

# TECNICA PARA REDUCIR EL ERROR DE MUESTREO

Jorge Troche Luna

Uno de los objetivos de los estudios por muestreo es reducir el intervalo de confianza del estimador de un parámetro. En el muestreo sistemático uno de los más utilizados, puede reducirse dicho intervalo por medio del re-ordenamiento de los elementos de la población adecuadamente con el fin de reducir específicamente la varianza. Para este propósito se considera que tiene que ver mucho la tendencia y comportamiento de la población, que se expresa mediante el orden del marco muestral. Sarndal, Swensson y Wretmam en el libro "Model Assisted Survey Sampling" indican textualmente: La eficiencia del muestreo sistemático depende grandemente del orden particular de los  $N$  elementos en el cual la selección sistemática es aplicada. Es decir esta en función de la población ordenada.

**EI MUESTREO SISTEMÁTICO** ofrece un conjunto de procedimientos ventajosos, particularmente en la selección, donde solo un primer elemento es obtenido en forma aleatoria y con igual probabilidad, de los primeros  $a$  elementos en el listado de la población. El entero positivo  $a$  es fijo llamado intervalo muestral. El resto de la muestra se determina sistemáticamente tomando cada  $a$ -ésimo elemento hasta el fin del listado. Así hay solamente  $a$  posibles muestras, cada una tiene la misma probabilidad  $1/a$  de ser seleccionada. Esta simplificación de solamente un ensayo aleatorio es una gran ventaja. Es decir la muestra esta conformada por:

$$s = \{k: k = r + (j - 1)a \leq N; j = 1, 2, \dots, n\}$$

Donde  $r$  es llamado arranque aleatorio, y  $k$  es la posición del elemento  $k$ -ésimo

En general si  $a$  es un entero positivo, tenemos:

$$N = an + c$$

Por tanto el tamaño de la muestra consideraremos siempre de tamaño  $n$ .

La varianza es:  $V_{S\bar{Y}}(\hat{t}_\pi) = a(a - 1)S_t^2$

donde  $s_t^2 = \frac{1}{a - 1} \sum_{r=1}^a (t_{s_r} - \bar{t})^2$

es la varianza del total muestral. La varianza es pequeña si los totales muestrales son aproximadamente iguales

Para el análisis de varianza, notemos que la sumatoria de la ecuación anterior, es una parte de la siguiente igualdad. Es decir, que la variación total en la población puede ser descompuesto en variación dentro de muestras sistemáticas, y la variación entre muestras sistemáticas, es decir un análisis de varianza. Esto es:

$$SST = SSW + SSB$$

$$SST_{Total} = SSD_{Dentro} + SSD_{Entre}$$

donde  $SST$  representa la suma de cuadrados del total,  $SSW$  representa la suma de cuadrados dentro las muestra y  $SSB$  representa la suma de cuadrados entre muestras. Para una población dada,  $SST = (N - 1) S^2$  y  $U$  es fija. Por tanto, un incremento



en la variación dentro SSW es acompañado por un correspondiente decremento en la variación entre SSB. Además de una constante multiplicativa fija SSB determina la varianza (1) la cual puede ser escrita como

$$V_{SY}(\hat{t}_\pi) = N \cdot SSB$$

En otras palabras, a más homogeneidad dentro de las muestras es menos eficiente el muestreo sistemático. la homogeneidad se usa aqui para conocer la tendencia para tener y-valores iguales. así para realizar un

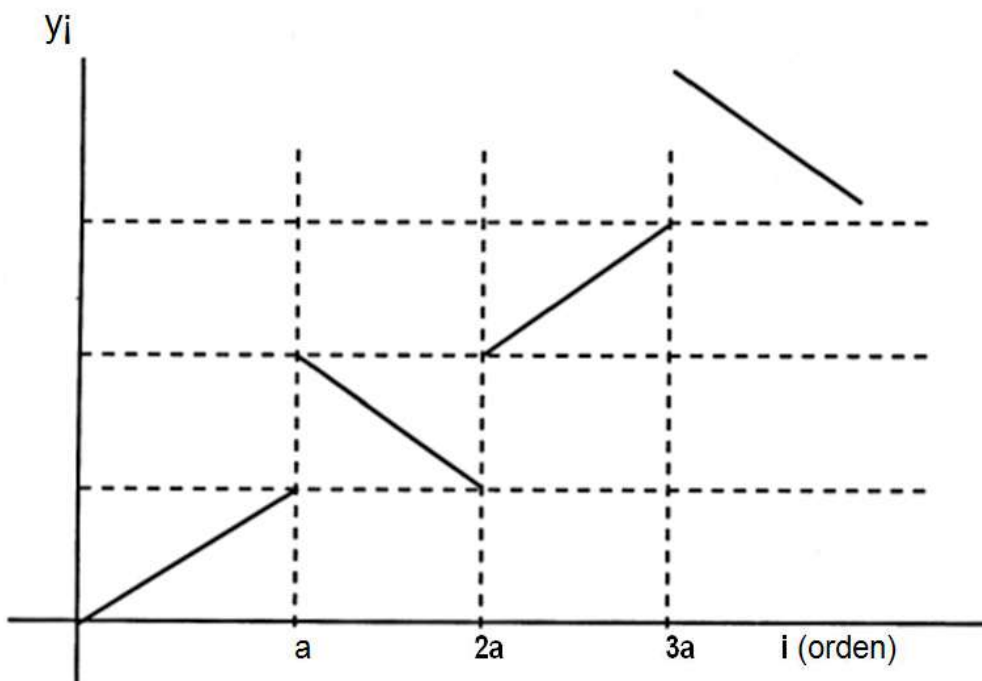
favorable ordenamiento de la población para muestras sistemáticas, podemos hacer lo posible, para un ordenamiento que vincula un bajo grado de homogeneidad alrededor de los elementos en la misma muestra sistemática.

La homogeneidad mediremos mediante:

$$\rho = 1 - \frac{n}{n-1} \frac{SSW}{SST}$$

la cual es llamada el coeficiente de correlación intraclásica.

ORDEN PROPUESTO





Por ejemplo, consideremos un marco muestral ordenado ascendentemente de tamaño 15 y se desea obtener una muestra de tamaño 5, como sigue:

Calculando tenemos que  $a$  (intervalo muestral) es  $a = N/n = 15/5 = 3$ , dando como resultado el siguiente orden propuesto.

Número de la observación	Valor de la variable
1	13
2	14
3	17
4	18
5	19
6	20
7	20
8	21
9	22
10	23
11	24
12	25
13	26
14	26
15	27

Número de la observación	Valor de la variable
1	13
2	14
3	17
4	18
5	19
6	20
7	20
8	21
9	22
10	23
11	24
12	25
13	26
14	26
15	27



### ALGORITMO PARA EL ORDENAMIENTO DEL MARCO MUESTRAL

El algoritmo que se presenta esta con sintaxis computacional (en Fox Pro concretamente), aunque la adaptación del algoritmo a otros medios de almacenamiento de marcos muestrales es similar.

*sort to transí on Edad*

*use transí*

*CLEAR*

*INPUT 'INTRODUZCA EL TAMAÑO DE LA POBLACION:' TO N*

*INPUT 'INTRODUZCA EL TAMAÑO DE LA MUESTRA:' TO M*

*goto 1*

*ini=0*

*k=N/M*

*do while not eofQ*

*for l=ROUND(ini+1t0) to*

*ROUND(ini+k, 0)*

*replace sis te with i*

*skíp*

*endfor*

*b=0*

*for 1= ROUND(ini+k+1,0) to*

*ROUND(iní+2\*k,0)*

*replace síste with ini+2\*k-b*

*skíp*

*b=b+1*

*endfor*

*ini=ini+2\*k*

*enddo*

*sort to sístem on síste*

### HUMOR ESTADÍSTICO

Cómo se ponen las notas

**Carrera de Estadística:** Se colocan los estudiantes por orden alfabético sobre una gráfica distribuida a lo largo de una Gaussiana.

**Carrera de Estadística:** Las notas son variables aleatorias.

**Carrera de Religión:** Dios pone las notas (inapelable)

**Carrera de Informática:** Se usa un generador de números aleatorios

**Carrera de Historia:** Cada Estudiante recibe la misma nota que el año anterior.

El 97.3 % de las estadísticas han sido claramente inventadas