad de las observaciones del estrato.
- El costo que representa muestrear cada estrato.

Es importante considerar que si el costo de obtener una observación varía entre estratos, se deberán tomar muestras pequeñas en estratos donde el costo sea alto y viceversa, con el fin de minimizar el costo total del muestreo.

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

Tamaño de la Muestra

Para garantizar la precisión deseada:

$$d = t_{(n-1, 1-\alpha/2)} \sqrt{V(\bar{\theta})}$$

donde $V(\bar{\theta})$ es la varianza del estimador de interés y $100(1-\alpha)\%$ es el nivel de confianza.

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

Tamaño de la Muestra para la MEDIA estratificada

$$n^* = \frac{N t_{(n-1, 1-\alpha/2)}^2 \sum_{i=1}^E W_i S_i^2}{N d^2 + N t_{(n-1, 1-\alpha/2)}^2 \sum_{i=1}^E W_i S_i^2}$$

En la ecuación anterior $W_i = N_i/N$, mientras que d es el margen de error (precisión) que se desea aceptar en la estimación de la media.

Calculadora Tamaño de Muestra Estratificada: Ver Ejemplo de cómo usarla [CLICK AQUÍ](#)

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

Asignación de la Muestra: Proporcional al tamaño estrato

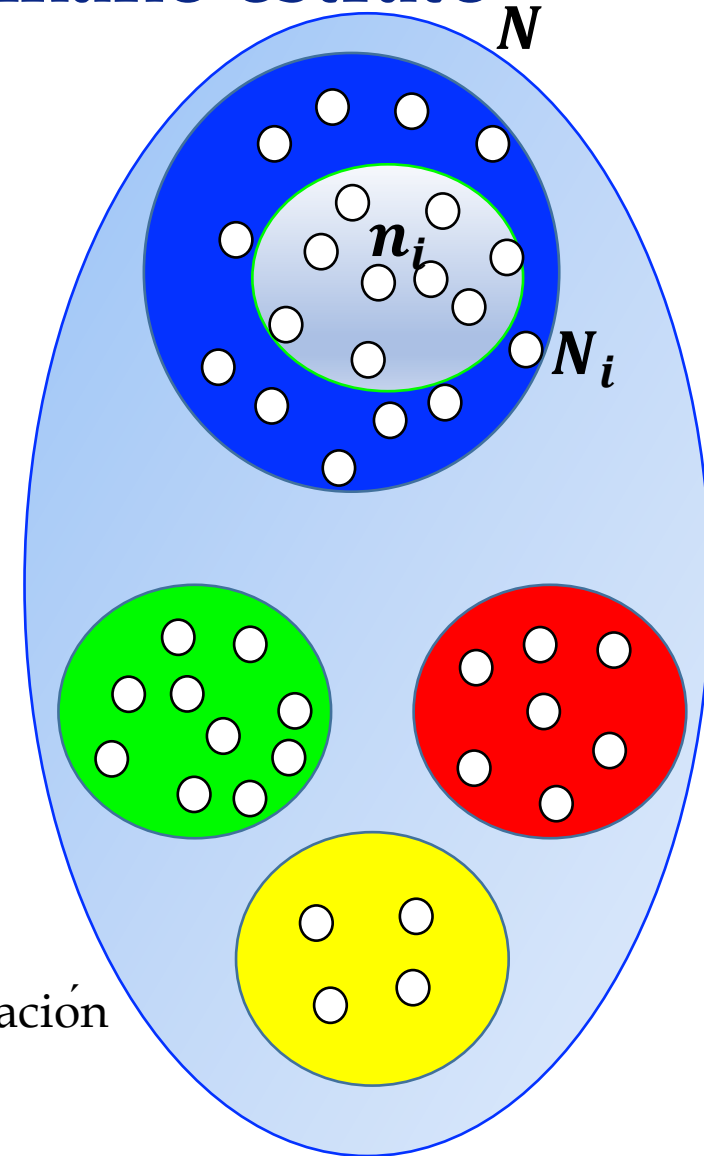
Hay que considerar que el número de unidades del estrato influye en el tamaño de muestra. Se asignará un tamaño de muestra mayor a los estratos más grandes y uno menor a los estratos más chicos. Existen diferentes métodos de asignación de la muestra.

El criterio de asignación proporcional, suponiendo que ya se ha calculado el tamaño de muestra n_m requerido, consiste en determinar una parte de n_m , la cual será proporcional al tamaño del estrato.

$n_{m_i} = [(N_i)/N]n_m = W_i n_m$, donde $i = 1, 2, \dots, E$, es decir, como:

$$n_{m_1} = \frac{N_1}{N} n_m, n_{m_2} = \frac{N_2}{N} n_m, \dots, n_{m_E} = \frac{N_E}{N} n_m$$

Los estratos más grandes requieren un tamaño de muestra mayor, es decir, la asignación de n entre los estratos es proporcional al tamaño del estrato.



ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

Asignación de la Muestra:

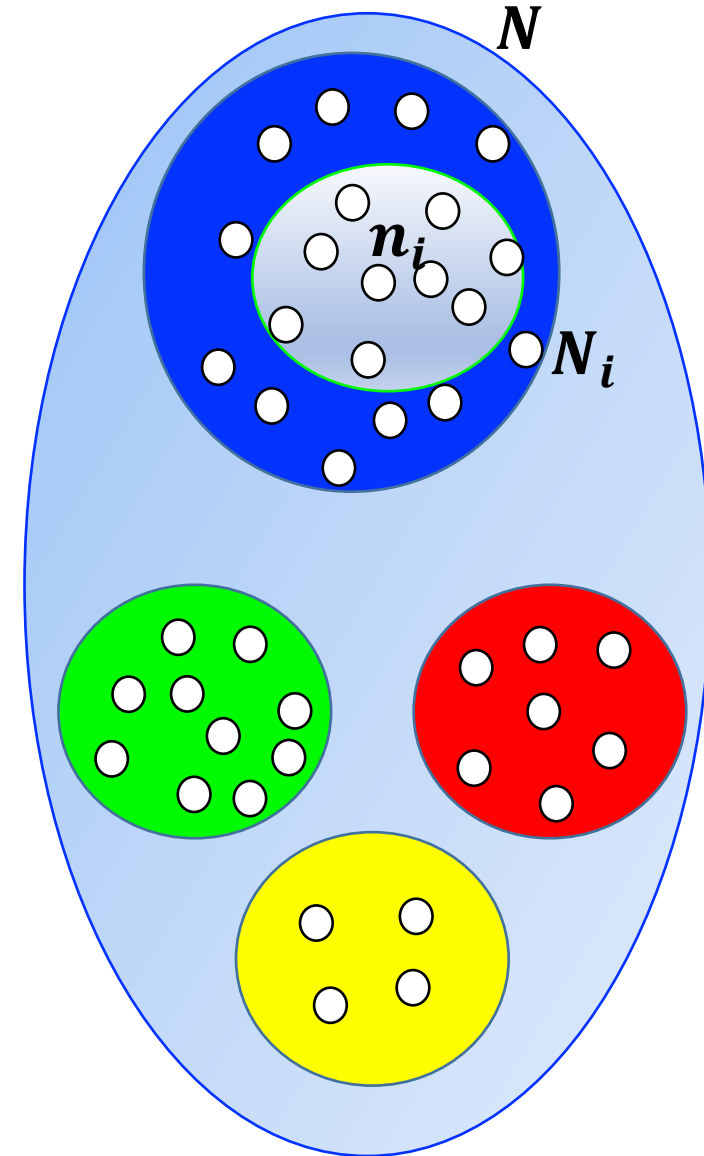
Muestreo estratificado óptimo (de Neyman)

Determina el tamaño de muestra óptimo en cada estrato minimizando la varianza del estimador de la media.

Fórmula de Neyman:

$$n_h = n \cdot \frac{N_h S_h}{\sum_{k=1}^L N_k S_k}$$

donde S_h es la desviación estándar del estrato h .



ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

Asignación de la Muestra:

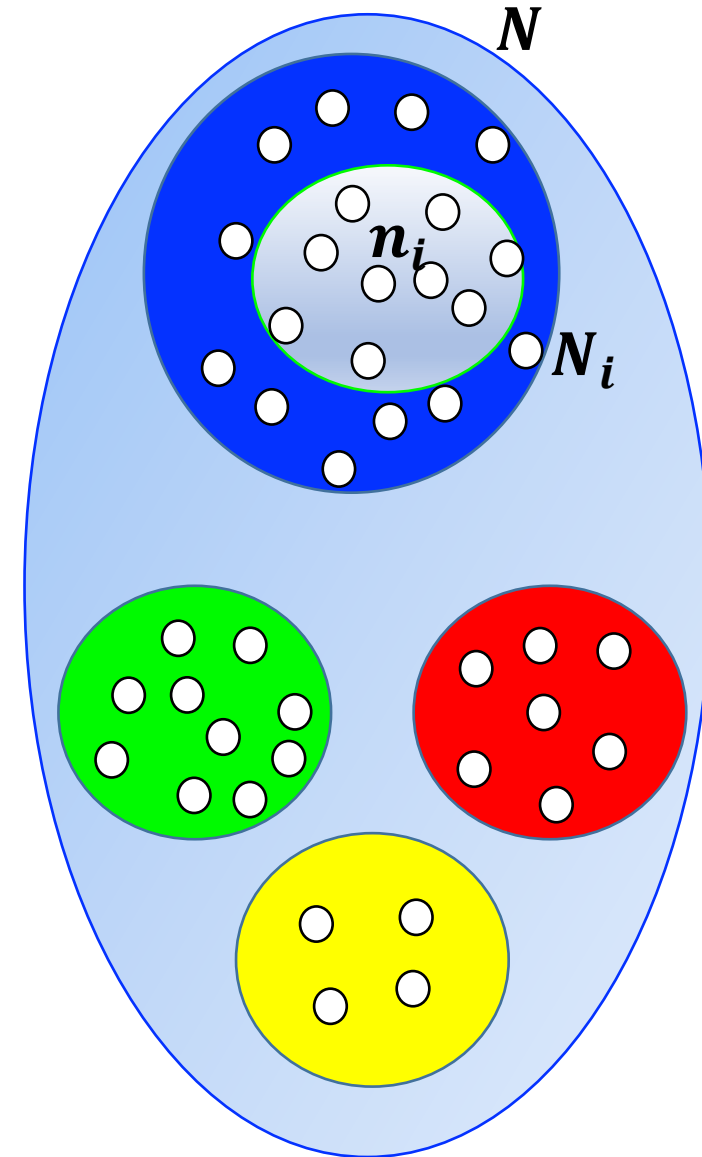
Muestreo estratificado óptimo (de Neyman)

Ejemplo: Una empresa desea estimar el **promedio de horas de capacitación recibidas** por sus empleados en tres áreas (estratos):

- Producción
- Ventas
- Administración

La población total es de $N=600$ empleados, divididos así:

Estrato	Tamaño N_h	Desviación estándar S_h (horas)
Producción	300	10
Ventas	180	15
Administración	120	20



ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

Asignación de la Muestra: Muestreo estratificado óptimo (de Neyman)

Estrato	Tamaño N_h	Desviación estándar S_h (horas)
Producción	300	10
Ventas	180	15
Administración	120	20

La asignación óptima es:

$$n_h = n \cdot \frac{N_h S_h}{\sum_{k=1}^L N_k S_k}$$

Numeradores:

- $N_1 S_1 = 300 \cdot 10 = 3000$
- $N_2 S_2 = 180 \cdot 15 = 2700$
- $N_3 S_3 = 120 \cdot 20 = 2400$

Denominador total:

$$\sum N_h S_h = 3000 + 2700 + 2400 = 8100$$

Asignaciones:

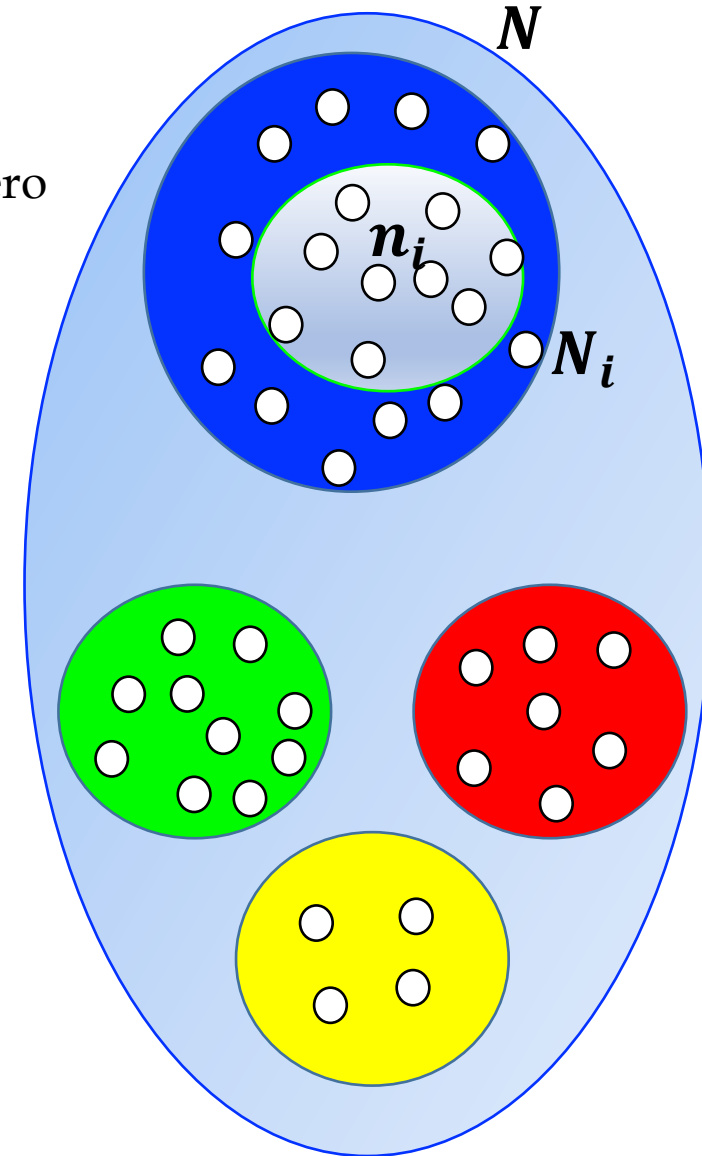
- $n_1 = 60 \cdot \frac{3000}{8100} = 22.22 \approx 22$
- $n_2 = 60 \cdot \frac{2700}{8100} = 20.00$
- $n_3 = 60 \cdot \frac{2400}{8100} = 17.78 \approx 18$

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

Selección de los Estratos

En ocasiones es sencillo delimitar los elementos que corresponden a cada estrato, pero ¿siempre es así? Definitivamente no.

- A veces es una tarea sencilla debido a que los estratos están implícitos y se conoce el comportamiento con base en registros antiguos, en nuestra experiencia o simplemente en la naturaleza de los resultados que deseamos obtener.
- ¿Con base en qué se delimitan los estratos? Una primera aproximación es el caso cuantitativo: *método acumulativo de la raíz cuadrada de la frecuencia* el cual se basa en las frecuencias de la categoría de la variable interés que determinaría los estratos. Ver EJEMPLO del Método [CLICK AQUÍ](#)



ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

Ejemplo:

Estadísticas por tema
AGROPECUARIO



En la region Caribe se conoce a partir del Censo agropecuario(DANE) que existen 780 fincas sembradas con aguacate. Se desea estimar el promedio de plantas por hectárea que en determinada etapa del cultivo se infectaron de alguna enfermedad. De acuerdo con las condiciones ecológicas en la región, se siembran tres variedades de aguacate. Considerando que el desarrollo de la enfermedad puede ser distinto de una variedad a otra, la población de parcelas se estratificó en $E = 3$ estratos. Los tamaños de los estratos son: $N_1 = 270$, $N_2 = 180$ y $N_3 = 330$; $N = N_1 + N_2 + N_3 = 780$. Suponga que para realizar las estimaciones se tomó una muestra de $n = 63$ parcelas

Hallar:

- a) La estimación puntual del promedio de plantas infectadas por hectárea.
- b) La estimación puntual del total estratificado de plantas infectadas.
- c) La varianza y desviación estándar del promedio estratificado.
- d) El IC de la media estratificada con una confiabilidad de 95%.
- e) El IC del total de plantas infectadas en la población con una confiabilidad de 95%.

Estrato 1 ($n_1 = 21$)			Estrato 2 ($n_2 = 21$)			Estrato 3 ($n_3 = 21$)		
48	53	64	20	31	45	74	68	77
62	45	47	36	17	26	70	72	73
59	65	54	15	30	18	78	76	69
45	48	46	40	25	35	69	80	74
50	60	63	24	29	30	80	78	71
55	57	46	19	42	27	72	71	79
64	61	54	33	51	48	76	75	68

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

Ejemplo:

Estadísticas por tema
AGROPECUARIO



Considerando que el desarrollo de la enfermedad puede ser distinto de una variedad a otra, la población de parcelas se estratificó en $E = 3$ estratos. Los tamaños de los estratos son: $N_1 = 270$, $N_2 = 180$ y $N_3 = 330$; $N = N_1 + N_2 + N_3 = 780$. Suponga que para realizar las estimaciones se tomó una muestra de $n = 63$ parcelas

Hallar:

a) La estimación puntual del promedio de plantas infectadas por hectárea.

$$\bar{y}_{estr} = \frac{\sum_{i=1}^E N_i \bar{y}_i}{N}$$

donde: N_i : $N_1 = 270$, $N_2 = 180$, $N_3 = 330$, $N = 780$, $E = 3$,

$$\bar{y}_1 = \frac{48+62+59+\dots+46+54}{21} = 54.5714,$$

$$\bar{y}_2 = \frac{20+36+15+\dots+27+48}{21} = 30.5238 \text{ y}$$

$$\bar{y}_3 = \frac{74 + 70 + 78 + \dots + 79 + 68}{21} = 73.8095$$

Estrato 1 ($n_1 = 21$)			Estrato 2 ($n_2 = 21$)			Estrato 3 ($n_3 = 21$)		
48	53	64	20	31	45	74	68	77
62	45	47	36	17	26	70	72	73
59	65	54	15	30	18	78	76	69
45	48	46	40	25	35	69	80	74
50	60	63	24	29	30	80	78	71
55	57	46	19	42	27	72	71	79
64	61	54	33	51	48	76	75	68

$$\bar{y}_{estr} = \frac{(270)(54.5714) + (180)(30.5238) + (330)(73.8095)}{780}$$

$$\bar{y}_{estr} = \frac{44,581.697}{780} = 57.1612 \text{ plantas infectadas por parcela.}$$

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

Ejemplo:

Estadísticas por tema
AGROPECUARIO



Considerando que el desarrollo de la enfermedad puede ser distinto de una variedad a otra, la población de parcelas se estratificó en $E = 3$ estratos. Los tamaños de los estratos son: $N_1 = 270$, $N_2 = 180$ y $N_3 = 330$; $N = N_1 + N_2 + N_3 = 780$. Suponga que para realizar las estimaciones se tomó una muestra de $n = 63$ parcelas

Hallar:

b) La estimación puntual del total estratificado de plantas infectadas.

$$\hat{\tau}_{estr} = N\bar{y}_{estr} \quad \text{donde: } N = 780 \text{ y } \bar{y}_{estr} = 57.1612$$

Estrato 1 ($n_1 = 21$)			Estrato 2 ($n_2 = 21$)			Estrato 3 ($n_3 = 21$)		
48	53	64	20	31	45	74	68	77
62	45	47	36	17	26	70	72	73
59	65	54	15	30	18	78	76	69
45	48	46	40	25	35	69	80	74
50	60	63	24	29	30	80	78	71
55	57	46	19	42	27	72	71	79
64	61	54	33	51	48	76	75	68

$$\hat{\tau}_{estr} = 780(57.1612) = 44,585.736 \text{ plantas infectadas.}$$

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

Ejemplo:

Estadísticas por tema
AGROPECUARIO



Considerando que el desarrollo de la enfermedad puede ser distinto de una variedad a otra, la población de parcelas se estratificó en $E = 3$ estratos. Los tamaños de los estratos son: $N_1 = 270$, $N_2 = 180$ y $N_3 = 330$; $N = N_1 + N_2 + N_3 = 780$. Suponga que para realizar las estimaciones se tomó una muestra de $n = 63$ parcelas

Hallar:

c) La varianza y desviación estándar del promedio estratificado.

$$S^2_{\bar{y}_{estr}} = \sum_{i=1}^E \left(\frac{N_i}{N} \right)^2 \left(\frac{N_i - n_i}{N_i} \right) \left(\frac{S_i^2}{n_i} \right)$$

donde: $N = 780$, $E = 3$, $N_i: N_1 = 270$, $N_2 = 180$, $N_3 = 330$, $n_i: n_1 = n_2 = n_3 = 21$,

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} \quad S_1^2 = 50.3571 \quad S_2^2 = 107.2619 \quad S_3^2 = 15.5619$$

$$S^2_{\bar{y}_{estr}} = \left(\frac{270}{780} \right)^2 \left(\frac{270 - 21}{270} \right) \left(\frac{50.3571}{21} \right) + \left(\frac{180}{780} \right)^2 \left(\frac{180 - 21}{180} \right) \left(\frac{107.2619}{21} \right) + \left(\frac{330}{780} \right)^2 \left(\frac{330 - 21}{330} \right) \left(\frac{15.5619}{21} \right)$$

$$S_{\bar{y}_{estr}} = \sqrt{S^2_{\bar{y}_{estr}}} = \sqrt{0.6295} = 0.7934$$

Estrato 1 ($n_1 = 21$)			Estrato 2 ($n_2 = 21$)			Estrato 3 ($n_3 = 21$)		
48	53	64	20	31	45	74	68	77
62	45	47	36	17	26	70	72	73
59	65	54	15	30	18	78	76	69
45	48	46	40	25	35	69	80	74
50	60	63	24	29	30	80	78	71
55	57	46	19	42	27	72	71	79
64	61	54	33	51	48	76	75	68

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

Ejemplo:

Estadísticas por tema
AGROPECUARIO



Considerando que el desarrollo de la enfermedad puede ser distinto de una variedad a otra, la población de parcelas se estratificó en $E = 3$ estratos. Los tamaños de los estratos son: $N_1 = 270$, $N_2 = 180$ y $N_3 = 330$; $N = N_1 + N_2 + N_3 = 780$. Suponga que para realizar las estimaciones se tomó una muestra de $n = 63$ parcelas

Hallar:

d) El IC de la media estratificada con una confiabilidad de 95%.

$$\bar{y}_{estr} \pm t_{(n-1, 1-\alpha/2)} S_{\bar{y}_{estr}}$$

donde: $\bar{y}_{estr} = 57.1612$, $t_{(n-1, 1-\alpha/2)} = Z_{1-\alpha/2} = Z_{1-0.025} = 1.96$ y $S_{\bar{y}_{estr}} = 0.7934$.

$$57.1612 \pm (1.96)(0.7934)$$

$$57.1612 \pm 1.5551$$

$$55.6061 \leq \mu_{estr} \leq 58.7163$$

Estrato 1 ($n_1 = 21$)			Estrato 2 ($n_2 = 21$)			Estrato 3 ($n_3 = 21$)		
48	53	64	20	31	45	74	68	77
62	45	47	36	17	26	70	72	73
59	65	54	15	30	18	78	76	69
45	48	46	40	25	35	69	80	74
50	60	63	24	29	30	80	78	71
55	57	46	19	42	27	72	71	79
64	61	54	33	51	48	76	75	68

El promedio de plantas infectadas por hectárea en la población está entre 55.6061 y 58.7163.

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

Ejemplo:

Estadísticas por tema
AGROPECUARIO



Considerando que el desarrollo de la enfermedad puede ser distinto de una variedad a otra, la población de parcelas se estratificó en $E = 3$ estratos. Los tamaños de los estratos son: $N_1 = 270$, $N_2 = 180$ y $N_3 = 330$; $N = N_1 + N_2 + N_3 = 780$. Suponga que para realizar las estimaciones se tomó una muestra de $n = 63$ parcelas

Hallar:

e) El IC del total de plantas infectadas en la población con una confiabilidad de 95%.

$$\hat{\tau}_{estr} \pm N t_{(n-1, 1-\alpha/2)} S_{\bar{y}_{estr}}$$

donde: $\hat{\tau}_{estr} = 44,585.736$, $N = 780$, $t_{(n-1, 1-\alpha/2)} = Z_{1-\alpha/2} = Z_{1-0.025} = 1.96$ y

$$S_{\bar{y}_{estr}} = 0.7934.$$

$$44,585.736 \pm (780)(1.96)(0.7934)$$

$$44,585.736 \pm 1,212.9499$$

Estrato 1 ($n_1 = 21$)			Estrato 2 ($n_2 = 21$)			Estrato 3 ($n_3 = 21$)		
48	53	64	20	31	45	74	68	77
62	45	47	36	17	26	70	72	73
59	65	54	15	30	18	78	76	69
45	48	46	40	25	35	69	80	74
50	60	63	24	29	30	80	78	71
55	57	46	19	42	27	72	71	79
64	61	54	33	51	48	76	75	68

$$43,372.7861 \leq \tau_{estr} \leq 45,798.6859$$

El total de plantas infectadas por hectárea en la población está entre 43,372.7861 y 45,798.6859.

Implementación del ME con Python

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

EJERCICIO:

A comienzos del COVID 19 la inasistencia a las clases de un Colegio fue un elemento recurrente(debido en parte a las limitaciones tecnológicas de muchos estudiantes). Si la población de $N = 640$ estudiantes que alberga la totalidad de los estudiantes quedó estratificada de la siguiente manera: Estrato 1 (primer año): $N_1 = 200$ alumnos, Estrato 2 (segundo año): $N_2 = 150$ alumnos, Estrato 3 (tercer año): $N_3 = 150$ alumnos y Estrato 4 (cuarto año): $N_4 = 140$ alumnos. Se seleccionó una muestra de $n = 40$ alumnos: 12 para el Estrato 1, 10 para el Estrato 2 y 9 para los estratos 3 y 4.

Hallar:

Datos de Inasistencias

- a) Intervalo de Confianza del promedio de inasistencia estratificado al 95%.
- b) Intervalo de Confianza de para el total estudiantes que faltaron al curso 95%.

Estrato 1 ($n_1 = 12$)		Estrato 2 ($n_2 = 10$)		Estrato 3 ($n_3 = 9$)		Estrato 4 ($n_4 = 9$)	
7	6	4	4	3	4	3	3
6	7	5	6	3	4	2	2
7	7	4	4	3	3	3	5
8	8	5	6	4	4	4	2
5	5	6	6	4		5	
4	6						

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

EJERCICIO: Muestreo estratificado en uso de plataformas de streaming

Netflix quiere estimar el promedio de horas semanales de *streaming* en sus tres grupos de usuarios según edad:

- 18–29 años ($N_1=350$)
- 30–49 años ($N_2=450$)
- 50+ años ($N_3=200$)



Datos en Savio archivo “usuarios_streaming.csv”

Punto 1:

Se toma una muestra de $n=90$ usuarios (**30 por estrato**). Se pide:

1. Estimar el promedio estratificado de horas de uso.
2. Estimar el total de horas de streaming en la población.
3. Calcular varianza, desviación estándar y los ICs al 95% para la media y el total.

Punto 2:

Se toma una muestra de $n=90$ usuarios (**asignación proporcional al estrato**). Se pide:

1. Estimar el promedio estratificado de horas de uso.
2. Estimar el total de horas de streaming en la población.
3. Calcular varianza, desviación estándar y los ICs al 95% para la media y el total.

CASO APLICACIÓN

Artículo: Estimación de Muestra para un estudio

http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1794-44492018000200223

Artículo original / Original article / Artigo original

DOI: 10.22507/rli.v15n2a17

Muestreo para el levantamiento de datos acerca de la enseñanza de física experimental en Guayaquil¹

Víctor Manuel Barros², Diana Ercilia Gallegos Zurita³, Christian Antonio Pavón Brito⁴

CASO APLICACIÓN

Artículo: Estimación de Muestra para un estudio

http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1794-44492018000200223

Por las características de la población, la cual está dividida en estratos o conjuntos homogéneos (distritos), respecto a la variable objetivo (instituciones con BGU) y el tamaño de esta, la selección de los elementos de la población se la realizó utilizando el muestreo estratificado.

Se ha considerado cada distrito educativo como una categoría (estrato) conformado por instituciones educativas de características similares, tales como currículo, régimen, modalidad, jornada, ente otras.

CASO APLICACIÓN

Artículo: Estimación de Muestra para un estudio

http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1794-44492018000200223

MATERIALES Y MÉTODOS

El muestreo estratificado es adecuado puesto que la población de estudio, en este caso el cantón Guayaquil ya está dividido en distritos educativos (estratos) cada uno de diferente tamaño. La ventaja de las muestras estratificadas es que cuando se diseñan bien, reflejan más

exactamente las características de la población de donde se extrajeron, que otras clases de muestreo (UNEG, 2016).

Para aplicar el muestreo estratificado, se divide la población de interés en grupos homogéneos (estratos), los grupos constituyen los diez distritos educativos descritos en la Tabla 1, los cuales son

CASO APLICACIÓN

Artículo: Estimación de Muestra para un estudio

http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1794-44492018000200223

10 distritos con sus respectivos barrios

CODIGO	NOMBRE	Dirección	Teléfonos	Parroquias-Barrios
09D02	Ximena 2	Junto al colegio Francisco Orellana calles Arturo Serrano y Carlos Garcés.	(04) 2440295 / 2440296 / 2440298	Santiago Roldós, Fertisa, Esteros, Fragata, Santa Mónica, Coviem, Sopeña, Acacias, Huancavilca, Guangala, Isla Trinitaria, Esmeraldas Chiquita, Jacobo Bucarám. Eje vial comunicante: Av. 25 de Julio.
09D03	Centro	Antigua Dirección Provincial de Educación calle Aguirre junto al Colegio Vicente Rocafuerte	(04) 3715500	Astillero, Centro, Las Peñas, Garay, La Chala, Cristo del Consuelo, Eje vial comunicante: Portete, Av. Quito, Lizardo García.
09D04	Portete	Junto al Colegio Patria Ecuatoriana calle 40 entre Portete y Camulo Destruge.	(04) 2462618 / 2462269 ext, 104	Barrio Lindo, El Cisne, Puerto Liza, Batallón del Suburbio, Fe y Alegría. Eje vial comunicante: Porte, Rosendo Avilés (Francisco Segura), Av. 29.

Ver el resto de distritos en el artículo

CASO APLICACIÓN

Artículo: Estimación de Muestra para un estudio

http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1794-44492018000200223

“calculó el tamaño de la muestra usando la fórmula para determinar la muestra en estudios descriptivos donde la variable es de tipo cualitativa en una población finito”

Donde:

n = tamaño de la muestra

N = tamaño de la población

p = proporción aproximada del fenómeno en estudio en la población de referencia

q = proporción de la población de referencia que no presenta el fenómeno en estudio (1 -p).

$$n = \frac{NpqZ^2}{(N-1)e^2 + pqZ^2} \quad (1)$$

Z = valor de Z crítico, calculado en las tablas del área de la curva normal. Llamado también nivel de confianza.

e = nivel de precisión absoluta. Referido a la amplitud del intervalo de confianza deseado en la determinación del valor promedio de la variable en estudio.

CASO APLICACIÓN

Artículo: Estimación de Muestra para un estudio

http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1794-44492018000200223

RESULTADOS

$$n = \frac{(384)(0,25)(1,96)^2}{(384 - 1)(0,05)^2 + (0,25)(1,96)^2} = 192$$

Tabla 2. Distribución de Instituciones Educativas y estudiantes por distrito

Código de Distrito	Nombre del Distrito	Total, Instituciones educativas	Total, estudiantes	Total, Instituciones con Bachillerato	Total, estudiantes de Bachillerato
09D01	XIMENA 1	217	91082	44	40355
09D02	XIMENA 2	121	60017	24	28909
09D03	CENTRO	175	90464	62	59980
09D04	PORTETE	156	76159	28	34045
09D05	TARQUI-TENGUEL	205	86159	74	70952
09D06	BELLAVISTA –MAPASINGUE	157	98224	53	56285
09D07	LOS VERGELES	131	72823	35	41906
09D08	MONTE SINAI	116	91626	36	48515
09D09	CHONGON	29	15085	19	13026
09D10	GOMEZ RENDON (PROGRESO)	46	19128	9	9818
TOTAL	10	1353	700767	384	403791

CASO APLICACIÓN

Artículo: Estimación de Muestra para un estudio

http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1794-44492018000200223

RESULTADOS

Asignación de la Muestra

$$n = \frac{(384)(0,25)(1,96)^2}{(384 - 1)(0,05)^2 + (0,25)(1,96)^2} = 192$$

DISCUSIÓN

Basados en los resultados, se debe realizar la encuesta en 192 instituciones educativas de nivel bachillerato como muestra representativa de las 384 que conforman la población. Con esta información se calculó la proporción de instituciones por cada distrito. La columna proporción, de la Tabla 3, detalla la cantidad de instituciones por distrito que serán investigadas, luego con el muestreo aleatorio simple se definió cuales instituciones de la muestra se han de seleccionar de cada uno de los distritos.

Tabla 3. Muestreo estratificado en relación con los 10 distritos educativos

CÓDIGO	NOMBRE	CANTIDAD IE(POBLACIÓN)	PORCENTAJE	PROPORCION (MUESTRA ni)
09D01	XIMENA 1	44	11%	22
09D02	XIMENA 2	24	6%	12
09D03	CENTRO	62	16%	31
09D04	PORTETE	28	7%	14
09D05	TARQUI-TENGUEL	74	19%	37
09D06	BELLAVISTA– MAPASINGUE	53	14%	27
09D07	LOS VERGELES	35	9%	18
09D08	MONTE SINAI	36	9%	18
09D09	CHONGON	19	5%	10
09D10	GOMEZ RENDON (PROGRESO)	9	2%	5
TOTAL		384	100%	192

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO ESTRATIFICADO

EJERCICIO:

Replicar metodología para el caso de Cartagena sabiendo que existen tres localidades con los datos sobre instituciones dado en la table de abajo.

- 1- Leer el artículo anterior (http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1794-44492018000200223)
- 2- Estime el tamaño de la muestra con la formula y valores de parámetros sugerida en el artículo.
- 3- Replicar la tabla 3 del artículo para la asignación del muestreo estratificado a las localidades de acuerdo al tamaño de la muestra estimado en el punto anterior.

Distribución de instituciones por Localidades

Localidad	Total Instituciones Educativas	Total Estudiantes	Total Inst. con Bachillerato	Total Est. de Bachillerato
Histórica y del Caribe Norte	80	85000	50	60000
De la Virgen y Turística	77	55000	40	42000
Industrial y de la Bahía	32	30000	18	26550
TOTAL	189	170000	108	128550