Muestreo Estadístico

Muestreo por Conglomerado

Cluster Sampling (SS)

Andy Domínguez

adominguez@utb.edu.co

Jul 2025

FACULTAD DE CIENCIAS BÁSICAS

Especialización en Estadística Aplicada y Ciencia de Datos

Contenido

- Introducción
- Muestreo por Conglomerado -MC
- Ventajas/Limitantes del MC
- Estimación de parámetros bajo MC
- Tamaño de muestra para MC
- Práctica en Python

Muestreo por Conglomerado

La muestra por conglomerados

Una muestra obtenida aleatoriamente de conglomerados (de la misma forma que en el muestreo simple aleatorio), en donde a las unidades de muestreo primarias definidas les llamaremos conglomerados, las cuales son grupos de elementos (o unidades de muestreo secundarias), sobre las que se hará la medición o evaluación de la característica de interés (Pérez, 2000). Es decir, en éste diseño se extrae bajo MAS una muestra de tamaño n de conglomerados donde cada conglomerado es una colección de elementos o conglomerados.

El muestreo por conglomerados es muy adecuado cuando los grupos en que dividimos la población son muy similares entre sí, por lo que no hay gran diferencia entre estudiar individuos de un grupo o de otro.

En este tipo de muestreo, las unidades observacionales (Unidades Muestrales Secundarias- UMS) son agregadas en mayores unidades muestrales llamadas **Cluster** (Unidades Muestrales Primarias- UMP)

Osval A Montesino, et al. *Muestreo Estadístico*, Ed. Universidad Colima, 2009

Muestreo por Conglomerado

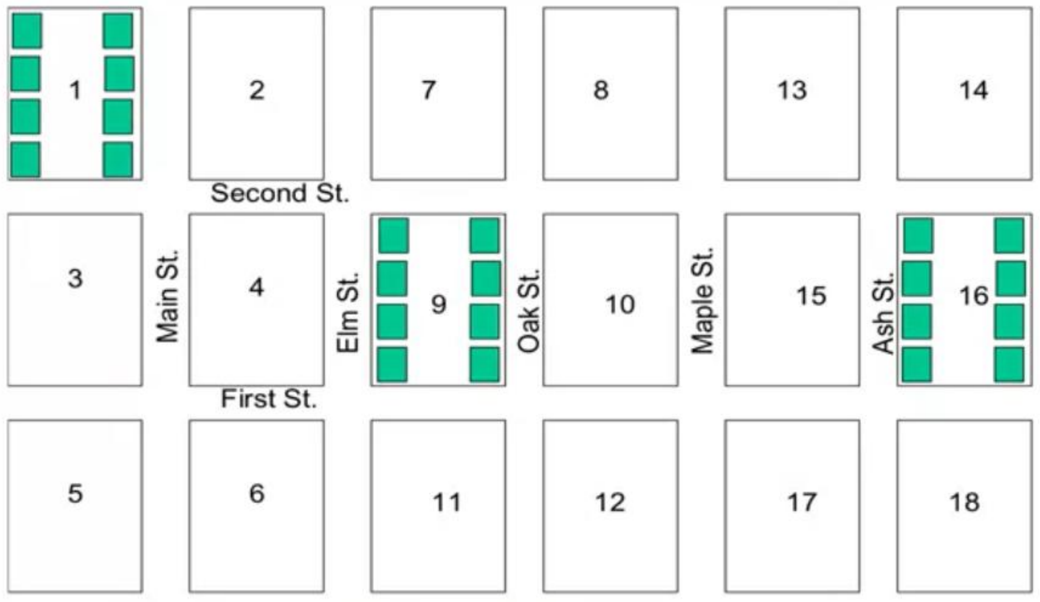
Tipos:

•**Una etapa:** se seleccionan aleatoriamente conglomerados y se estudian **todos** los elementos dentro de ellos.

•**Dos etapas:** se seleccionan aleatoriamente conglomerados, y luego **una muestra aleatoria** dentro de cada conglomerado.

Osval A Montesino, et al. *Muestreo Estadístico*, Ed. Universidad Colima, 2009

Muestreo por Conglomerado



Ventajas Muestreo por Conglomerados

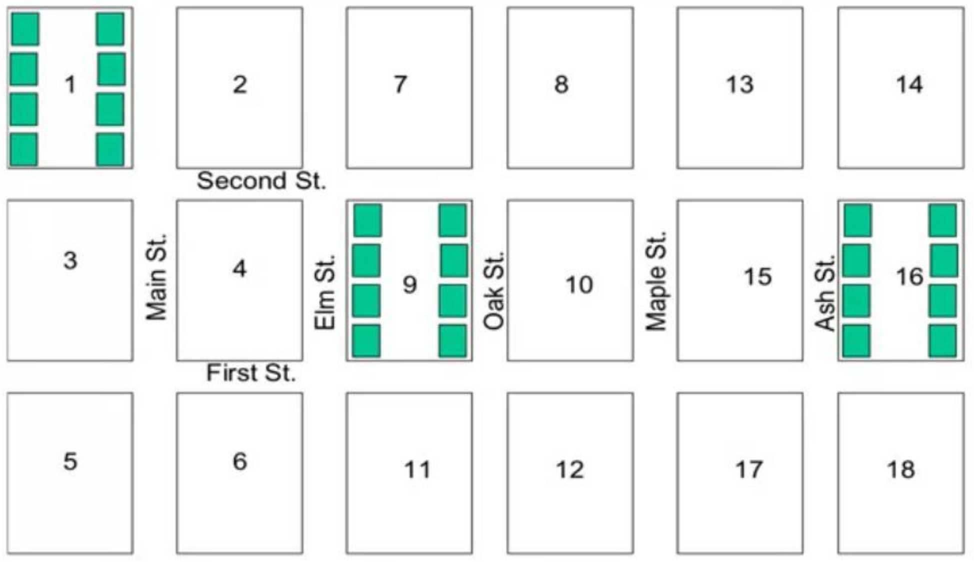
- Resuelve el problema del marco de muestreo: Con frecuencia no es posible tener el marco de muestreo u obtenerlo es costoso, además de que el costo crece al tener que medir unidades separadas entre sí por una gran distancia física.

- En el muestreo por conglomerados el costo se reduce sustancialmente, ya que al levantar la información de elementos contiguos o muy cercanos entre sí se evita el costo de transportación y puede operarse también aun sin tener un marco de muestreo completo

- Suele ser más fácil que el MAS y MS: Usar conglomerados geográficos puede representar importantes ahorros en desplazamientos de personas.

Osval A Montesino, et al. *Muestreo Estadístico*, Ed. Universidad Colima, 2009

Muestreo por Conglomerado

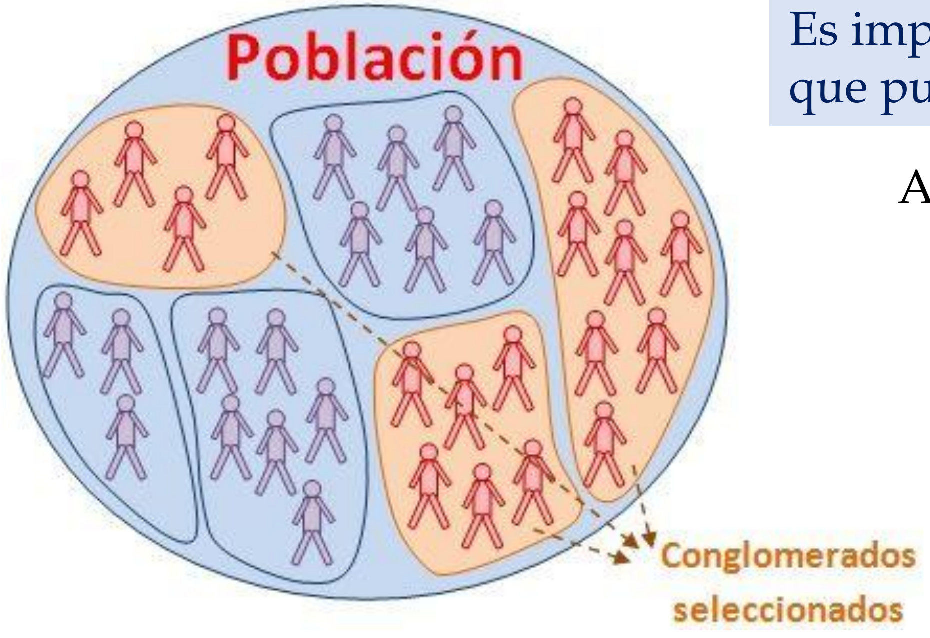


Limitantes Muestreo por Conglomerados

- Como principal inconveniente, al usar muestreo por conglomerados corremos un riesgo importante: que los conglomerados no sean realmente homogéneos entre ellos. Por ejemplo: Podría suceder que un barrio o municipio o ciudad exista una propensión mayor a observar valores de la características de interes en relación a otro barrio o municipio o ciudad sea por razones geográficas, urbanas, culturales, etc.

Osval A Montesino, et al. *Muestreo Estadístico*, Ed. Universidad Colima, 2009

Muestreo por Conglomerado



¿Qué puede ser un conglomerado?

Es importante tener claro lo que será considerado como conglomerados, ya que pueden ser naturales o convenientemente determinados.

Aspectos generales que es necesario que satisfagan los conglomerados:

- Que las unidades que conforman cada conglomerado sean lo más diferentes entre sí, y que estén lo más próximo posible unas de otras, es decir, que las unidades dentro de cada conglomerado sean lo más heterogéneas y cercanas entre sí.

- Que los conglomerados sean lo más similares entre sí, es decir, homogéneos.

Ej: si en un municipio, ciudad deseamos conocer cierto parámetro, los conglomerados podrían agrupar manzanas o barrios

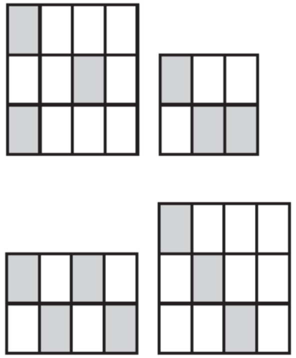
La decisión se toma de acuerdo con la precisión que se quiera, la información disponible, los objetivos o cualquier criterio de interés para el investigador.

Osval A Montesino, et al. *Muestreo Estadístico*, Ed. Universidad Colima, 2009

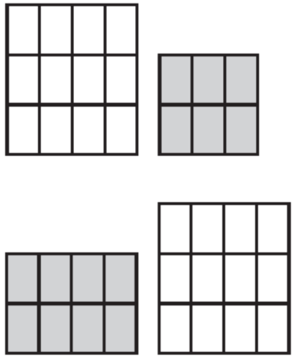
Muestreo por Conglomerado

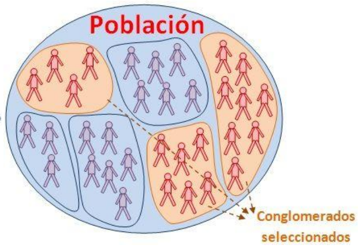
Comparación Muestreo Estratificado y Muestreo por Conglomerado

Muestreo Estratificado



Muestreo por Conglomerado

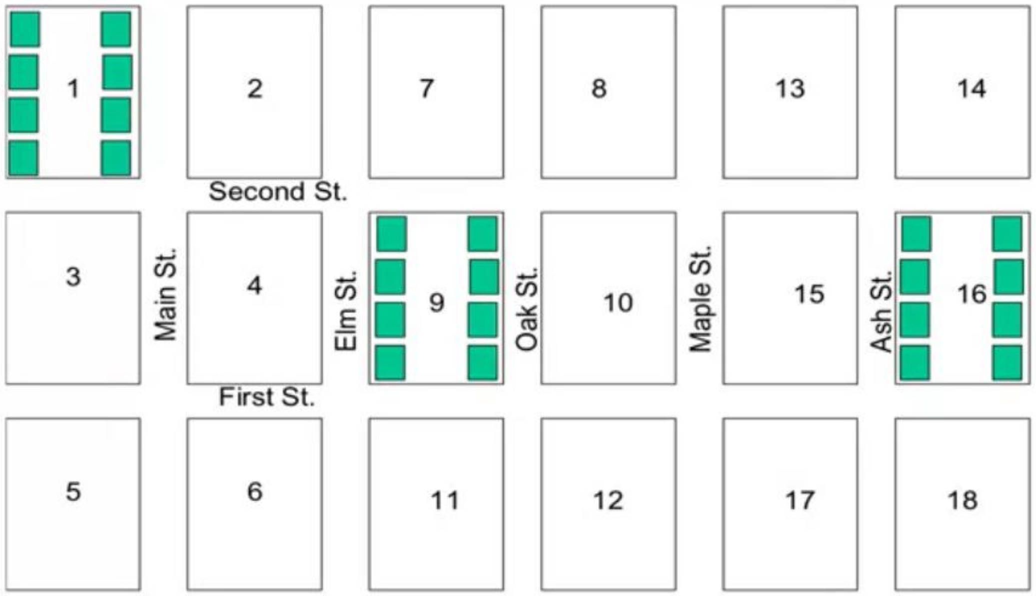




Muestreo estratificado	Muestreo por conglomerados
Mayor precisión con relación al muestreo simple aleatorio.	Mayor precisión con relación al muestreo simple aleatorio.
Los estratos deben contener elementos que sean muy homogéneos entre sí.	Los conglomerados deben contener elementos lo más heterogéneo posible entre sí.
Para obtener mayor precisión, la diferencia debe ser grande entre estratos.	Para mayor precisión, los conglomerados deben ser muy similares.
La varianza de la estimación de la media depende de la variabilidad de los valores dentro del estrato.	La varianza de la estimación de la media depende de la variabilidad que existe entre las medias de los conglomerados.

Osval A Montesino, et al. *Muestreo Estadístico*, Ed. Universidad Colima, 2009

Muestreo por Conglomerado de una Etapa



Muestreo por Conglomerado de una Etapa

Se asume que todos los elementos incluidos en los conglomerados seleccionados y que constituyen la muestra serán estudiados.

Osval A Montesino, et al. *Muestreo Estadístico*, Ed. Universidad Colima, 2009

Muestreo por Conglomerado de una Etapa

Notación

N: el número de conglomerados en la población o unidades de muestreo primarias (UMP) que cubre a toda la población, sin traslapes.

n: el número de conglomerados seleccionados de una muestra simple aleatoria.

M_i: el número de elementos o Unidades de Muestreo Secundarias (UMS) en el conglomerado, $i = 1, 2, \ldots, N$.

$M = \sum_{i=1}^{N} M_i$: el número de elementos o Unidades de Muestreo Secundarias en la población.

$\bar{M}$: el número promedio de UMS por UMP (o conglomerado) en la población.

τ_i: el total del conglomerado i.

$\bar{y}_i = \frac{\sum_{i=1}^{M_i} y_{ij}}{M_i}$: la media a nivel de UMS del conglomerado i.

$\bar{y} = \frac{\sum_{i=1}^{N} y_i}{N}$: el total del promedio por UMP.

$\tau = \sum_{i=1}^{N} \tau_i = \sum_{i=1}^{N} \sum_{j=1}^{M_i} y_{ij}$: el total de la población.

$\mu = \frac{\tau}{M}$: la media a nivel de UMS.

y_{ij}: el valor de la j-ésima UMS en el i-ésimo conglomerado.

Osval A Montesino, et al. *Muestreo Estadístico*, Ed. Universidad Colima, 2009

ESTIMACIÓN DE PARÁMETROS CON M CONOCIDA BAJO MUESTREO POR CONGLOMERADO

Estimación de la MEDIA con M conocida

Los estimadores suponiendo una muestra aleatoria de n conglomerados y que cada uno contiene Mi elementos

Estimador para la Media Poblacional:

$$\hat{\mu} = \bar{y}_s = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} M_i} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{M_i} y_{ij}}{\sum_{i=1}^{n} M_i}$$

Estimador de la varianza de la Media :

$$\hat{V}(\bar{y}_s) = \left(\frac{N-n}{N}\right)\left(\frac{1}{n}\right)\left(\frac{1}{\bar{M}^2}\right)\left(\frac{\sum_{i=1}^{n}(y_i - \bar{y}_s M_i)^2}{n-1}\right)$$

Nota: los estimadores del promedio son de UMS en toda la población. Si se sustituye n por N se obtendría el parámetro μ . El estimador de la varianza es sesgado. pero pueden ser aceptables si n es grande (digamos n > 30) y el sesgo desaparecería si los tamaños de los conglomerados fueran iguales (todas las Mi iguales).

ESTIMACIÓN DE PARÁMETROS CON M CONOCIDA BAJO MUESTREO POR CONGLOMERADO

Estimación Intervalo de Confianza de la MEDIA

El IC suponiendo una muestra aleatoria de n conglomerados y que cada uno contiene Mi elementos

Error de Muestreo: $t_{(n-1,1-\alpha/2)}\sqrt{\hat{V}(\bar{y}_s)}$

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 MEDIA :

$$\boxed{\bar{y}_s \pm t_{(n-1,1-\alpha/2)}\sqrt{\hat{V}(\bar{y}_s)}}$$

ESTIMACIÓN DE PARÁMETROS CON M CONOCIDA BAJO MUESTREO POR CONGLOMERADO

Estimación del TOTAL con M conocida

Los estimadores suponiendo una muestra aleatoria de n conglomerados y que cada uno contiene Mi elementos

Estimador para el TOTAL Poblacional:

$$\boxed{\hat{\tau}_s = M\bar{y}_s = M\left(\frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} M_i}\right)}$$

Estimador de la varianza el TOTAL:

$$\begin{aligned}\hat{V}(\bar{\tau}_s) &= \hat{V}(M\bar{\tau}_s) = M^2\hat{V}(\bar{y}_s)\\ &= M^2\left(\frac{N-n}{N}\right)\left(\frac{1}{n}\right)\left(\frac{1}{\bar{M}^2}\right)\frac{\sum_{i=1}^{n}(y_i-\bar{y}_sM_i)^2}{n-1}\\ &= (N\bar{M})^2\left(\frac{N-n}{N}\right)\left(\frac{1}{n}\right)\left(\frac{1}{\bar{M}^2}\right)\frac{\sum_{i=1}^{n}(y_i-\bar{y}_sM_i)^2}{n-1}\end{aligned}$$

$$\boxed{\hat{V}(\bar{\tau}_s) = N^2\left(\frac{N-n}{N}\right)\left(\frac{1}{n}\right)\frac{\sum_{i=1}^{n}(y_i-\bar{y}_cM_i)^2}{n-1}}$$

Nota: los estimadores del Total son de UMS en toda la población. Si se sustituye n por N se obtendría el parámetro τ. El estimador de la varianza es sesgado. pero pueden ser aceptables si n es grande (digamos $n >$ 30) y el sesgo desaparecería si los tamaños de los conglomerados fueran iguales (todas las Mi iguales).

ESTIMACIÓN DE PARÁMETROS CON M CONOCIDA BAJO MUESTREO POR CONGLOMERADO

Estimación Intervalo de Confianza el TOTAL

El IC suponiendo una muestra aleatoria de n conglomerados y que cada uno contiene Mi elementos

Error de Muestreo: $t_{(n-1,1-\alpha/2)}\sqrt{\hat{V}(\hat{\tau}_s)}$

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 :

$$\boxed{\hat{\tau}_s \pm t_{(n-1,1-\alpha/2)}\sqrt{\hat{V}(\hat{\tau}_s)}}$$

ESTIMACIÓN DE PARÁMETROS CON M CONOCIDA BAJO MUESTREO POR CONGLOMERADO

Tamaño de la Muestra

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

- el parámetro a estimar
- la precisión admissible
- la varianza poblacional y
- el nivel de confianza de la inferencia
- Tamaño de los conglomerados

En el diseño por conglomerados se busca la situación inversa al diseño estratificado, pues formaremos conglomerados homogéneos entre ellos, pero heterogéneos en su interior.

- En algunas ocasiones los conglomerados ya están definidos por algún esquema y no es posible hacerlos más eficientes, lo cual es una desventaja en cuanto a la precisión.
- La anterior condición también puede ser una ventaja, ya que al utilizar un muestreo por conglomerado no requerimos de un marco de muestreo de elementos.

ESTIMACIÓN DE PARÁMETROS CON M CONOCIDA BAJO MUESTREO POR CONGLOMERADO

Tamaño de la Muestra

Para garantizar la precisión deseada:

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

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

donde $\hat{\theta}$ representa el estimador del parámetro de interés. De la expresión anterior y con $\hat{\theta} = \bar{y}_s$, se despeja n para obtener el tamaño de muestra.

ESTIMACIÓN DE PARÁMETROS CON M CONOCIDA BAJO MUESTREO POR CONGLOMERADO

Tamaño de la Muestra para la MEDIA

$$\boxed{n^* = \frac{N t^2_{(n-1,1-\alpha/2)} \sigma_c^2}{N \bar{M}^2 d^2 + t^2_{(n-1,1-\alpha/2)} \sigma_c^2}}$$

donde σ_c^2 es estimada por $s_c^2 = [\sum_{i=1}^{n} (y_i - \bar{y}_s M_i)^2]/(n-1)$.

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO POR CONGLOMERADO

Ejemplo:

Un Colegio tiene $M = 10{,}000$ estudiantes inscritos en $N = 220$ grupos. Con la finalidad de estimar el gasto promedio por estudiante en útiles escolares, se toma una muestra aleatoria simple de $n = 5$ grupos y se pregunta a cada integrante de los grupos sobre su gasto en útiles escolares.

Hallar:

a) La estimación de la media.

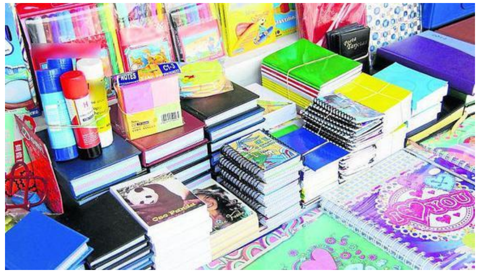
b) La estimación del total.

c) La varianza y la desviación estándar de la media.

d) El IC del 90% para la media poblacional (μ_c).

e) Un IC de 90% para el total.

f) Si $n = 5$ grupos es una muestra preliminar. El tamaño de muestra necesario para estimar el promedio poblacional con una precisión de 4% de la media preliminar y una confiabilidad de 90% es:



Gasto en útiles escolares por estudiante

Grupo 1 ($M_1 = 30$)		Grupo 2 ($M_2 = 32$)		Grupo 3 ($M_3 = 31$)		Grupo 4 ($M_4 = 36$)		Grupo 5 ($M_5 = 34$)	
104	81	107	116	96	102	91	100	113	124
86	78	106	111	108	112	84	104	118	74
114	121	101	93	114	116	70	87	105	122
106	93	97	67	124	101	79	81	96	87
74	114	64	94	103	106	92	101	119	89
125	92	109	79	98	114	131	94	118	132
114	107	97	91	96	94	88	126	113	94
90	114	102	114	103	109	96	102	97	88
98	101	93	109	124	91	99	69	127	134
120	101	121	109	103	96	100	78	119	111
97	98	130	121	105	99	77	122	115	141
99	92	90	112	104	83	69	73	100	91
112		98	103	99	115	83	102	80	127
112		107	79	104		70	123	94	123
104		114		100		81	109	113	136
125		89		110		67	122	128	114
93		89		102		70	90	92	
129		72		107		112	94	82	
$y_1 = 3{,}094$		$y_2 = 3{,}184$		$y_3 = 3{,}238$		$y_4 = 3{,}336$		$y_5 = 3{,}716$	

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO POR CONGLOMERADO

Ejemplo:

Un Colegio tiene $M = 10{,}000$ estudiantes inscritos en $N = 220$ grupos. Con la finalidad de estimar el gasto promedio por estudiante en útiles escolares, se toma una muestra aleatoria simple de $n = 5$ grupos y se pregunta a cada integrante de los grupos sobre su gasto en útiles escolares.

Hallar:

a) La estimación de la media.

$$\hat{\mu} = \bar{y}_s = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} M_i} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{M_i} y_{ij}}{\sum_{i=1}^{n} M_i}$$

donde: $M = 10{,}000,\ n = 5,\ y_i\!: y_1 = 3{,}094,\ y_2 = 3{,}184,\ y_3 = 3{,}238,\ y_4 = 3{,}336,$

$y_5 = 3{,}716,\ M_i\!: M_1 = 30,\ M_2 = 32,\ M_3 = 31,\ M_4 = 36$ y $M_5 = 34$.

$$\bar{y}_s = \frac{3{,}094 + 3{,}184 + 3{,}238 + 3{,}336 + 3{,}716}{30 + 32 + 31 + 36 + 34}$$

$$\bar{y}_s = \frac{16{,}568}{163} = 101.6442.$$

Gasto promedio es de 1’016.442 pesos

Gasto en útiles escolares por estudiante

Grupo 1 ($M_1 = 30$)		Grupo 2 ($M_2 = 32$)		Grupo 3 ($M_3 = 31$)		Grupo 4 ($M_4 = 36$)		Grupo 5 ($M_5 = 34$)	
104	81	107	116	96	102	91	100	113	124
86	78	106	111	108	112	84	104	118	74
114	121	101	93	114	116	70	87	105	122
106	93	97	67	124	101	79	81	96	87
74	114	64	94	103	106	92	101	119	89
125	92	109	79	98	114	131	94	118	132
114	107	97	91	96	94	88	126	113	94
90	114	102	114	103	109	96	102	97	88
98	101	93	109	124	91	99	69	127	134
120	101	121	109	103	96	100	78	119	111
97	98	130	121	105	99	77	122	115	141
99	92	90	112	104	83	69	73	100	91
112		98	103	99	115	83	102	80	127
112		107	79	104		70	123	94	123
104		114		100		81	109	113	136
125		89		110		67	122	128	114
93		89		102		70	90	92	
129		72		107		112	94	82	
$y_1 = 3{,}094$		$y_2 = 3{,}184$		$y_3 = 3{,}238$		$y_4 = 3{,}336$		$y_5 = 3{,}716$	

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO POR CONGLOMERADO

Ejemplo:

Un Colegio tiene $M = 10{,}000$ estudiantes inscritos en $N = 220$ grupos. Con la finalidad de estimar el gasto promedio por estudiante en útiles escolares, se toma una muestra aleatoria simple de $n = 5$ grupos y se pregunta a cada integrante de los grupos sobre su gasto en útiles escolares.

Hallar:

b) La estimación del total.

$$\hat{\tau}_s = M\bar{y}_s$$

donde: $M = 10{,}000$ y $\bar{y}_s = 101.6442.$

$$\hat{\tau}_s = (10{,}000)(101.6442) = 1{,}016{,}442$$

Gasto Total es de 1016’644.200 pesos

Gasto en útiles escolares por estudiante

Grupo 1 ($M_1 = 30$)		Grupo 2 ($M_2 = 32$)		Grupo 3 ($M_3 = 31$)		Grupo 4 ($M_4 = 36$)		Grupo 5 ($M_5 = 34$)	
104	81	107	116	96	102	91	100	113	124
86	78	106	111	108	112	84	104	118	74
114	121	101	93	114	116	70	87	105	122
106	93	97	67	124	101	79	81	96	87
74	114	64	94	103	106	92	101	119	89
125	92	109	79	98	114	131	94	118	132
114	107	97	91	96	94	88	126	113	94
90	114	102	114	103	109	96	102	97	88
98	101	93	109	124	91	99	69	127	134
120	101	121	109	103	96	100	78	119	111
97	98	130	121	105	99	77	122	115	141
99	92	90	112	104	83	69	73	100	91
112		98	103	99	115	83	102	80	127
112		107	79	104		70	123	94	123
104		114		100		81	109	113	136
125		89		110		67	122	128	114
93		89		102		70	90	92	
129		72		107		112	94	82	
$y_1 = 3{,}094$		$y_2 = 3{,}184$		$y_3 = 3{,}238$		$y_4 = 3{,}336$		$y_5 = 3{,}716$	

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO POR CONGLOMERADO

Ejemplo:

Un Colegio tiene $M = 10{,}000$ estudiantes inscritos en $N = 220$ grupos. Con la finalidad de estimar el gasto prom por estudiante en útiles escolares, se toma una muestra aleatoria simple de $n = 5$ grupos y se pregunta a cada integrante de los grupos sobre su gasto en útiles escolares.

Hallar: c) La varianza y la desviación estándar de la media.

$$\hat{V}(\bar{y}_s) = \left(\frac{N-n}{N}\right)\left(\frac{1}{n\bar{M}^2}\right)\sigma_c^2 \quad \text{donde: } M = 10{,}000,\ N = 220,\ n = 5, \bar{y}_s = 101.6442,$$

$\bar{M} = \frac{M}{N} = \frac{10{,}000}{220} = 45.45$: el número promedio de estudiantes por grupo,

σ_c^2: estimada por $s_c^2 = \frac{\left[\sum_{i=1}^{n}(y_i - \bar{y}_s M_i)^2\right]}{n-1}$:

$$s_c^2 = \frac{(3{,}094-(101.6442)(30))^2+\cdots+(3{,}716-(101.6442)(34))^2}{(5-1)} = 46{,}595.$$

$$\hat{V}(\bar{y}_s) = \left(\frac{220-5}{220}\right)\left(\frac{1}{(5)(45.45)^2}\right)\left(\frac{(3{,}094-(101.6442)(30))^2+\ldots+(3{,}716-(101.6442)(34))^2}{5-1}\right)$$

$$\hat{V}(\bar{y}_s) = 4.4079.$$

Desviación estándar: $\sqrt{\hat{V}(\bar{y}_s)} = \sqrt{4.4079} = 2.0995.$

Gasto en útiles escolares por estudiante

Grupo 1 ($M_1 = 30$)		Grupo 2 ($M_2 = 32$)		Grupo 3 ($M_3 = 31$)		Grupo 4 ($M_4 = 36$)		Grupo 5 ($M_5 = 34$)	
104	81	107	116	96	102	91	100	113	124
86	78	106	111	108	112	84	104	118	74
114	121	101	93	114	116	70	87	105	122
106	93	97	67	124	101	79	81	96	87
74	114	64	94	103	106	92	101	119	89
125	92	109	79	98	114	131	94	118	132
114	107	97	91	96	94	88	126	113	94
90	114	102	114	103	109	96	102	97	88
98	101	93	109	124	91	99	69	127	134
120	101	121	109	103	96	100	78	119	111
97	98	130	121	105	99	77	122	115	141
99	92	90	112	104	83	69	73	100	91
112		98	103	99	115	83	102	80	127
112		107	79	104		70	123	94	123
104		114		100		81	109	113	136
125		89		110		67	122	128	114
93		89		102		70	90	92	
129		72		107		112	94	82	
$y_1 = 3{,}094$		$y_2 = 3{,}184$		$y_3 = 3{,}238$		$y_4 = 3{,}336$		$y_5 = 3{,}716$	

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO POR CONGLOMERADO

Ejemplo:

Un Colegio tiene $M = 10{,}000$ estudiantes inscritos en $N = 220$ grupos. Con la finalidad de estimar el gasto prom por estudiante en útiles escolares, se toma una muestra aleatoria simple de $n = 5$ grupos y se pregunta a cada integrante de los grupos sobre su gasto en útiles escolares.

Hallar:

d) El IC del 90% para la media poblacional (μ_c).

$$\bar{y}_s \pm t_{(n-1,1-\alpha/2)}\sqrt{\hat{V}(\bar{y}_s)}$$

donde: $\bar{y}_s = 101.6442$, $t_{(n-1,1-\alpha/2)} = t_{(5-1,0.95)} = 2.1318$ y $\sqrt{\hat{V}(\bar{y}_s)} = 2.0995$.

$$101.6442 \pm (2.1318)(2.0995)$$

$$101.6442 \pm 4.4757$$

$$97.1685 \leq \mu_s \leq 106.1199.$$

Con 90% de confianza se estima que la media poblacional está entre 971685 y 1’061.199 pesos.

Gasto en útiles escolares por estudiante

Grupo 1 ($M_1 = 30$)		Grupo 2 ($M_2 = 32$)		Grupo 3 ($M_3 = 31$)		Grupo 4 ($M_4 = 36$)		Grupo 5 ($M_5 = 34$)	
104	81	107	116	96	102	91	100	113	124
86	78	106	111	108	112	84	104	118	74
114	121	101	93	114	116	70	87	105	122
106	93	97	67	124	101	79	81	96	87
74	114	64	94	103	106	92	101	119	89
125	92	109	79	98	114	131	94	118	132
114	107	97	91	96	94	88	126	113	94
90	114	102	114	103	109	96	102	97	88
98	101	93	109	124	91	99	69	127	134
120	101	121	109	103	96	100	78	119	111
97	98	130	121	105	99	77	122	115	141
99	92	90	112	104	83	69	73	100	91
112		98	103	99	115	83	102	80	127
112		107	79	104		70	123	94	123
104		114		100		81	109	113	136
125		89		110		67	122	128	114
93		89		102		70	90	92	
129		72		107		112	94	82	
$y_1 = 3{,}094$		$y_2 = 3{,}184$		$y_3 = 3{,}238$		$y_4 = 3{,}336$		$y_5 = 3{,}716$	

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO POR CONGLOMERADO

Ejemplo:

Un Colegio tiene $M = 10{,}000$ estudiantes inscritos en $N = 220$ grupos. Con la finalidad de estimar el gasto prom por estudiante en útiles escolares, se toma una muestra aleatoria simple de $n = 5$ grupos y se pregunta a cada integrante de los grupos sobre su gasto en útiles escolares.

Hallar: e) Un IC de 90% para el total.

$$\hat{\tau}_s \pm t_{(n-1,1-\alpha/2)}\sqrt{\hat{V}(\hat{\tau}_s)}$$

$$\sqrt{\hat{V}(\hat{\tau}_s)} = \sqrt{M^2\hat{V}(\bar{y}_s)} = \sqrt{(10{,}000)^2(4.4079)} = 20{,}994.9994.$$

$$1{,}016{,}442 \pm (2.1318)(20{,}994.9994)$$

$$1{,}016{,}442 \pm 44757.1397$$

$$971{,}684.8603 \leq \tau_c \leq 1{,}061{,}199.1397.$$

Con 90% de confianza se estima que la media poblacional está entre 976 850 000 y 1061’200 000 pesos.

Gasto en útiles escolares por estudiante

Grupo 1 ($M_1 = 30$)		Grupo 2 ($M_2 = 32$)		Grupo 3 ($M_3 = 31$)		Grupo 4 ($M_4 = 36$)		Grupo 5 ($M_5 = 34$)	
104	81	107	116	96	102	91	100	113	124
86	78	106	111	108	112	84	104	118	74
114	121	101	93	114	116	70	87	105	122
106	93	97	67	124	101	79	81	96	87
74	114	64	94	103	106	92	101	119	89
125	92	109	79	98	114	131	94	118	132
114	107	97	91	96	94	88	126	113	94
90	114	102	114	103	109	96	102	97	88
98	101	93	109	124	91	99	69	127	134
120	101	121	109	103	96	100	78	119	111
97	98	130	121	105	99	77	122	115	141
99	92	90	112	104	83	69	73	100	91
112		98	103	99	115	83	102	80	127
112		107	79	104		70	123	94	123
104		114		100		81	109	113	136
125		89		110		67	122	128	114
93		89		102		70	90	92	
129		72		107		112	94	82	
$y_1 = 3{,}094$		$y_2 = 3{,}184$		$y_3 = 3{,}238$		$y_4 = 3{,}336$		$y_5 = 3{,}716$	

ESTIMACIÓN DE PARÁMETROS MUESTREO POR CONGLOMERADO

Ejemplo:

Un Colegio tiene $M = 10{,}000$ estudiantes inscritos en $N = 220$ grupos. Con la finalidad de estimar el gasto prom por estudiante en útiles escolares, se toma una muestra aleatoria simple de $n = 5$ grupos y se pregunta a cada integrante de los grupos sobre su gasto en útiles escolares.

Hallar: f) Si $n = 5$ grupos es una muestra preliminar. El tamaño de muestra necesario para estimar el promedio poblacional con una precisión de 4% de la media preliminar y una confiabilidad de 90% es:

$$n^* = \frac{N t^2_{(n-1,1-\alpha/2)} \sigma_c^2}{N \bar{M}^2 d^2 + t^2_{(n-1,1-\alpha/2)} \sigma_c^2}$$

donde: $N = 220,\ n = 5,\ \bar{M} = \frac{M}{N} = 45.45,\ t_{(n-1,1-\alpha/2)} = 2.1318$

σ_c^2: estimada por $s_c^2 = 46{,}595$ y $d = (0.04)(101.6442) = 4.0658.$

$$n^* = \frac{(220)(2.1318)^2(46{,}595)}{(220)(45.45)^2(4.0658)^2 + (2.1318)^2(46{,}595)}$$

$$n^* = \frac{46{,}585{,}945.3241}{7{,}724{,}219.5966} \qquad n^* = 6.0311 = 7 \text{ grupos.}$$

Gasto en útiles escolares por estudiante

Grupo 1 ($M_1 = 30$)		Grupo 2 ($M_2 = 32$)		Grupo 3 ($M_3 = 31$)		Grupo 4 ($M_4 = 36$)		Grupo 5 ($M_5 = 34$)	
104	81	107	116	96	102	91	100	113	124
86	78	106	111	108	112	84	104	118	74
114	121	101	93	114	116	70	87	105	122
106	93	97	67	124	101	79	81	96	87
74	114	64	94	103	106	92	101	119	89
125	92	109	79	98	114	131	94	118	132
114	107	97	91	96	94	88	126	113	94
90	114	102	114	103	109	96	102	97	88
98	101	93	109	124	91	99	69	127	134
120	101	121	109	103	96	100	78	119	111
97	98	130	121	105	99	77	122	115	141
99	92	90	112	104	83	69	73	100	91
112		98	103	99	115	83	102	80	127
112		107	79	104		70	123	94	123
104		114		100		81	109	113	136
125		89		110		67	122	128	114
93		89		102		70	90	92	
129		72		107		112	94	82	
$y_1 = 3{,}094$		$y_2 = 3{,}184$		$y_3 = 3{,}238$		$y_4 = 3{,}336$		$y_5 = 3{,}716$	

ESTIMACIÓN DE PARÁMETROS BAJO MUESTREO POR CONGLOMERADO

EJERCICIO:

MinSalud cuenta con 200 hospitals públicos distribuidos en el territorio nacional, dentro de los cuales tiene laborando a 6,000 médicos con especialidades. Debido al COVID 19 se ha incrementando los niveles de estres en la comunidad médica. Para medir el nivel de satisfacción de los médicos se toma una muestra aleatoria simple de seis hospitales y un censo en cada uno. El nivel de satisfacción se mide de 0 (nada satisfecho) a 10 (muy satisfecho)

Hallar:

a) Intervalo de Confianza del promedio del nivel de satisfación de los médicos al 95%

b) Intervalo de Confianza de para el Total de medicos 95%.

c) Basado en el anterior estimación: ¿Cuál es el tamaño de muestra para estimar la media poblacional de tal manera que sean estimados con una precisión de 5% de la media, una confiabilidad de 95%?

Nivel de Satisfación de los Médicos

Hospital 1			**Hospital 2**			**Hospital 3**			**Hospital 4**			**Hospital 5**			**Hospital 6**		
6	9	6	9	8	8	8	9	8	8	9	8	6	8	9	9	8	6
7	7	6	8	7		7	10	8	8	7	9	6	7		8	7	6
8	8	7	7	9		6	9	6	7	8	8	9	6		7	6	4
7	7	7	8	8		5	9	7	9	9	8	8	9		6	7	8
8	6	7	6	7		9	7	8	8	6		9	8		5	8	8
7	5	6	5	6		6	8	6	8	5		6	10		4	9	7
9	8	6	9	6		5	8	7	9	9		7	10		8	9	6
6	7	5	7	6		8	7		8	8		9	9		9	9	9
6	4	9	6	7		9	6		8	7		8	8		9	8	
6	7	10	6	7		7	6		7	6		7	7		7	8	
7	6	8	8	8		9	6		6	5		6	6		7	7	
8	6		8	8		8	7		5	4		9	5		9	7	