Tema 5: Muestreo sistemático

Contenido

1- Definición, Ventajas y Desventajas, Variantes del muestreo sistemático, selección de una muestra sistemática.
2- Muestreo sistemático como Estratificado o Conglomerado
3- Cuando N no es un múltiplo exacto de n
4- Estimación de Y ($\bar{y}_{sy}$) insesgabilidad y varianza
5- Notas sobre la Varianza y la Precisión
6- Tipos de Población: Aleatoria ordenada y periódica
7- Estimación del total, de la proporción y tamaño de muestra
8- Estimación de la Varianza
9- Muestreo sistemático Replicado

Muestreo sistemático

El muestreo sistemático es un tipo de muestreo alternativo al muestreo aleatorio simple.

Definición

El Muestreo Sistemático consiste en seleccionar aleatoriamente una unidad muestral entre las K primeras y a continuación de modo sistemático o rígido se va seleccionando una unidad muestral cada k unidades hasta completar n.

Así si el primer elemento seleccionado es i ($1 \leq i \leq k$) el segundo será $i + k$ el tercero $i + 2k$ así hasta completar los n elementos al seleccionar el elemento $i + (n-1)k$ de la población.

Hemos supuesto que N es múltiplo de n esto es conocido y k determinado por $k = N/n$. En principio supondremos N/n es un entero.

Ejemplo: Seleccionar una muestra sistemática de n=128 elementos de un archivo de 3.200 items (N=3200).

Muestreo sistemático

Vemos que con este método solo hay k muestras posible y son todos equiprobables, con probabilidad 1/k de ser seleccionada.

Ventajas del muestreo sistemático

- La selección de la muestra es más sencilla y rápida que en el m.a.s, lo cual permite realizarlo en el campo.
- Puede ser más representativo que el aleatorio simple donde sea necesario no dejar segmentos sin representación ya que se extiende uniformemente en toda la población.
- En algunos casos puede ser tan preciso y de menor costo que el m.a.s.

Desventajas del muestreo sistemático

- No se puede estimar la varianza de los estimadores con los datos muéstrales.
- No se puede aplicar indistintamente a todo tipo de población ya que los estimadores puede presentar varianzas que difieren mucho a la varianza del aleatorio simple, que es la disponible.

Muestreo sistemático

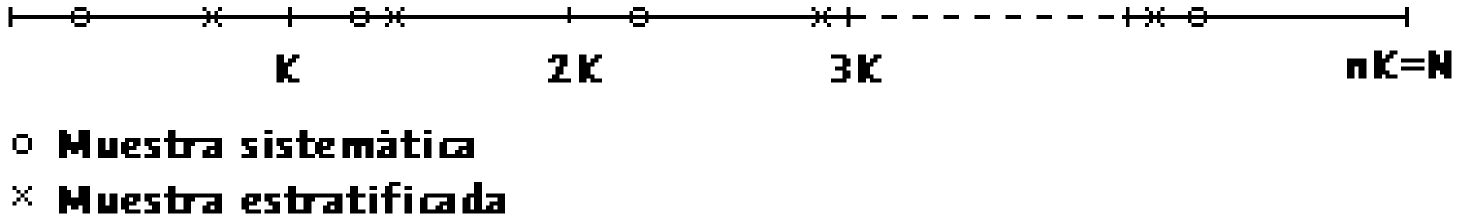
Variantes del sistematico

1-Elegir rígidamente cada unidad muestral en el centro de k (no aleatoria)

2-Sustituir el recuento de k por una medida aproximada.

Muestreo sistemático como estratificado

El Muestreo Sistemático puede asociarse a un estratificado de una población de n estratos, de igual tamaño k, con una unidad seleccionada por cada uno, donde las unidades seleccionadas estan en la misma posición relativa en el estrato.



Muestreo sistemático

Muestreo sistemático como conglomerados

El muestreo sistemático puede también verse como una población de N unidades clasificada en k subpoblaciones con n elementos (conglomerados) cada una donde sólo seleccionamos una y se observan todos sus elementos.

Número de subpoblaciones o conglomerados					
1	**2**	**...**	**i**	**...**	**k**
y_1	y_2	...	y_i	...	y_k
y_{k+1}	y_{k+2}	...	y_{k+i}	...	y_{2k}
:	:		:		:
$y_{(n-1)k+1}$	$y_{(n-1)k+2}$	...	$y_{(n-1)k+i}$	...	y_{nk}
$\bar{y}_1$	$\bar{y}_2$		$\bar{y}_i$		$\bar{y}_k$

Muestreo sistemático

Si N no es múltiplo exacto de n

1. Se toma k como el mayor entero menor que N/n y el residuo r de N/n se distribuye entre los primeros r serán de tamaño k+1 y los restantes n-r de tamaño k.
2. Permitir que la muestra pueda tener un tamaño n o n+1, completar r hasta k y seleccionar el origen i; si $i \leq r$ la muestra será de tamaño n+1, y si $i \rhd r$ será de tamaño n.
3. Eliminar al azar entre los N los r de modo N = nk o llenar con elementos de N al azar, los k-r faltantes y tomar la muestra n+1.
4. Considerar la lista como circular y tomar el arranque al azar entre 1 y N (ahora hay N muestras posibles con probabilidad 1/N) y tomar cada K hasta completar n.
5. Usar intervalos fraccionarios, multiplicando por 10a de modo K=N/n x 10a sea entero y tomar el origen entre 1 – k10a y seguirlo cada 10k.

Muestreo Sistemático

Estimación de la media

Notación: $y_{i,j}$ es el j-esimo elemento de la i-esima muestra sistemática i=1,2,...,k j=1,2...,n

Primera muestra sistemática	y_{11}	y_{12}	...	y_{1j}	...	y_{1n}
Segunda muestra sistemática	y_{21}	y_{22}	...	y_{2j}	...	y_{2n}
	:	:		:		:
i-esima muestra sistemática	y_{i1}	y_{i2}	...	y_{ij}	...	y_{in}
	:	:		:		:
k-esima muestra sistemática	y_{k1}	y_{k2}	...	y_{kj}	...	y_{kn}

$$\overline{y}_{sy} = \frac{\sum y_{ij}}{n} = \overline{y}_{i\bullet}$$

Si N = nk es fácil ver que $\overline{y}_{sy}$ es insesgada, pues:

$$E(\overline{y}_{sy}) = E(\overline{y}_{i\bullet}) = \frac{\sum \overline{y}_{i\bullet}}{k} = \frac{\sum^{k}\sum^{n} y_{ij}}{nk} = \frac{Y}{N} = \overline{Y}$$

Muestreo Sistemático

Varianza de la media

$$V(\bar{y}_{sy}) = \frac{(N-1)}{N} S^2 - \frac{k(n-1)}{N} S^2_{wsy} = \frac{\sum \left(\bar{y}_{i\bullet} - \bar{Y}\right)^2}{k}$$

donde $$S^2 = \frac{\sum\sum \left(y_{ij} - \bar{Y}\right)^2}{N-1}$$ y $$S^2_{wsy} = \frac{\sum\sum \left(y_{ij} - \bar{y}_{i\bullet}\right)^2}{k(n-1)}$$

S^2_{wsy} es la varianza media entre las unidades de la misma muestra sistemática

Muestreo Sistemático

Varianza de la media (continuación)

$$(N-1)S^2=\sum_i^k\sum_j^n\left(y_{ij}-\overline{Y}\right)^2=\sum_i^k\sum_j^n\left(y_{ij}-\overline{y}_{i\bullet}+\overline{y}_{i\bullet}-\overline{Y}\right)^2=$$

$$n\sum_i^k\left(\overline{y}_{i\bullet}-\overline{Y}\right)^2+\sum_i^k\sum_j^n\left(y_{ij}-\overline{y}_{i\bullet}\right)^2=$$

$$(N-1)S^2=nkV(\overline{y}_{sy})+k(n-1)S_{wsy}^2$$

La media de una muestra aleatoria sistemática es más precisa que la de una m.a.s. sii $S_{wsy}^2>S^2$

La varianza de $\overline{y}_{sy}$ se puede expresar: $V(\overline{y}_{sy})=\dfrac{S^2}{n}\dfrac{(N-1)}{N}\left(1+(n-1)\rho_w\right)$

donde ρ_w es el coeficiente de correlación entre pares de unidades que están en la misma muestra sistemática

$$\rho_w=\frac{E\left[\left(y_{ij}-\overline{Y}\right)\left(y_{iu}-\overline{Y}\right)\right]}{E\left[y_{ij}-\overline{Y}\right]^2}=\frac{2}{(n-1)(N-1)S^2}\sum_i^k\sum_{j<u}^n\left(y_{ij}-\overline{Y}\right)\left(y_{iu}-\overline{Y}\right)$$

Muestreo Sistemático

Varianza de la media (continuación)

También se puede escribir como:

$$V(\bar{y}_{sy}) = \frac{S_{wst}^2}{n}\frac{(N-n)}{N}\left(1+(n-1)\rho_{wst}\right)$$

donde

$$S_{wst}^2 = \frac{1}{n(k-1)}\sum\sum(y_{ij}-\bar{y}_{\bullet j})^2$$

$$\rho_{wst} = \frac{E\left[\left(y_{ij}-\bar{y}_{\bullet j}\right)\left(y_{iu}-\bar{y}_{\bullet u}\right)\right]}{E\left[y_{ij}-\bar{y}_{\bullet j}\right]^2}$$

Muestreo Sistemático

Notas sobre la varianza y la precisión

1- La Varianza de $\bar{y}_{sy}$ depende de parámetros que en general no pueden ser estimados con los valores de una sola muestra sistemática $(S^2_{wsy}, S^2_{wst}, \rho_{wsy}, \rho_{wst})$. Ello obliga a utilizar aproximaciones no siempre confiables y estudiar la forma de la población.

2- Al igual que con $V(\bar{y}_{sy})$ sucedera con $V(\hat{Y}_{sy})$ y con $V(p_{sy})$

3- La $\bar{y}_{sy}$ es mas precisa que $\bar{y}$ cuando $S^2_{wsy} > S^2$; esto es el muestreo Sistemático es preciso cuando la muestra sistemática recoge toda la heterogeneidad de la población y es impreciso cuando no reflejan en heterogeneidad.

Muestreo Sistemático

Notas sobre la varianza y la precisión

4. Una muestra sistemática tiene igual precisión que la correspondiente estratificado con una unidad por estrato si $\rho_{wst} = 0$

5. La correlación positiva ρ_w infla la varianza del estimador
   - Si $\rho_w \approx 1$ los elementos de la muestra sistemática son homogéneos y $V(\bar{y}_{sy}) >> V(\bar{y}_{ran})$
   - Si $\rho_w \approx 0$ y N es grande el sistemático es tan preciso como el aleatorio
   - Si $\rho_w \approx -1$ los elementos de la muestra sistemática son heterogéneos y $V(\bar{y}_{sy}) < V(\bar{y}_{ran})$

6. Una muestra sistemática tiene la misma precisión que la correspondiente estratificada con una unidad por estrato $\rho_w = 0$.

Tipos de población en el M. sistemático

El éxito del muestreo sistemático frente a otros diseños muéstrales (aleatorio simple y estratificado) depende de la estructura de orden de los elementos de la población. Cochran distingue dos líneas de trabajo: con Poblaciones Teóricas (artificiales) y con poblaciones naturales.

Con poblaciones teóricas se estudia:

a. Poblaciones con orden aleatorio

El muestreo sistemático se utiliza favorablemente en poblaciones en las cuales el orden de los elementos es aleatorio, no habiendo en la población tendencias, estratificación, ni correlación entre los valores vecinos, (por tanto $\rho_w \approx 0$) En este caso se espera que el muestreo sistemático sea tan preciso como el aleatorio simple

$$s_{sy}^2 = \frac{N-n}{N}\frac{1}{n}\frac{\sum (y_{ij} - \bar{y}_{sy})^2}{n-1}$$

Tipos de población en el M. sistemático

b. Poblacion ordenada

Si los elementos dentro de la población están ordenados en magnitud de acuerdo a un esquema lineal u otro. Se dice que la población esta ordenada.

Una muestra sistemática extraída de una población ordenada es generalmente heterogénea con $\rho_w < 0$ y $V(\bar{y}_{sy}) < V(\bar{y}_{ran})$ proporcionando así la muestra estratificada más información que una aleatoria por unidad de costo. Por lo tanto si utilizamos $V(\bar{y}_{ran}) = \frac{S^2}{n}\frac{(N-n)}{N}$ por $V(\bar{y}_{sy})$ (que podemos calcular con una sola muestra) estamos sobreestimando $V(\bar{y}_{sy})$.

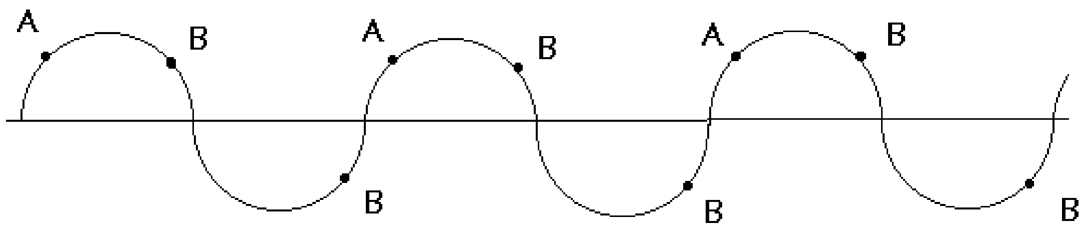
Si el orden es lineal existen métodos para mejorar los estimadores y sus varianzas, utilizando medias donde se ponderan los valores extremos i=1 e i=k y utilizar como estimador de $V(\bar{y}_{sy})$ a

$$S_{sy}^2 = \frac{N-n}{Nn}\frac{\sum(y_i - 2y_{i+k} + y_{i+2k})^2}{n-2}$$ para (1<=i<=n-2)

Tipos de población en el M. sistemático

Poblaciones con variación periódica

Si los elementos de la población siguen una variación cíclica, como una curva sinusoidal, la efectividad del muestreo puede depender del valor de k como lo indica la gráfica:



A. Con k = período, caso menos favorable. Todas las observaciones son iguales, es decir, homogeneidad.

B. Con k = semiperíodo (o múltiplo impar), caso más favorable.

Los elementos de una muestra sistemática de una población periódica puede ser homogéneos (Caso A) esto es $\rho_w > 0$ y cuando N es grande $V(\bar{y}_{sy}) > V(\bar{y}_{ran})$, luego el muestreo sistemático proporciona menos información por unidad de costo que el aleatorio simple.

Tipos de población en el M. sistemático

Poblaciones con variación periódica (continuación)

Los elementos de una muestra sistemática de una población periódica puede ser homogéneos (Caso A) esto es $\rho_w > 0$ y cuando N es grande $V(\bar{y}_{sy}) > V(\bar{y}_{ran})$, luego el muestreo sistemático proporciona menos información por unidad de costo que el aleatorio simple.

Una forma de evitar ese problema es tomando durante la selección de la muestra nuevos orígenes aleatorios, con el fin de mezclar las tendencias cíclicas de la población. Con este procedimiento (aleatorizar la muestra) podemos suponer que $V(\bar{y}_{sy}) \approx V(\bar{y}_{ran})$ y proceder a utilizar $\frac{\hat{S}^2}{n}\frac{(N-n)}{N}$ por $V(\bar{y}_{sy})$

Muestreo Sistemático

Estimación del total

Si conocemos el tamaño de la población (no siempre conocida cuando se inicia el sistemático) podemos estimar el total poblacional Y por $\hat{Y}_{sy} = N\bar{y}_{sy}$ y su varianza $V(\hat{Y}_{sy}) = N^2 V(\bar{y}_{sy})$ la cual debemos estimar bajo las mismas consideraciones de la vista para $V(\bar{y}_{sy})$, esto cuando es igual o próxima a la del muestreo aleatorio.

Estimación de la proporción

El estimador de la proporción al igual que en el aleatorio es un caso particular de $\bar{y}$ así: $p_{sy} = \frac{a_i}{n} = \frac{\sum y_i}{n} = \bar{y}_{sy}$ y su varianza la aproximaremos por la varianza de la proporción como si la muestra fuese aleatoria simple.

Muestreo Sistemático

Tamaño de la muestra:

Para calcular el tamaño de la muestra podemos disponer de dos alternativas:

1- Disponemos estudios anteriores de estimación de $V(\bar{y}_{sy})$ válidas.

2- Usamos las fórmulas del muestreo aleatorio simple en cuyo caso debemos tener presente podríamos sobre-dimensionar el tamaño de muestra si la población es ordenada y sub-dimensionar si la población es periódica.

Muestreo sistemático replicado

Consiste en seleccionar varias muestras sistemáticas simultáneamente originadas por la partición del tamaño de la muestra sistemática inicial y así poder estimar la varianza (Vsy).
El método funciona así: Si originalmente tenemos k=N/n, entonces definimos el número de muestras replicadas en que vamos a dividir la muestra original y la denominamos n_s (por lo general n_s =10) pero tal que n/n_s sea entero), luego calculamos el largo de los nuevos intervalos que es $k_r=n_s k$ y en ese intervalo seleccionamos n_s orígenes aleatorios (para n_s muestras sistemáticas) y cada muestra de tamaño $n_r=n/n_s$ (así el total de elementos muéstrales sigue siendo $n=n_r.n_s$)

Ejemplo (7.1 pag 183 S.M& O)
N=960 n=60 k=16
si n_s=10 entonces $k_r=n_s k=10*16=160$, $n_r=60/10=6$

Muestreo sistemático replicado

Estimación de la media $\overline{Y}$

Sea $\overline{y}_i$ la media de la i-ésima muestra sistemática replicada i=1,2,..,n_s

donde $\overline{y}_i = \dfrac{\sum y_{ij}}{n_r}$

$$\overline{y}_{sy_r} = \frac{\sum \overline{y}_i}{n_s} \qquad \hat{V}\left(\overline{y}_{sy_r}\right) = \frac{N-n}{N}\frac{\sum(\overline{y}_i - \overline{y}_{sy_r})^2}{n_s(ns-1)}$$

Estimación del total Y

$$\hat{Y}_{sy_r} = N\overline{y}_{sy_r}$$

y

$$\hat{V}(\hat{Y}_{sy_r}) = \hat{V}(N\overline{y}_{sy_r}) = N^2\hat{V}(\overline{y}_{sy_r})$$