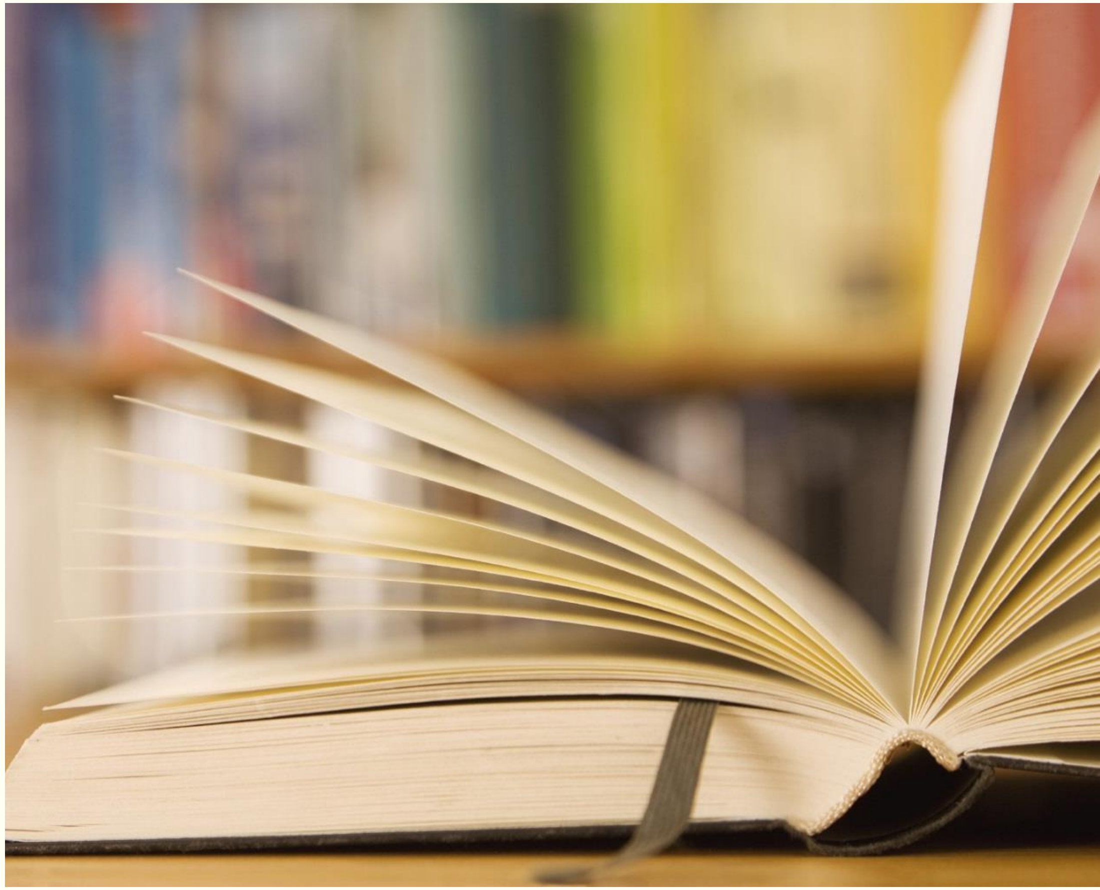
UNIVERSIDAD NACIONAL DE CHIMBORAZO

FACULTAD DE INGENIERÍA

CARRERA DE INGENIERÍA EN TECNOLOGÍAS DE LA INFORMACIÓN

ASIGNATURA: Probabilidad y Estadística

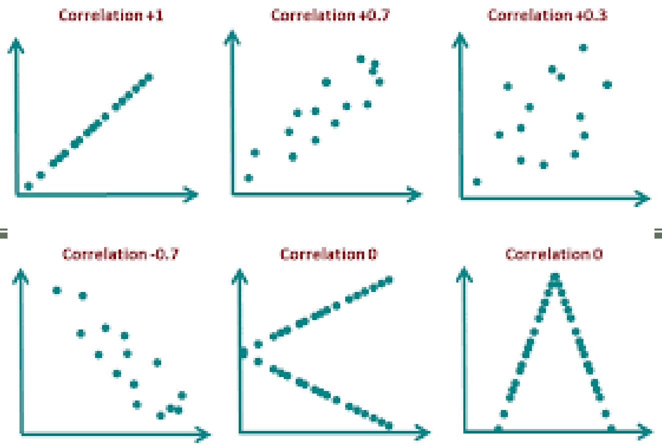
DOCENTE: Ing. Lidia Castro M.Sc



UNIDAD 3 → ESTADÍSTICA INFERENCIAL: ANÁLISIS DE REGRESIÓN Y CORRELACIÓN

OBJETIVO DE LA UNIDAD

Identificar problemas que requieren de análisis estadístico para su entendimiento, aplicando los análisis de regresión y correlación.



1. Teoría del muestreo.
2. Estimación de Parámetros.
3. Regresión y Correlación.

Introducción

Población $\rightarrow$ Muestras $\rightarrow$ Estimadores $\rightarrow$ Calidad

Objetivo

- Conocer una característica θ
- de una población de N individuos
- a partir de n muestras
- con una calidad determinada

Población

Población: Conjunto de elementos sobre los que se quiere obtener una determinada característica

Población → Muestras → Estimadores → Calidad

Parámetros

- Media poblacional

$$\mu = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

- Proporción poblacional

$$P = \frac{1}{N}\sum_{i=1}^{N} a_i,\ a_i = 1 \text{ si el elemento tiene la característica}$$

$N \uparrow\uparrow$

- Coste $\uparrow\uparrow$
- Calidad $\uparrow\uparrow$

Muestreo: $n < N$

- Coste: objetivo $\downarrow\downarrow$
- Calidad: objetivo $\uparrow\uparrow$

Muestras

Muestreo: Selección de n elementos de la población

Población → **Muestras** → Estimadores → Calidad

Probabilísticos

Todos los elementos tienen una probabilidad de ser seleccionados

- Aleatorio simple
- Estratificado
- Por conglomerados
- Sistemático

No probabilístico

No todos los elementos tienen una probabilidad de ser seleccionados (preferencias, subjetividad)

- Cuotas
- Intencional
- Bola de nieve
- Accidental

Estimadores

Población → Muestras → **Estimadores** → Calidad

Queremos conocer la característica de la población de tamaño N a partir de n muestras.

$$\boxed{\text{Estimador } \hat{\theta}}$$

El estimador es una función matemática que depende de la característica a conocer y del muestreo elegido.

Estimadores a utilizar

- Estimador de la media poblacional: $\hat{\mu}$
- Estimador de la proporción poblacional que cumple una característica: $\hat{P}$
- Estimador de la población que cumple una característica: $\hat{A}$

Intervalos de confianza

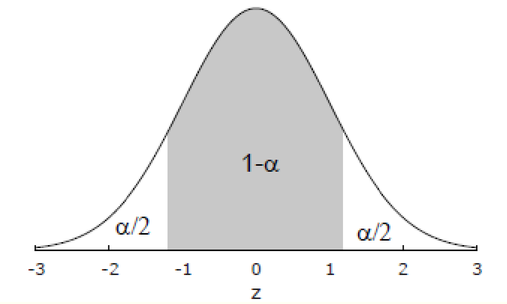
Población → Muestras → Estimadores → Calidad

Intervalo de confianza

$$\hat{\theta} \rightarrow N\left(\theta, V\left(\hat{\theta}\right)\right) \longrightarrow \frac{\hat{\theta} - \theta}{\sqrt{V\left(\hat{\theta}\right)}} \rightarrow N(0,1)$$

$$P\left\{a < \frac{\hat{\theta} - \theta}{\sqrt{V\left(\hat{\theta}\right)}} < b\right\} = 1 - \alpha$$

$$\theta \in \left(\hat{\theta} \pm k\hat{V}\left(\hat{\theta}\right)\right)$$



Calidad

Población → Muestras → Estimadores → **Calidad**

La calidad de los estimadores se mide para determinar el error cometido al realizar un determinado muestreo.

Error del muestreo e_m y error de la estimación e

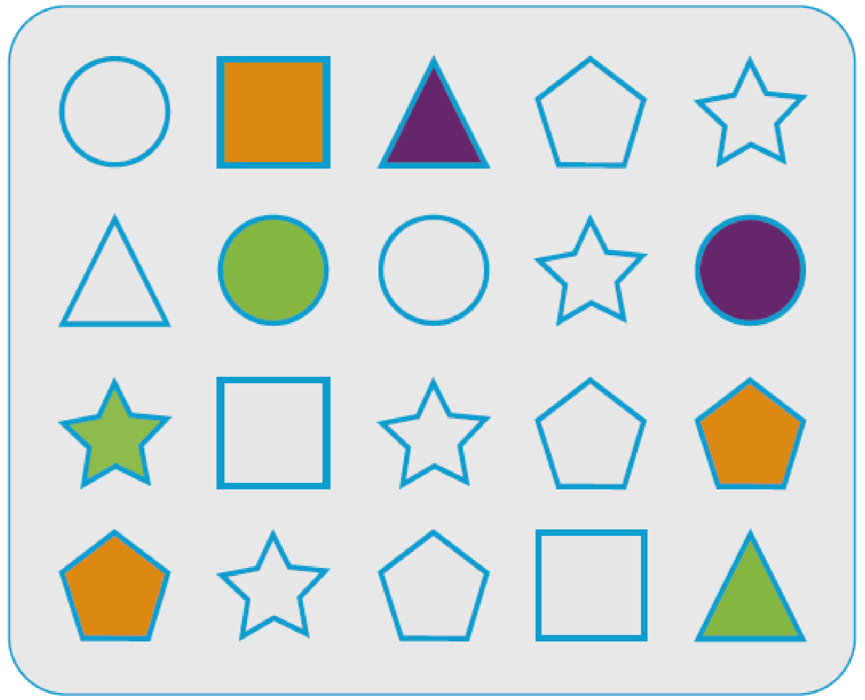
$$e_m = \hat{V}(\hat{\theta}), e = k\sqrt{\hat{V}(\hat{\theta})},$$

donde k representa el nivel de confianza y $\hat{V}(\hat{\theta})$ es el estimador de la varianza del estimador $\hat{\theta}$.

$1-\alpha$	0.68	0.90	0.95	0.955	0.99
k	1	1.64	1.96	2	2.58

Muestreo aleatorio simpe

- Selección aleatoria de las muestras



Muestreo aleatorio simple

- **Definición:** Muestreo aleatorio
- **Características:** Planteamiento sencillo, implementación complicada, costo alto.
- Combinación de muestras elegibles:

$$\binom{N}{n} = \frac{N!}{(N-n)!n!}$$

- Probabilidad de cada muestra

$$\frac{n}{N}$$

Muestreo aleatorio simple – Estimación media poblacional

- Estimador de la media poblacional ⟹ $$\hat{\mu} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i .$$
- Error de muestreo ⟹ $$\hat{V}(\hat{\mu}) = \hat{V}(\bar{y}) = \frac{s^2}{n}\left(\frac{N-n}{N-1}\right) \xrightarrow{N\uparrow} \frac{s^2}{n},$$ $$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 .$$
- Intervalo de confianza ⟹ $$\mu \in \left(\bar{y} \pm k\sqrt{\hat{V}(\bar{y})}\right)$$
- Selección del tamaño de la muestra ⟹ $$e = k\sqrt{\hat{V}(\bar{y})} = k\sqrt{\frac{s^2}{n}\frac{N-n}{N}} \leftrightarrow n = \frac{k^2 s^2}{\frac{k^2 s^2}{N-1} + \frac{e^2(N-1)}{N}} \xrightarrow{N\uparrow} \frac{k^2 s^2}{e^2}.$$

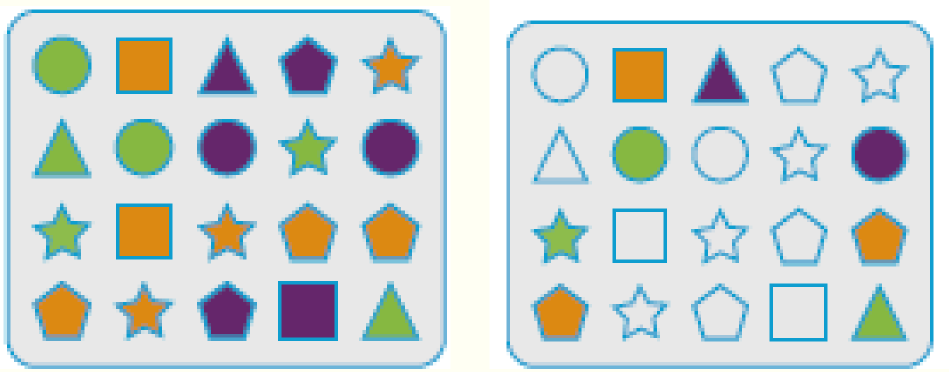
Ejemplo 1

Estimación de la media del número de ángulos a partir del muestreo aleatorio simple, con un 95.5% de confianza (k = 2)

Para ello se tienen los siguientes valores

$\mu = 3{,}55$

$\sigma^2 = 3{,}6475$



Muestreo aleatorio simple – Estimación proporción poblacional

Concepto: En una población de tamaño N, A elementos tienen una característica y N-A elementos no la tienen. Cada elemento i estaría determinado por a_i y b_i de la siguiente manera:

a_i	b_i	Descripción
1	0	El elemento i tiene la característica
0	1	El elemento i no tiene la característica

La proporción poblacional P es la proporción de elementos que contienen la característica.

$$P = \frac{A}{N} = \frac{1}{N}\sum_{i=1}^{N} a_i .$$

Llamamos Q a la proporción de elementos que no contienen la característica, siendo:

Q = 1-P.

La proporción muestral p se calcula como:

$$p = \frac{1}{n}\sum_{i=1}^{n} a_i = \frac{1}{n}a.$$

Muestreo aleatorio simple – Estimación proporción poblacional

- Estimadores de la proporción poblacional y del número total de individuos

$$\hat{P} = p = \frac{1}{n}\sum_{i=1}^{n} a_i = \frac{a}{n}, \qquad \hat{A} = Np = \frac{N}{n}a.$$

- Error de muestreo

$$\hat{V}(\hat{P}) = \left(\frac{N-n}{N-1}\right)\frac{pq}{n-1}, \qquad \hat{V}(\hat{A}) = N^2\hat{V}(\hat{P}) = N^2\left(\frac{N-n}{N-1}\right)\frac{pq}{n-1}.$$

- Intervalos de confianza

$$P \in \left(\hat{P} \pm k\sqrt{\hat{V}(\hat{P})}\right), A \in \left(\hat{A} \pm k\sqrt{\hat{V}(\hat{A})}\right).$$

- Selección del tamaño de la muestra

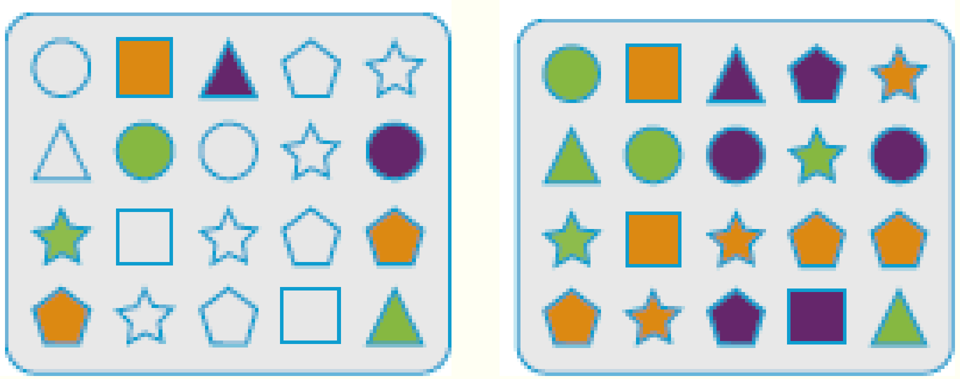
$$n = \frac{\frac{N}{N-1}PQ}{D + \frac{1}{N+1}PQ} \xrightarrow{N\uparrow} \frac{PQ}{D}.$$

Ejemplo 2.

Estimación de la proporción de figuras con un número de ángulos mayor o igual a 4 a partir del muestreo aleatorio simple, con un 95.5% de confianza (k = 2).

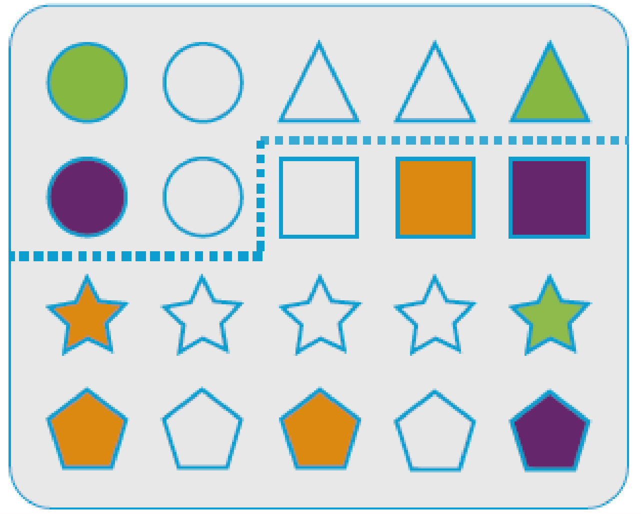
P = 0,65

A = 13



Muestreo estratificado

- Agrupamos los elementos por estratos
- Seleccionamos aleatoriamente elementos de cada estrato



Muestreo estratificado

La población de N elementos se divide en L estratos de tamaño N_j lo más homogéneas posibles.

Características: La muestra se obtiene más rápida e incluso más precisa.

Minimizar errores: Varianza intra-estrato mínima, varianza inter-estrato máxima.

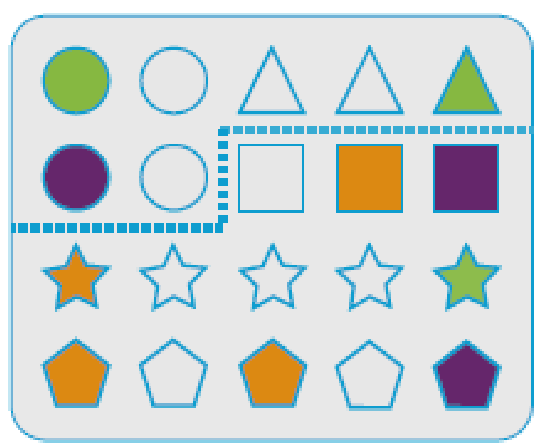
Usos

- Se desea resaltar un subgrupo específico de la población
- Observar relaciones entre dos o más grupos
- Se puede probar de forma representativa a los subgrupos más inaccesibles

Muestreo estratificado

- Peso de cada estrato:

$$W_j = \frac{N_j}{N}.$$



- $N = 20, L = 2$
- $N_1 = 7, N_2 = 13$
- $W_1 = 0.35, W_2 = 0.65$

Muestreo estratificado - Estimación media poblacional

- Estimador de la media poblacional

$$\hat{\mu} = \bar{y}_{st} = \sum_{j=1}^{L} W_j \bar{y}_j.$$

- Error de muestreo

$$\hat{V}(\bar{y}_{st}) = \sum_{j=1}^{L} W_j^2 \hat{V}(\bar{y}_j).$$

- Intervalo de confianza

$$\mu \in \left(\bar{y}_{st} \pm k\sqrt{\hat{V}(\bar{y}_{st})} \right).$$

Muestreo estratificado - Estimación media poblacional

Selección del tamaño de la muestra: Dependerá de la forma en la que se reparten los estratos (afijación)

- **Afijación uniforme** Los elementos de la muestra se reparten por igual en todos los estratos: $n_j = \frac{n}{L}.$

$$n = \frac{L\sum_{j=1}^{L} W_j^2 s_j^2}{D + \frac{1}{N}\sum_{j=1}^{L} W_j s_j^2}.$$

- **Afijación proporcional** Los elementos de la muestra se reparten proporcionalmente al tamaño del estrato: $\frac{n_j}{N_j}$ =constante.

$$n = \frac{\sum_{j=1}^{L} W_j s_j^2}{D + \frac{1}{N}\sum_{j=1}^{L} W_j s_j^2}.$$

Muestreo estratificado - Estimación media poblacional

- **Afijación óptima** Los elementos de la muestra se reparten de forma que la varianza de los estimadores sea mínima: $n_j = n \frac{N_j s_j}{\sum_{j=1}^{L} N_j s_j}$

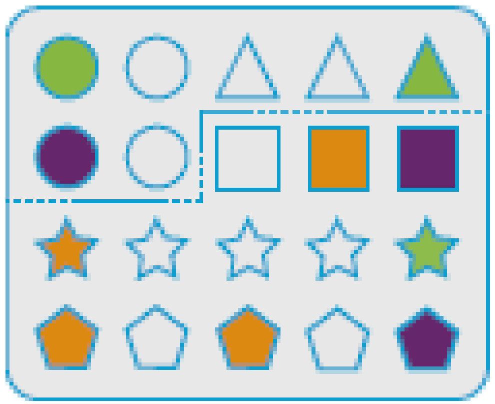
$$n = \frac{\left(\sum_{j=1}^{L} W_j s_j^2\right)^2}{D + \frac{1}{N} \sum_{j=1}^{L} W_j s_j^2}.$$

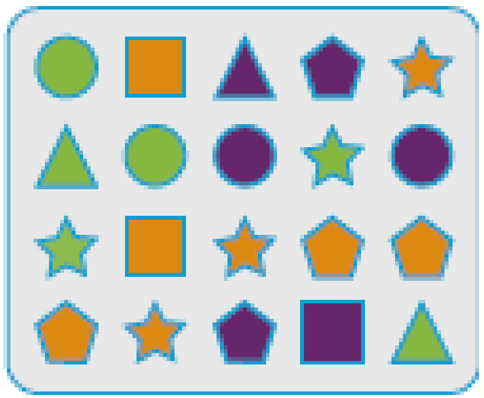
Ejemplo 3

Estimación de la media del número de ángulos a partir del muestreo estratificado, con un 95.5% de confianza (k = 2)

$\mu = 3{,}55$

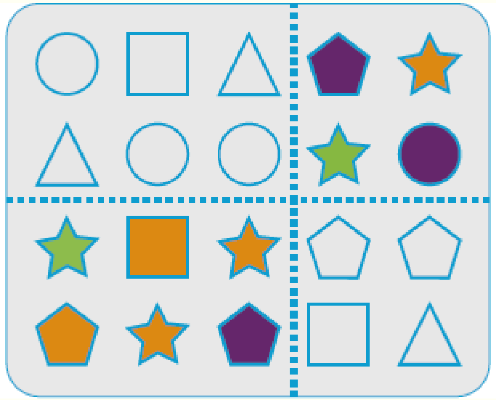
$\sigma^2 = 3{,}6475$

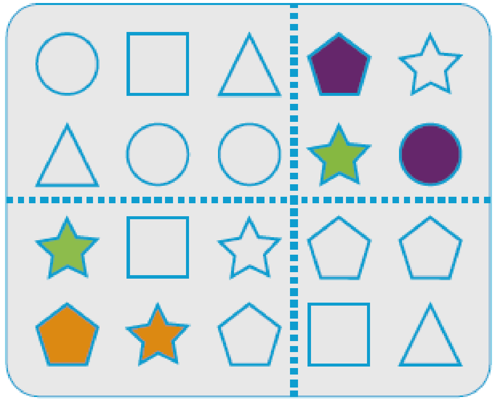




Muestreo por conglomerados

- Obtención de conglomerados.
- Mono-etápico: trabajamos con todas las muestras de los conglomerados seleccionados.
- Bi-etápico: trabajamos con algunas muestras de los conglomerados seleccionados.





Muestreo por conglomerados monoetápico

La población de M elementos se divide en N conglomerados lo más heterogéneos posibles. Se seleccionan aleatoriamente n conglomerados, de tamaño m_i .

Características Cada conglomerado se parece mucho a la población. Bajo coste.

Minimizar errores Varianza intra-clúster máxima, varianza inter-clúster mínima.

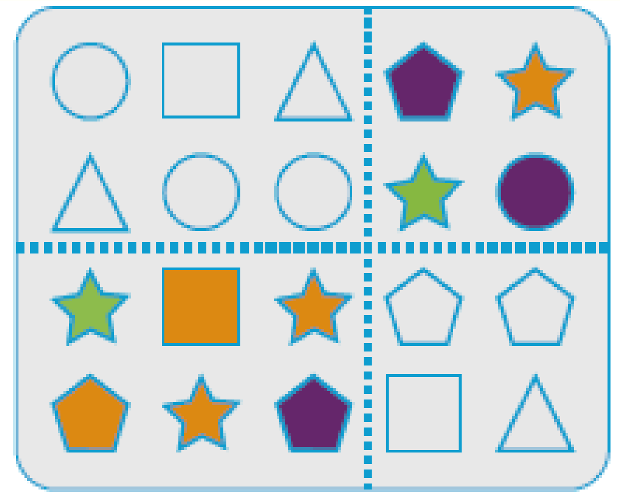
Usos

- Los elementos de la población son muy dispares
- No es posible un muestreo aleatorio

Muestreo por conglomerados monoetápico

- Tamaño medio del clúster:

$$\bar{M} = \frac{M}{N}.$$



- $M = 20, N = 4, n = 2$
- $\bar{M} = M/N = 5$
- $m_1 = 4, m_2 = 6$

Muestreo por conglomerados monoetápico - Estimación media poblacional

- Estimador de la media poblacional

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} m_i}, x_i = \sum_{j=1}^{n} y_{ij}.$$

- Error de muestreo

$$\hat{V}(\hat{\mu}) = \left(1 - \frac{n}{N}\right)\left(\frac{1}{n\bar{M}^2}\right)\frac{1}{n-1}\sum_{i=1}^{n}(x_i - m_i\bar{x})^2$$

- Intervalo de confianza

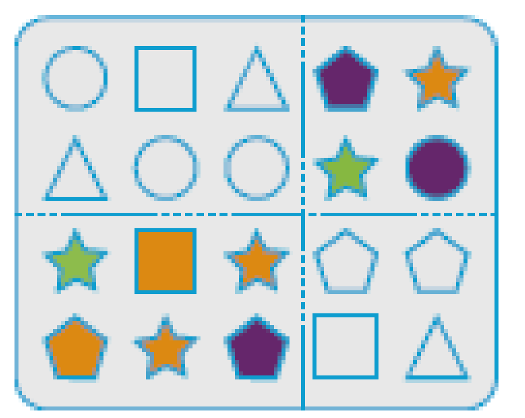
$$\mu \in \left(\bar{x} \pm k\sqrt{\hat{V}(\bar{x})}\right).$$

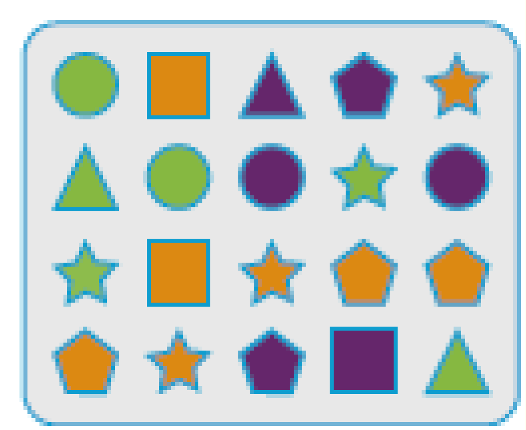
- Tamaño de la muestra

$$n = \frac{s_c^2}{D\bar{M}^2 + \frac{s_c^2}{N}}, s_c^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - m_i\bar{x})^2.$$

Ejemplo 4

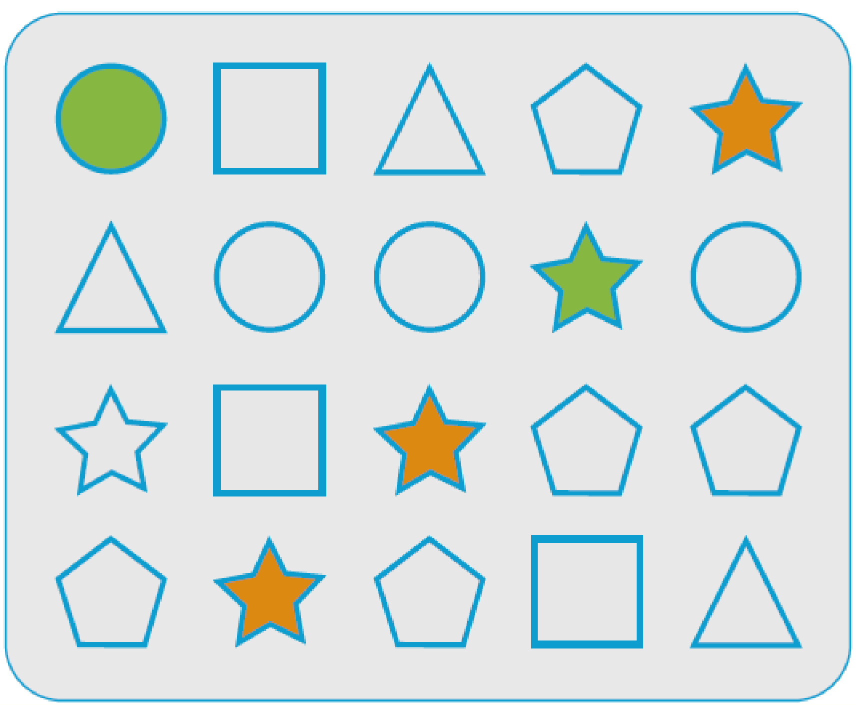
Estimación de la media del número de ángulos a partir del muestreo por conglomerados monoetápico, con un 95.5% de confianza (k = 2)





Muestreo sistemático

Del conjunto ordenado tomamos 1 muestra de cada t (en este caso, t = 4)



Muestreo sistemático

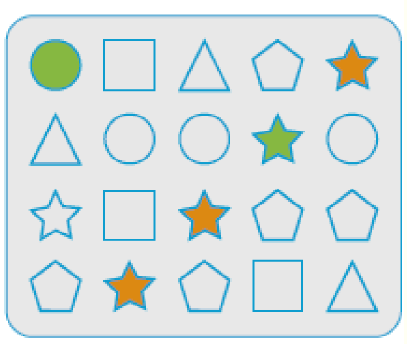
De los N elementos de la población, se seleccionan n elementos de t en t, empezando desde el elemento R.

Características: Ningún elemento de un subgrupo grande queda sin representación. Se puede combinar con el muestreo por estratos o por conglomerados.

Usos

- Control de calidad en cadenas de montaje

- $N = 20, n = 5$
- $R = 1, t = 4$



Muestreo sistemático - Estimación media poblacional

- Estimador de la media poblacional

$$\hat{\mu} = \bar{y}_{sy} = \frac{1}{n}\sum_{j=1}^{n} y_{R+(j-1)t}.$$

- Calidad del estimador

$$\hat{V}(\hat{\mu}) = \hat{V}(\bar{y}) = \frac{s^2}{n}\left(\frac{N-n}{N-1}\right), s^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(y_{R+(j-1)t} - \bar{y}_{sy}\right)^2.$$

- Población aleatoria: $\hat{V}(\bar{y}_{sy}) \approx \hat{V}(\bar{y}_{as})$.
- Población ordenada: $\hat{V}(\bar{y}_{sy}) \leq \hat{V}(\bar{y}_{as})$.
- Población periódica: $\hat{V}(\bar{y}_{sy}) \geq \hat{V}(\bar{y}_{as})$.

- Intervalo de confianza

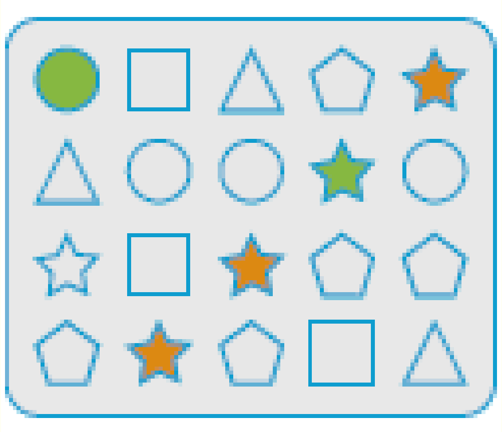
$$\bar{y}_{sy} \pm k\sqrt{\hat{V}(\bar{y}_{sy})}.$$

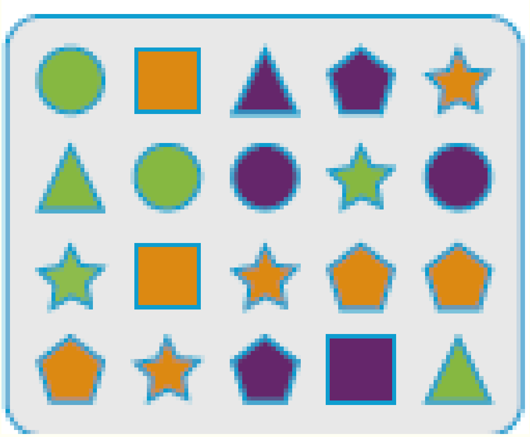
Ejemplo 5

Estimación de la media del número de ángulos a partir del muestreo sistemático, con un 95.5% de confianza (k = 2)

$\mu = 3{,}55$

$\sigma^2 = 3{,}6475$






¡GRACIAS!