

00365



**UNIVERSIDAD NACIONAL  
AUTÓNOMA DE MEXICO**

---

**POSGRADO EN CIENCIAS  
MATEMÁTICAS**

**FACULTAD DE CIENCIAS**

**EFICIENCIA RELATIVA DEL MUESTREO  
SISTEMÁTICO BAJO DIVERSOS  
ORDENAMIENTOS DE LAS  
UNIDADES POBLACIONALES**

**TESIS**

**QUE PARA OBTENER EL GRADO ACADÉMICO DE  
MAESTRA EN CIENCIAS MATEMÁTICAS**

**PRESENTA**

**MARIA CRISTINA MÚZQUIZ FRAGOSO**

**DIRECTORA DE TESIS**

**DRA. GUILLERMINA ESLAVA GOMEZ**

**MEXICO, D.F.**

**MAYO, 2004**



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

---

---

# ***Agradecimientos***

## ***A Dios***

Por darme la oportunidad de vivir

## ***A mis padres y hermanos***

Por su gran amor y apoyo para este proyecto

## ***A Antonio***

Por el tiempo y el apoyo totalmente incondicional

## ***A Hugo***

Por la amistad y facilidades para lograr este trabajo

---

---

---

---

---

# ***Agradecimientos***

***A mi directora de tesis, Guillermina Eslava***

Por su tiempo y guía, que para mi han sido muy valiosos

***A mis sinodales,  
Dr. Ignacio Méndez  
Dr. Alberto Contreras  
Mtra. Patricia Romero  
Mtro. Salvador Zamora***

Por sus acertados y oportunos comentarios para mejorar este  
trabajo

***A mi UNAM***

Por todas las oportunidades que me ha ofrecido

---

---

# INDICE GENERAL

---

<b>Resumen .....</b>	<b>3</b>
<b>Capítulo 1. Introducción</b>	
<b>1.1. Objetivo de la tesis.....</b>	<b>6</b>
<b>1.2. Introducción al Muestreo Sistemático.....</b>	<b>9</b>
<b>1.2.1. Concepto.....</b>	<b>9</b>
<b>1.2.2. Manejando el tamaño de muestra.....</b>	<b>13</b>
<b>1.2.3. La Eficiencia del Muestreo Sistemático.....</b>	<b>16</b>
<b>1.2.4. Orden de la muestra.....</b>	<b>19</b>
<b>1.2.5. Relación con el muestreo de Conglomerados.....</b>	<b>20</b>
<b>1.2.6. Comparación de un muestreo sistemático y un aleatorio simple.....</b>	<b>21</b>
<b>1.2.7. Poblaciones con orden "aleatorio".....</b>	<b>22</b>
<b>1.3. El uso del muestreo sistemático en la Estereología.....</b>	<b>23</b>
<b>Capítulo 2. Estimadores</b>	
<b>2.1. Error Cuadrático Medio, ECM.....</b>	<b>31</b>
<b>2.2. Entropía.....</b>	<b>35</b>
<b>2.3. Coeficiente de Homogeneidad, Delta.....</b>	<b>37</b>

**Capítulo 3. Ejemplos**

3.1. Números del 1 al 100.....	39
3.2. Elecciones presidenciales de México en el 2000 a nivel estatal y distrito electoral .....	57
3.3. Matrícula de las escuelas primarias de Guerrero y Distrito Federal (1999-2000).....	103

**Capítulo 4**

4.1. Conclusiones.....	117
------------------------	-----

<b>Anexo A.....</b>	<b>120</b>
---------------------	------------

<b>Bibliografía.....</b>	<b>125</b>
--------------------------	------------

## RESUMEN

---

Es sabido que el orden de las unidades poblacionales es importante para la eficiencia del muestreo sistemático, sin embargo no se especifica qué tipo de orden sería más eficiente. Por otro lado se tiene que existen ciencias como la Estereología (ciencia que estudia objetos a través de un análisis geométrico-estadístico) que con frecuencia utilizan el muestreo sistemático por la precisión en las estimaciones de parámetros tales como el volumen de un núcleo cerebral y la longitud de las prolongaciones neuronales, entre otras. En esta ciencia se tiene el supuesto de que si se ordenan las unidades poblacionales, por una variable auxiliar  $X$  con valores conocidos para toda la población en un orden similar a la normal, al inferir sobre una variable de interés  $Y$  que está altamente correlacionada con  $X$ , se minimiza la varianza del estimador. Es decir el orden tipo curva es el orden óptimo dentro del muestreo sistemático para la Estereología.

Por lo anterior se decidió realizar este trabajo, que tiene como objetivo explorar la hipótesis de que el orden tipo curva es más eficiente que otros tipos de orden en bases de datos de poblaciones humanas.

Para lograr lo anterior se trabajaron tres ejemplos de naturaleza diferente. En cada ejemplo se ordenan las unidades poblacionales en tres tipos de orden: creciente, curva y aleatorio, de acuerdo a los valores de la variable auxiliar  $X$ . Posteriormente se mide el desempeño del diseño evaluando los errores cuadráticos medios de estimadores puntuales al considerar el caso hipotético de conocer los valores muestrales de todas las unidades de la población.

También se evaluó la entropía relativa (medida de certidumbre) y una medida de homogeneidad, delta, con la finalidad de ver cómo se comportan estos estimadores con los diferentes ordenamientos.

---

El primer ejemplo, Números del 1 al 100, supone que los valores de las unidades  $y_i$  son una secuencia de números del 1 al 100, con una  $N = 100$  y una  $n = 10$ , donde  $N$  es el tamaño de la población y  $n$  el tamaño de la muestra. Se ordenan a las unidades poblacionales en diez diferentes maneras, y en cada tipo de orden se evaluó la entropía, el error cuadrático medio y la delta, el estudio fue sobre el estimador de la media, además se trabajó bajo un muestreo aleatorio simple, y se calculó su error cuadrático medio.

Concluyendo que el orden tipo curva fue muy eficiente con una varianza cero, pero no fue el único, ya que otros tres tipos de orden tuvieron la misma varianza. Este fue uno de los puntos más importantes de este ejemplo: *No necesariamente existe un solo orden que minimice la varianza para este caso*. En los cuatro ordenes de varianza cero, la delta alcanzó su valor mínimo y la entropía se maximizó en el orden tipo curva. El caso menos eficiente de los once, fue el que registró el muestreo aleatorio simple.

El segundo ejemplo fue el de las Elecciones Presidenciales de México en el 2000 a nivel estatal y distrito electoral. En éste se manejó un estimador de razón que era el porcentaje de votantes hacia un grupo político y se trabajó con los partidos más representativos en cuanto a número de votos. Se ordenó con la variable de Lista Nominal y se infirió sobre el porcentaje de votantes. Se seleccionaron a tres entidades federativas, Guanajuato, Michoacán, Sinaloa y algunos distritos electorales para este estudio. Es interesante hacer notar que en los estados de Michoacán y Guanajuato, el 73% de los casos (diseños con diferente tamaño de muestra) el orden curva fue más eficiente que el orden creciente y el muestreo aleatorio simple (llamado dentro del trabajo como *aleatorio*). También es importante anotar que la diferencia entre el error cuadrático medio entre el orden creciente y el orden tipo curva es muy pequeño (orden del  $10^{-4}$ ), no así con el muestreo aleatorio simple que en todos los casos fue mayor, llegando alcanzar diferencias del orden  $10^7$ .

Otro resultado interesante en el ejemplo es que al cambiar el tamaño de muestra puede cambiar el tipo de orden más eficiente. Por lo que la eficiencia de un tipo de orden con un tamaño de muestra específico no garantiza que se conserve si se cambia éste.



El último ejemplo es referente a la Matrícula y número de profesores de escuelas primarias de los estados de Guerrero y Distrito Federal (1999-2000). En este caso se ordenó por la variable de número de alumnos y se infirió sobre la variable de número de profesores, ambas variables con una correlación mayor a .7. Se pudo concluir que el orden tipo curva no tuvo una tendencia a ser el orden más eficiente, sin embargo, sigue siendo, junto con el orden tipo creciente, mejor en todos los casos que el muestreo aleatorio simple.

Por tanto, podemos mencionar que no existe un orden óptimo para los ejemplos mostrados, aunque sí se sugiere que se ordenen, ya sea en forma creciente o curva, cuando la variable auxiliar tenga una correlación alta con la variable de interés. También se tiene que cuando se ordena por tipo curva y el histograma tiene una forma suave parecida a la normal, este tipo de orden es más eficiente.

# *Capítulo 1*



## INTRODUCCIÓN

---

Este capítulo se divide en tres secciones, la primera se refiere al objetivo de la tesis 1.1, en este apartado se encontrarán los motivos de realizar este trabajo así como los medios que se utilizan para lograrlo.

En la sección 1.2 se da una introducción a lo que es el muestreo sistemático, así como la notación que se manejará en todos los capítulos. A su vez esta parte se divide en siete puntos, que consideran desde el concepto del muestreo sistemático hasta la comparación con otros tipos de muestreo como el aleatorio simple.

En la última sección 1.3 se desarrolla el tema del uso del muestreo sistemático en la Estereología, este último es muy interesante debido a que algunos supuestos utilizados en esta ciencia fueron el motivo para realizar este trabajo.

---

## **1.1. OBJETIVO DE LA TESIS**

---

El muestreo sistemático es un diseño que se utiliza ampliamente en diversas tareas de recolección de información. Una de las razones de su amplio uso, es porque operativamente no resulta tan complejo, como otros tipos de muestreo y además puede producir estimaciones muy precisas.

La precisión de las estimaciones depende de varios factores, uno de ellos es el orden de las unidades muestrales que decide utilizarse; dentro de este trabajo, se consideran distintos tipos de ordenamientos de las unidades poblacionales, para medir la eficiencia relativa del muestreo sistemático.

El muestreo sistemático es un diseño que ha sido utilizado en diversas encuestas, por ejemplo las realizadas por el Instituto Nacional de Estadística Geografía e Informática, INEGI, donde frecuentemente el muestreo sistemático es usado en la última ó penúltima etapa de muestreo.

La Estereología es una de las áreas que suele utilizar el muestreo sistemático con muy buenos resultados, debido a la precisión obtenida en las estimaciones, por lo que deseamos explorar más al respecto. Se estudian algunas propiedades teóricas, para poder saber si los mismos supuestos utilizados en la Estereología pueden ser aplicados en encuestas de poblaciones humanas.

Desde el punto de vista metodológico, para estimaciones de medias, totales, proporciones y razones, para varios diseños, existen expresiones bien definidas de los estimadores puntuales, de su varianza o de su error

cuadrático medio. Sin embargo, también es conocido que para el diseño del muestreo sistemático no existe expresión alguna para el estimador de la varianza o error cuadrático medio del estimador puntual. En la práctica se obtienen aproximaciones de los estimadores de la varianza o error cuadrático medio usando métodos de remuestreo o simplemente usando las expresiones correspondientes a un muestreo aleatorio simple.

El desempeño de los distintos órdenes de las unidades poblacionales se mide evaluando los errores cuadráticos medios de estimadores puntuales al considerar el caso hipotético de conocer los valores muestrales de todas las unidades de la población.

Uno de los estímulos para realizar este trabajo fue el que los expertos en Estereología en la práctica aplican el supuesto de que si tienen una unidad poblacional  $u_i = \{x, y, a, z, \dots\}$ , donde conocen, para toda la población, el valor de una variable de referencia  $x$ , y se desea inferir sobre una variable de interés  $y$ , basta con que la correlación entre ambas variables sea alta, para concluir que el orden óptimo de las unidades poblacionales  $u$  corresponde al de ordenar por la variable  $x$  en forma de una curva (similar al de la distribución normal).

Para fines de evaluación se usaron tres conjuntos de datos, el primero es con números generados, el segundo corresponde a los resultados de las Elecciones Presidenciales del 2000 a nivel estatal y distrito electoral y el tercero es sobre la matrícula de las escuelas primarias de Guerrero y el Distrito Federal del año 1999-2000.

Además del error cuadrático medio, también se calcula la entropía relativa y la medida de homogeneidad, delta.

## 1.2. INTRODUCCIÓN AL MUESTREO SISTEMÁTICO.

---

### 1.2.1. Concepto.

El muestreo sistemático (*SY*) es un diseño de muestreo probabilístico que tiene como ventaja la simplicidad de su aplicación.

Como primer paso se determina a  $k$  que es igual a dividir el tamaño de la población entre el tamaño de la muestra, después se selecciona de manera aleatoria y con igual probabilidad, uno de los primeros  $k$  elementos de una población. El resto de la muestra es obtenido de manera sistemática, determinando a todos los  $k$  elementos múltiplos.

Todas las  $k$  muestras tienen la misma probabilidad de ser seleccionadas de  $1/k$ . Un procedimiento para seleccionar una muestra sistemática aparece en Särndal (1972, p. 74).

Sea  $n$  es el tamaño de la muestra y  $N$  es el tamaño de la población.

Sea  $k$  un intervalo de muestreo fijo y sea  $n$  el entero de  $(N/k)$ . Entonces

$$N = nk + c$$

Donde  $k$  es un entero, y  $c$  satisface  $0 \leq c < k$ . Si  $c$  es cero,  $k$  es el intervalo de muestreo fijo, si  $c$  no es cero existen varias alternativas, una de ellas es sugerida por Lahiri en 1952 (ver Murthy (1967), p. 139), donde se toma una selección como sigue:

Seleccionar con igual probabilidad  $(1/k)$  un entero aleatorio, que se denominará  $i$ , entre 1 y  $k$ . La muestra seleccionada se compondrá como sigue:

$$s_i = \{u_m : m = i + (j-1)k \leq N; j = 1, 2, \dots, n_s\}$$

Donde  $n_s$  es el tamaño de una muestra, que puede tomar los valores:

$$n_s = \begin{cases} n+1 & \text{cuando } i \leq c \\ n & \text{cuando } c < i \leq k \end{cases}$$

Al entero  $i$  se le llama el principio aleatorio o arranque aleatorio y  $k$  es el salto.

Ejemplo: Sea  $N=23$  y  $k=5$ , de tal manera que se tienen tres muestras con  $n=5$  y otras dos con  $n=4$ , como se esquematiza en el siguiente cuadro,

I	II	III	IV	V
$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$
$x_{11}$	$x_{12}$	$x_{13}$	$x_{14}$	$x_{15}$
$x_{16}$	$x_{17}$	$x_{18}$	$x_{19}$	$x_{20}$
$x_{21}$	$x_{22}$	$x_{23}$		

Lo anterior aunque genera una alteración en la teoría del muestreo sistemático, puede ser insignificante si  $n$  excede 50 y podría ser ignorado, por simplicidad en la presentación de la teoría.



También existe la alternativa de que para una  $N$  muy grande se discriminen las  $c$  unidades que no se encuentren ubicadas en alguna muestra de las  $k$  existentes. Ésta fue la opción que se utilizó en dos ejemplos del capítulo tres.

El conjunto de posibles muestras  $\{s_1, s_2, \dots, s_i, \dots, s_k\}$ , se denota como  $S_{SY}$  que consiste en los  $k$  diferentes conjuntos que pueden ser obtenidos como se mencionó anteriormente.

$$S_{SY} = \{s_1, s_2, \dots, s_i, \dots, s_k\}$$

Por otro lado se tiene que, debido a que cada elemento pertenece sólo a una muestra, y la probabilidad de selección de cada una de las muestras es igual, entonces para cada elemento  $u_m$  que pertenezca a  $U$ ,

$$P(u_m \in S_{SY}) = \pi_m = 1/k \quad u_m \in \{u_1, u_2 \dots u_N\}$$

Para determinar el estimador  $\pi$  y su varianza, se necesitan tener las probabilidades de inclusión de primer y segundo orden. Dado que se tiene que  $\pi_m = 1/k$  para  $u_m \in U$ . Tenemos que para  $\forall u_m \neq u_l \in U$ ,

$$\pi_{ml} = \begin{cases} 1/k & \text{Si } u_m \text{ y } u_l \text{ pertenecen a la misma muestra } s \\ 0 & \text{En otro caso} \end{cases}$$

Es importante resaltar que no existen intersecciones entre dos muestras cualesquiera. Esto es,  $s_1, s_2, \dots, s_i, \dots, s_k$  representan una partición de  $U$ . Por tanto la unión de las  $k$  muestras, es la población total  $U$ , es decir,

$$U = \bigcup_{i=1}^k s_i$$

podemos escribir al total poblacional de nuestra variable de interés  $y$ , como

$$t = \sum_{m=1}^N y_m \text{ donde,}$$

$$t = \sum_{i=1}^k t_{s_i} \text{ con } t_{s_i} = \sum_{s_i} y_m$$

Ejemplo, cuando  $c = 0$ ,

	Muestra s				
	$s_1$	...	$s_i$	...	$s_k$
Valores de $y$	$y_1$	...	$y_i$	...	$y_k$
	$y_{k+1}$	...	$y_{k+i}$	...	$y_{2k}$
	...	...	...	...	...
	$y_{(n-1)k+1}$	...	$y_{(n-1)k+i}$	...	$y_N$
Totales	$t_{s_1}$	...	$t_{s_i}$	...	$t_{s_k}$

El estimador del total de la población  $t = \sum_U y_m$ , toma la forma de

$$\hat{t}_\pi = k \sum_{s_i} y_k = kt_{s_i}.$$

Y la varianza está dada por:

$$Var_{SY}(\hat{t}_\pi) = k^2 \sum_{i=1}^k \frac{(t_{s_i} - \bar{t})^2}{k} = k \sum_{i=1}^k (t_{s_i} - \bar{t})^2$$

Donde,  $\bar{t} = \sum_{i=1}^k t_{s_i} / k$  es el promedio de los totales muestrales.

La varianza también puede escribirse como:

$$Var_{SY}(\hat{t}_\pi) = k(k-1)S_t^2$$

Donde,

$$S_t^2 = \frac{1}{k-1} \sum_{i=1}^k (t_{s_i} - \bar{t})^2$$

es la varianza de los totales muestrales. Como se puede observar la varianza irá decreciendo en la medida que los totales muestrales sean similares.

### 1.2.2. Manejando el tamaño de muestra.

Por definición del muestreo sistemático, se tiene que si  $c = 0$ , entonces  $N = nk$ ,  $k \in \mathbb{N}$ , y cada muestra  $s_k$  tiene tamaño  $n$ . Mientras que si

$c > 0$  el tamaño de una muestra, puede ser de tamaño  $n+1$ , cuando  $i \leq c$ , o  $n$  cuando  $c < i \leq k$ . Lo anterior es cierto si  $k$  es entero, pero existen casos donde no es así.

En Särndal et al (1972, p.77) se presentan los siguientes dos métodos para manejar el tamaño de muestra en un muestreo sistemático.

### **1.2.2.1. Método de intervalo fraccional.**

En algunos casos,  $k$  puede ser una fracción, se puede utilizar el método de Intervalo fraccional, que consiste en obtener un número aleatorio  $\xi$  de una distribución uniforme en el intervalo  $(0, k)$ . La muestra seleccionada consistirá en elementos  $m$ , que se toman como:

$$m-1 < \xi + (j-1)k \leq m; \quad j = 1, \dots, n$$

Este método es equivalente a obtener con probabilidad  $1/N$ , un entero aleatorio  $i$  entre 1 y  $N$ , y seleccionar los primeros  $m$  elementos como,

$$(m-1)n < i + (j-1)N \leq mn; \quad j = 1, \dots, n$$

Por ejemplo, si  $m = 1$ ,  $i$  es seleccionado si satisface  $0 < i \leq n$ , y esto ocurre con probabilidad  $n/N$ .

### 1.2.2.2. Método circular.

En el muestreo sistemático cuando el tamaño de la población  $N$  no es múltiplo del tamaño de muestra  $n$ , el estimador de la media es sesgado.

Si  $N/n$  no es entero, una manera de tratar este problema consiste en considerar la lista como si fuese circular (el elemento  $N + 1$  coincide con el primero) y:

- Sea  $k$  el entero más cercano a  $N/n$ ;
- Se selecciona un número al azar  $i$ , entre 1 y  $N$ ;
- Se toma como muestra los elementos de la lista como sigue:

$$m = i + (j-1)k \quad \text{Si} \quad i + (j-1)k \leq N \quad j = 1, \dots, n$$

o

$$m = i + (j-1)k - N \quad \text{Si} \quad i + (j-1)k > N \quad j = 1, \dots, n$$

Todos los elementos de la lista tienen la misma probabilidad de selección

$$\pi_m = n/N \quad m = 1, 2, \dots, N.$$

### 1.2.3. La Eficiencia del Muestreo Sistemático.

La varianza  $V_{SY}(\hat{t}_\pi) = k \sum_{i=1}^k (t_{s_i} - \bar{t})^2$ , es cercana a cero si los totales de las muestras posibles son aproximadamente iguales. Es por esto que es importante el orden de las unidades muestrales en la población.

Más adelante se mostrarán resultados interesantes que se obtuvieron al ordenar las unidades poblacionales en distintos órdenes.

Consideremos el caso de que  $N = nk$  cuando  $k$  es entero y  $\bar{y}_U = \sum_U y_m / N$ , es la media de la población, se tiene lo siguiente:

$$\hat{t}_\pi = \sum_{s_i} y_m / \pi_m = k \sum_{s_i} y_m = \sum_{s_i} y_m / (n/N) = N \sum_{s_i} y_m / n = N \bar{y}_{s_i}$$

y la varianza es

$$Var(\hat{t}_\pi) = N^2 \frac{1}{k} \sum_{i=1}^k (y_{s_i} - \bar{y}_U)^2$$

Además la suma de cuadrados corregidos por la media es:

$$\sum_U (y_m - \bar{y}_U)^2 = \sum_{i=1}^k \sum_{S_i} (y_m - \bar{y}_{S_i})^2 + \sum_{i=1}^k n (\bar{y}_{S_i} - \bar{y}_U)^2$$

Särndal (1972, p. 78), lo que representa que la variación total se puede descomponer en la variación dentro de cada muestra, y la que existe entre las muestras, es decir:

$$SST = SSW + SSB$$

Donde,

$$SST = \sum_U (y_m - \bar{y}_U)^2, \quad SSW = \sum_{i=1}^k \sum_{S_i} (y_m - \bar{y}_{S_i})^2, \quad SSB = \sum_{i=1}^k n(\bar{y}_{S_i} - \bar{y}_U)^2$$

Donde  $SS$  es la suma de cuadrados,  $T$  es de total,  $W$  de within (dentro) y  $B$  de between (entre).

Para medir el grado de homogeneidad entre los elementos al interior de las muestras o conglomerados, Särndal et al (1972, p.78) propone utilizar la medida siguiente,

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

que es llamado el coeficiente de correlación intraclase. Puede ser interpretado como una medida de correlación entre pares de elementos dentro de la misma muestra sistemática. Se obtiene un valor positivo de  $\rho$  cuando elementos en la misma muestra tienden a tener valor similares de  $y$ . Por el otro lado  $\rho = 1$  si  $SSW = 0$ , es decir existe una completa homogeneidad (no variación) dentro de las muestras sistemáticas. En el extremo contrario  $\rho = -1/(n-1)$  si  $SSB = 0$  cuando existe una completa

heterogeneidad dentro de las muestras. Otra medida de homogeneidad

similar a  $\rho$  es  $\delta = 1 - \left( \frac{N-1}{N-k} \right) \frac{SSW}{SST}$ .

Los valores extremos de  $\delta$  son:

$$\delta = \begin{cases} -\frac{k-1}{N-k} & \text{si } SSB = 0 \\ 1 & \text{si } SSW = 0 \end{cases}$$

En el capítulo dos, se hablará un poco más sobre  $\delta$ .



#### **1.2.4. Orden de la muestra.**

La teoría dice que cuando el criterio de ordenación de los elementos en la lista es tal que los elementos más parecidos en cuanto a la variable de interés tienden a estar más cercanos, el muestreo sistemático suele ser más preciso que el aleatorio simple, ya que recorre la población de un modo más uniforme.

El punto anterior es muy importante y el tema central de este trabajo es considerar el orden de los datos para mejorar la eficiencia del muestreo sistemático.

El problema en la práctica consiste en identificar cuál es la variable sobre la cual medir lo parecido de los datos, cuando se manejan varias variables de interés.

Sobre este tema en específico vendrán posteriores apartados ilustrados con ejemplos.

**1.2.5. Relación del muestreo sistemático con el muestreo por Conglomerados.**

Con  $N = nk$ , las  $k$  posibles muestras se muestran en la siguiente tabla:

		Muestras posibles $s$				
		1	$i$	$k$		
		$s_1$	...	$s_i$	...	$s_k$
Valores de $y$		$y_1$	...	$y_i$	...	$y_k$
		$y_{k+1}$	...	$y_{k+i}$	...	$y_{2k}$
		...	...	...	...	...
		$y_{(n-1)k+1}$	...	$y_{(n-1)k+i}$	...	$y_N$
Media		$\bar{y}_{s_1}$	...	$\bar{y}_{s_i}$	...	$\bar{y}_{s_k}$

La población ha sido dividida en  $k$  muestras, que pueden considerarse como conglomerados cada uno con  $n$  elementos. Una muestra sistemática corresponde a elegir una de las  $k$  muestras de manera aleatoria. El muestreo sistemático equivale a seleccionar un solo conglomerado y constituye una muestra completa.

Un muestreo sistemático es un muestreo aleatorio simple de un solo conglomerado de una población dividida en  $k$  conglomerados.

### **1.2.6. Comparación de un muestreo sistemático y un aleatorio simple.**

El desempeño de un muestro sistemático en relación con un aleatorio simple es sumamente dependiente de las propiedades de la variable de interés de la población. Hay poblaciones para las cuales el muestreo sistemático tiene un alto grado de precisión y otros donde es menos preciso que un aleatorio simple. Esto dificulta el que se tengan recomendaciones sobre qué muestreo utilizar en cada caso. Sin embargo siempre se recomienda tener conocimiento de la estructura de la población para una decisión efectiva.

Se exploran dos líneas de investigación en este trabajo, una es comparar los diferentes tipos de muestreo en poblaciones artificiales donde  $y_i$  es una función simple de  $i$ , el orden de las unidades muestrales. El otro es hacer una comparación para poblaciones reales.

En el artículo Madow, W.G. y Madow, L.H.(1944, p. 2) se deducen los siguientes resultados:

- Si las correlaciones seriales tienen una suma positiva, el muestreo sistemático es menos eficiente que el aleatorio simple.
- Si las correlaciones seriales tienen una suma aproximadamente igual a cero, el muestreo sistemático es aproximadamente equivalente a un aleatorio simple.
- Si las correlaciones seriales tienen una suma negativa, el muestreo sistemático es mejor que el aleatorio simple.

Donde la correlación serial está definida como:  $\sigma^2 = \sum_{S_i} (\bar{x}_{S_i} - \bar{x})/kn$ .

### **1.2.7. Poblaciones con orden "aleatorio"**

En algunas ocasiones el muestreo sistemático es usado por conveniencia y por que es operativamente más fácil que algunos otros tipos de muestreo. El orden de las unidades muestrales que manejan puede considerarse como aleatorio porque no depende de ninguna variable relacionada con lo que se estimará.

En estos casos no habrá una tendencia en las variables, y se podrá esperar que un muestreo sistemático sea equivalente en términos de varianza a un muestreo aleatorio  $V_{MAS}$ .

#### **Teorema**

Considere todas las  $N!$  poblaciones finitas que son formadas por las  $N!$  permutaciones de un conjunto de números  $y_1, y_2, \dots, y_N$ . Entonces el promedio de todas estas poblaciones finitas,

$$E(V_{SY}) = V_{MAS}$$

Es importante notar que  $V_{MAS}$  es la misma para todas las permutaciones.

Este resultado fue demostrado por Madow, W.G. y Madow, L.H.(1944), mostrando que si el orden de las unidades en una población finita puede considerarse aleatorio proveniente de  $N!$  permutaciones, el muestreo sistemático es equivalente en promedio a un muestreo aleatorio simple.

## **1.3. EL USO DEL MUESTREO SISTEMÁTICO EN LA ESTEREOLOGÍA**

---

La estereología se define como un conjunto de métodos útiles para obtener información cuantitativa de carácter geométrico-estadístico de un objeto de interés, a partir de secciones del objeto.

En los últimos años ha habido un fuerte avance de los métodos probabilísticos de cuantificación estereológica. Con ellos, fácilmente se pueden estimar parámetros tales como el volumen de un núcleo cerebral, el volumen de las células nerviosas, la longitud de las prolongaciones neuronales, entre otras. Gracias a la aplicación de estos métodos, en el campo de los estudios biológicos y de las neurociencias, se han podido estudiar parámetros morfológicos con una gran precisión y fiabilidad.

En general, los métodos estereológicos gozan de una serie de ventajas metodológicas y prácticas de indudable interés para el estudio del sistema nervioso. Como ventajas metodológicas se podrían destacar, por un lado, que son métodos insesgados, carentes de error sistemático, y por otro, que la estimación de los parámetros no está influida ni por la forma del objeto, ni por su distribución y/o tamaño. Además, una clara ventaja metodológica y práctica es que los métodos estereológicos son altamente eficientes. En otras palabras, no exigen un gran esfuerzo para la cuantificación ya que con una muestra muy pequeña de objetos y pocos cortes también llamadas secciones (alrededor de 6-7) se obtienen coeficientes de error y de variación aceptablemente bajos.

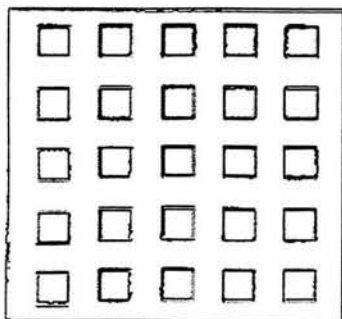
En algunos artículos del *MaPhySto*, *Centre for Mathematical Physics and Stochastics*, de la Universidad de Aarhus, se explica que algunos de los parámetros estereológicos de interés en los estudios psicobiológicos es el número de poblaciones neuronales y volúmenes de estructuras cerebrales. Mediante estos estudios, se ha observado una disminución de neuronas pigmentadas en la sustancia negra en pacientes con Parkinson y en el núcleo basal de Meynert en el caso de pacientes con Alzheimer. En pacientes esquizofrénicos también se ha constatado una reducción de entre un 40% y 50 % en el número neuronal en los núcleos talámicos frente a sujetos normales. Estos cambios estructurales pueden relacionarse con las disfunciones cerebrales y neurocognitivas observadas en estas patologías. De este modo, la cuantificación neuronal y del volumen del área, nos informaría de la existencia de cambios morfofuncionales cerebrales que pudieran relacionarse con deterioros de funciones cognitivas.

Metodológicamente, la estimación del número de células nerviosas en un determinado núcleo, puede realizarse empleando numerosos métodos estereológicos. De ellos, el fraccionador óptico es uno de los más recomendables. Este método se basa en la combinación de un muestreo sistemático de las secciones y el disector óptico.

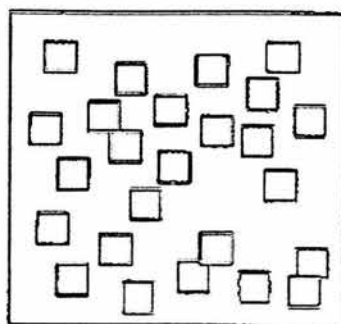
Previamente a la aplicación del fraccionador óptico se ha de proceder a una seriación sistemática de la estructura a cuantificar. Esta seriación permite conocer la distancia exacta entre las secciones que después será recogida en el cálculo de los parámetros estereológicos. De entre todas las secciones obtenidas se seleccionarán algunas, sobre las que se realizarán las cuantificaciones. Un requisito importante para realizar las cuantificaciones estereológicas del número de células nerviosas es el muestreo sistemático

de la zona elegida y para ello la orientación de las secciones debe ser al azar.

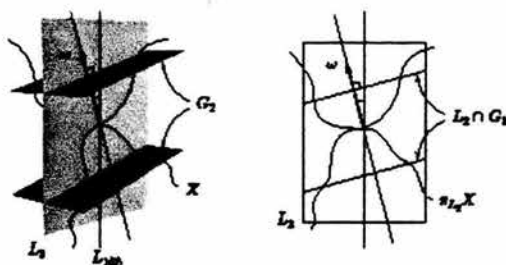
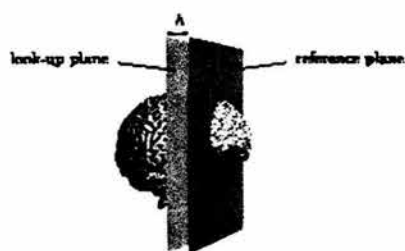
En el muestreo sistemático, la cuantificación no se realiza en todo el tejido sino que se contarán el número de células que se encuentren contenidas en disectores ópticos o cuadrantes (áreas específicas de tejido), que se hallan distribuidos sistemáticamente en una plantilla previamente diseñada, como se muestra en la figura 1.3.1. Esta plantilla, inicialmente, se coloca fuera de la zona cerebral de la sección y seguidamente, ésta se desplaza en una dirección determinada a priori hasta que se superpone sobre el núcleo de interés. Posteriormente, se toman en cuenta sólo las neuronas, células gliales, etc., que se encuentran dentro de los disectores que forman parte de la plantilla; aquellas neuronas/ glías que no se encuentran en los disectores de la plantilla no son cuantificadas. Así, se evita el sesgo de que el experimentador seleccione las áreas del tejido en las que aparezcan la mayor densidad de partículas o aquellas que mejor se visualicen, ya que el conteo depende del lugar en el que se encuentren los disectores. Además el muestreo sistemático no provoca la sobreestimación numérica, pues se evita que un mismo campo sea cuantificado en repetidas ocasiones, como podría suceder con el muestreo al azar, como se observa en la figura 1.3.2.



*Figura 1.3.1. Muestreo sistemático*



*Figura 1.3.2. Muestreo al azar*



*Figura 1.3.3. Un disector consiste en la referencia de un plano a una distancia  $h$ .*

*Figura 1.3.4. Ilustración de un muestreo para la estimación de medidas relevantes en  $R^3$ , basado en proyecciones de un plano vertical.*

La estereología también se usa en la estadística espacial y en el análisis digital de imágenes. De hecho, algunos científicos, como Adrian Baddeley profesor de estadística en la Universidad de Western Australia en la escuela de Matemáticas y Estadística, definen a la estereología como la versión espacial de la teoría del muestreo. La estereología es desarrolló inicialmente en la biología y ciencias de los materiales y como se mencionó, es un camino rápido para analizar materiales sólidos tridimensionales a partir de información que se puede tener en  $R^2$ , por medio de operaciones de muestreo y geométricas. La Estereología se aplica en ciencias como:

- Anatomía
- Biología y patología
- Ciencia de los materiales



- Mineralogía y metalurgia
- Botánica
- Ecología y Silvicultura
- Geología
- Computación gráfica

Es importante subrayar que dentro de la estereología hay varias técnicas que pueden ser utilizadas para lograr un muestreo apropiado con determinados propósitos cuando se usa un microscopio confocal.

Como se mencionó en el objetivo de la tesis, uno de los estímulos para realizar este trabajo, fue ejemplificar el supuesto que se aplica en la estereología que indica que el orden óptimo de las unidades poblacionales  $\mu_i = \{x, y, a, z, \dots\}$   $i = 1, 2, \dots, N$  corresponde al de ordenarlas por la variable de referencia  $x$  en forma de curva, dado que se conoce su valor para toda la población y está altamente correlacionada con la variable de interés  $y$ , sobre la cual se va a inferir.

Al respecto el autor Deming (1955) muestra en su libro *Some theory of sampling*, la siguiente tabla, donde se consideran seis distintas distribuciones y su respectiva varianza y desviación estándar.

Los resultados son muy interesantes y útiles para el muestreo.

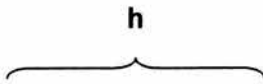


Figura 1.3.5. Distribución rectangular

**Varianza**

**Desviación estándar**

$$h^2 / 12$$

$$* 0.29h$$

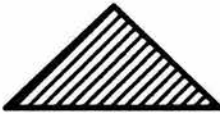


Figura 1.3.6. Distribución de triángulo equilátero.

$$h^2 / 24$$

$$* 0.20h$$

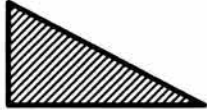


Figura 1.3.7. Distribución triángulo cuadrado.

$$h^2 / 18$$

$$* 0.24h$$



Figura 1.3.8. Distribución dos triángulos.

$$h^2 / 8$$

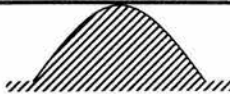
$$* 0.35h$$



Figura 1.3.9. Distribución semicírculo.

$$h^2 / 16$$

$$* 0.25h$$



$$h = 6\sigma$$

Figura 1.3.10. Distribución normal

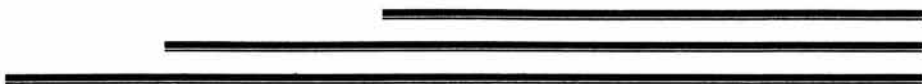
$$h^2 / 36$$

$$* 0.17h$$

\*Aproximaciones

Como se puede observar la distribución (dentro de estas seis) que tiene la menor varianza es precisamente la distribución normal.

# *Capítulo 2*



## ESTIMADORES

---

Para evaluar el desempeño del estimador bajo los distintos ordenamientos de las unidades poblacionales bajo el muestreo sistemático y el muestreo aleatorio simple, se utilizaron el error cuadrático medio, la entropía y la medida de homogeneidad, delta, en los tres ejemplos presentados en el siguiente capítulo.

Como se verá en el tercer capítulo, puede existir más de un ordenamiento que minimice el error cuadrático medio; en este tipo de situaciones, se hace necesario tener otros parámetros de referencia que apoyen en la decisión de elegir el tipo de orden más eficiente. Es por esto que se decidió incluir el cálculo de la entropía y la medida de homogeneidad, delta.

En la sección 2.1 se presenta la información referente al error cuadrático medio, en la 2.2 se muestra la entropía y por último en la 2.3 se explica el concepto y cálculo de la medida de homogeneidad, delta.

Es importante subrayar que la entropía no fue calculada en cada muestra sistemática debido a que no dependen explícitamente del valor de  $k$ . Por lo tanto, el estimador más utilizado fue el error cuadrático medio.

A continuación se da una breve descripción del concepto y el cálculo de cada uno de los estimadores.

## 2.1 ERROR CUADRÁTICO MEDIO, ECM

Para comparar un estimador insesgado con uno sesgado, o dos estimadores con un sesgo diferente, un criterio muy útil, es el error cuadrático medio (ECM) del estimador. Formalmente el ECM es como sigue.

Sea  $\hat{\mu}$  cualquier estimador de un parámetro desconocido  $\mu$ . Se define el error cuadrático medio de  $\hat{\mu}$  como el valor esperado del cuadrado de la diferencia entre  $\hat{\mu}$  y  $\mu$ . Es decir:

$$\begin{aligned}
 ECM(\hat{\mu}) &= E(\hat{\mu} - \mu)^2 = E(\hat{\mu}^2 - 2\hat{\mu}\mu + \mu^2) \\
 &= E(\hat{\mu}^2) - 2\mu E(\hat{\mu}) + E(\mu^2) \\
 &= (Var(\hat{\mu}^2) + E^2(\hat{\mu})) - 2\mu E(\hat{\mu}) + \mu^2 \\
 &= Var(\hat{\mu}^2) + (\mu - E(\hat{\mu}))^2 \\
 &= (Varianza de \hat{\mu}) + (Sesgo)^2
 \end{aligned}$$

Es importante subrayar que habrá que cuidar la interpretación del ECM como criterio de precisión. Por ejemplo, si se tienen dos tipos de estimadores diferentes del mismo parámetro, que tienen el mismo ECM, no se podrá concluir que son equivalentes, porque la distribución de la frecuencia de errores  $(\hat{\mu} - \mu)$  de diferente tamaño no será la misma, aunque tengan un mismo sesgo.

Las expresiones del error cuadrático medio para un estimador de razón y su estimador están dados por:

$$\begin{aligned}ECM(\hat{R}) &= \frac{1}{X^2} \{V(\hat{Y}) + R^2V(\hat{X}) - 2RCov(\hat{X}, \hat{Y})\} \\ \widehat{ECM}(\hat{R}) &= \frac{1}{X^2} \{V(\hat{Y}) + \hat{R}^2V(\hat{X}) - 2\hat{R}\widehat{Cov}(\hat{X}, \hat{Y})\} \\ &= \frac{1}{\hat{X}^2} \{V(\hat{Y} - \hat{R}\hat{X})\}\end{aligned}$$

Donde:

- $X$ : El denominador de la proporción que se desea tener, por ejemplo, número total de personas adultas.
- $Y$ : El numerador de la proporción que se desea tener, por ejemplo, número total de personas adultas que fuman.
- $R = Y/X$ : La proporción que será estimada, por ejemplo, proporción de personas adultas que fuman.
- $\hat{X}$ : Estimado de  $X$
- $\hat{Y}$ : Estimado de  $Y$
- $\hat{R}$ : El cociente de los totales estimados, estimador de  $R$

Para el muestreo sistemático:

$$\begin{aligned}
 ECM(\hat{R}_{SY}) &= \frac{k^2}{X^2} \left(1 - \frac{1}{k}\right) \left(\frac{1}{k-1}\right) \sum_{i=1}^k (y_i - Rx_i)^2 \\
 &= \frac{k}{X^2} \sum_{i=1}^k (y_i - Rx_i)^2
 \end{aligned}$$

$$\widehat{ECM}(\hat{R}_{SY}) = \frac{k}{X^2} \sum_{i=1}^k (y_i - \hat{R}x_i)^2$$

Donde:

$k = N/n$ , con  $N/n \in \mathbb{N}$

$y_i$ : Valor de  $Y$  en el grupo  $i$ ,  $i = 1, 2, \dots, k$

$x_i$ : Valor de  $X$  en el grupo  $i$ ,  $i = 1, 2, \dots, k$

$R_{SY}$ : Proporción de  $Y/X$ , en un muestreo sistemático.

$\hat{R}_{SY}$ : El cociente de los totales estimados, en un muestreo sistemático.

Para el muestreo aleatorio simple, Cochran (1977), expresión 2.39, p.31:



$$\begin{aligned}
 ECM(\hat{R}_{MAS}) = V(\hat{R}) &= \frac{1-f}{nX^2} \frac{\sum_{i=1}^N (y_i - Rx_i)^2}{N-1} \\
 &= \frac{\left(1 - \frac{n}{N}\right)}{n \left(\sum_{i=1}^N \frac{X_i}{N}\right)^2} \frac{\sum_{i=1}^N (y_i - Rx_i)^2}{N-1} \\
 &= \frac{N^2}{X^2} \left(1 - \frac{n}{N}\right) \left(\frac{1}{n}\right) \frac{\sum_{i=1}^N (y_i - Rx_i)^2}{N-1}
 \end{aligned}$$

Mientras que el ECM estimado es,

$$\widehat{ECM}(\hat{R}_{MAS}) = \hat{V}(\hat{R}_{MAS}) = \frac{N^2}{X^2} \left(1 - \frac{n}{N}\right) \left(\frac{1}{n}\right) \frac{\sum_{i=1}^n (y_i - \hat{R}x_i)^2}{n-1}$$

Donde

- $N$ : Número total de elementos de la población.
- $n$ : Tamaño de la muestra
- $R_{MAS}$ : Proporción de  $Y/X$ , en un muestreo aleatorio simple.
- $\hat{R}_{MAS}$ : El cociente de los totales estimados, en un muestreo aleatorio simple.

## 2.2. ENTROPÍA

El término de entropía es utilizado en varios sentidos, en la física es la magnitud termodinámica que mide la parte no utilizable de la energía no contenida en un sistema; en el análisis de la información sirve para dar una medida de certidumbre eliminada por la recepción de un mensaje, se utiliza mucho en mecánica estadística, llamándola entropía termodinámica.

La entropía de una variable aleatoria continua que tiene una función de

distribución  $F(X)$  es  $H(X) = - \int_{-\infty}^{\infty} f(x) \ln f(x) dx$ , donde

$F'(X) = f(x)$ , si es que existe, Rényi (1970, p.367).

Para el caso de una variable aleatoria discreta  $X$  con distribución  $P_x$ , la

entropía está dada por  $H(X) = - \sum_x P_x(x) \ln P_x(x)$ .

La entropía relativa o discriminación entre dos funciones de probabilidad  $q(x)$  y  $f(x)$  está definida como:

$$\begin{aligned} D(q | f) &= \sum_x q(x) \log \frac{q(x)}{f(x)} = \sum_x q(x) \log q(x) - \sum_x q(x) \log f(x) \\ &= \sum_x q(x) [\log q(x) - \log f(x)] \end{aligned}$$

El uso de entropía que se manejará en adelante corresponde a la entropía relativa, donde  $q(x_i)$  corresponde al orden y  $f(x_i)$  tomará los valores de la variable auxiliar de cada unidad poblacional.

La entropía en los ejemplos de este trabajo, resulta ser muy útil, para la selección de muestras.

La entropía como se puede observar no depende del valor de  $k$  (número posible de muestras generadas en el muestreo sistemático). Se calcula considerando a toda la población.

Un resultado interesante en los cálculos de la entropía relativa para uno de los ejemplos mostrados en este trabajo, fue que la entropía para los datos ordenados en curva es mayor o igual que aquella que sea para datos en orden creciente. Para ver la demostración, vea Anexo A.

## 2.3. MEDIDA DE HOMOGENEIDAD, DELTA

Este estimador es una medida de homogeneidad entre los elementos, y es similar a otro estimador llamado  $\rho$ .

Para los casos donde se estima la media de la población  $\bar{y}_U = \sum_U y_m / N$ , la suma de cuadrados se puede expresar como:

$$\sum_U (y_m - \bar{y}_U)^2 = \sum_{i=1}^k \sum_{S_i} (y_m - y_{S_i})^2 + \sum_{i=1}^k n (y_{S_i} - \bar{y}_U)^2$$

Lo que representa que la variación total se puede descomponer en la variación dentro de cada muestra, y la que existe entre ellas, es decir:

$$SST = SSW + SSB$$

Donde  $SS$  es la suma de cuadrados,  $T$  es total,  $W$  de within (dentro de las muestras) y  $B$  de between (entre las muestras).

Aquí es donde se utiliza  $\rho$  para medir el grado de homogeneidad entre los elementos,

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

Que es llamado el coeficiente de correlación intraclase. Puede ser interpretado como una medida de correlación entre pares de elementos

$$\delta = 1 - \frac{N-1}{N-k} \frac{SSW}{SST}$$

Los valores extremos de  $\delta$  son:

$$\delta_{\min} = -\frac{k-1}{N-k} \text{ y } \delta_{\max} = 1$$

Si  $SSW = 0$  entonces  $\delta = 1$  (Su valor máximo), es decir los datos tienen una homogeneidad total dentro de las muestras.

El mínimo valor de  $\delta$  se alcanza cuando  $SSW = SST$ , es decir, cuando existe una total heterogeneidad dentro de las muestras.

Algunas veces se prefiere calcular  $\delta$ , debido a que  $\rho$  puede estar representado en términos de los coeficientes de correlación de muestras de igual tamaño, lo que no siempre puede darse.

# *Capítulo 3*



## **INTRODUCCIÓN AL CAPÍTULO 3: EJEMPLOS**

---

Este capítulo sin lugar a dudas es clave, porque aquí es donde se muestran los tres ejemplos numéricos a los cuales se les evalúa los distintos estimadores para medir la eficiencia de cada tipo de ordenamiento. En este apartado se ejemplifica el supuesto de que el ordenamiento en forma de normal, tipo curva, de las unidades poblacionales  $u_i$  con  $i = 1, 2, \dots, N$ , es más eficiente que otros tipos de ordenamiento como el creciente y aleatorio, cuando existe una alta relación entre la variable de interés y la variable auxiliar.

En la sección 3.1 se muestra un ejemplo con números del 1 al 100, que ayuda a visualizar cómo es la metodología a usar para medir la eficiencia. En el ejemplo se consideran 11 tipos de ordenamientos.

En la sección 3.2 se trabaja con el ejemplo más complejo, las Elecciones Presidenciales de México del año 2000 a nivel estatal y distrito electoral. Los resultados son muy interesantes y reflejan lo complicado que resulta llegar a conclusiones generales.

En la última sección 3.3 se estudia la Matrícula de las escuelas primarias de Guerrero y Distrito Federal (1999-2000). En este caso se robustecen las conclusiones de los dos ejemplos anteriores.

---

## 3.1. NÚMEROS DEL 1 AL 100

---

Primer ejemplo numérico, el motivo para trabajar con estos datos, fue la mención que se hace en el libro "Model Assisted Survey Sampling" de Särndal(1972, p.80), en su capítulo de muestreo sistemático, precisamente para ejemplificar el efecto del orden de la población en los resultados obtenidos. En este ejemplo la unidad poblacional  $u_i = \{x_i\}$   $i = 1, 2, \dots, N$  sólo considera una variable, por lo tanto la variable auxiliar es la misma que la variable de interés  $x = y$ , aunque en la realidad esto no se da.

La metodología a seguir es:

1. Calcular la varianza bajo un muestreo aleatorio simple.
2. Ordenar de distintas maneras a las unidades poblacionales por la variable de referencia  $x$ .
3. Dividir en  $k$  muestras a la población.
4. Obtener las muestras resultantes.
5. Sumar los elementos de cada muestra.
6. Calcular la varianza de la suma de cada una de las muestras, bajo un muestreo sistemático.
7. Calcular la entropía relativa por cada tipo de orden.
8. Calcular la medida de homogeneidad delta por cada tipo de orden.



El ejemplo consiste en una población de  $N = 100$ , donde la variable  $x_i = i$ ,  $i = 1, 2, \dots, 100$  y  $n = 10$ . Por lo tanto el número posible de muestras es de  $k = 100/10 = 10$ .

Se tiene,

$$\begin{aligned}
 S_{xU}^2 &= \frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N-1} = \frac{\sum_{i=1}^N (x_i^2 - 2x_i\bar{X} + \bar{X}^2)}{N-1} \\
 &= \frac{\sum_{i=1}^N x_i^2 - 2\bar{X}\sum_{i=1}^N x_i + N\bar{X}^2}{N-1} = \frac{\sum_{i=1}^N x_i^2 - \frac{\left(\sum_{i=1}^N x_i\right)^2}{N}}{N-1} \\
 &= \frac{83,325}{99} = 841.66
 \end{aligned}$$

Sin hacer caso al orden de la población, la varianza bajo el muestreo aleatorio simple *MAS* sin reemplazo es,

$$\begin{aligned}
 Var_{MAS}(\hat{t}_\pi) &= N^2 \frac{1-f}{n} S_{xU}^2 = 100^2 \frac{1-(10/100)}{10} 841.66 \\
 &= 7.575 \cdot 10^5
 \end{aligned}$$

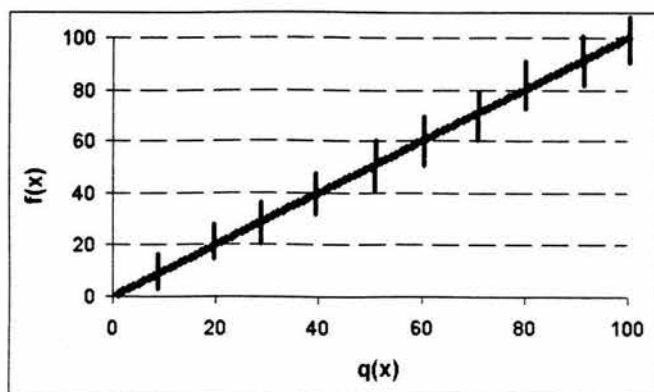
Examinemos otros tipos de orden. En este ejemplo  $f(x_i)$  son los valores  $x_i$  que se le asignarán a un orden dado  $q(x_i)$ .

Orden	$q(x_i)^*$	Orden	$f(x_i)^*$	Orden	$q(x_i)^*$	Orden	$f(x_i)^*$	Orden	$q(x_i)^*$	Orden	$f(x_i)^*$	Orden	$q(x_i)^*$	Orden	$f(x_i)^*$
1	50	11	40	21	30	31	20	41	10	51	51	91	91		
2	49	12	39	22	29	32	19	42	9	52	52	92	92		
3	48	13	38	23	28	33	18	43	8	53	53	93	93		
4	47	14	37	24	27	34	17	44	7	54	54	94	94		
5	46	15	36	25	26	35	16	45	6	...	...	95	95		
6	45	16	35	26	25	36	15	46	5	...	...	96	96		
7	44	17	34	27	24	37	14	47	4	87	87	97	97		
7	43	18	33	28	23	38	13	48	3	88	88	98	98		
9	42	19	32	29	22	39	12	49	2	89	89	99	99		
10	41	20	31	30	21	40	11	50	1	90	90	100	100		

\* Valor  $X_i$  que se le asignará al orden correspondiente.

De esta manera al orden 1 se le proporciona el valor 50.

A continuación se muestra un ejemplo donde se ordenan todas las unidades de manera creciente, donde  $q(x_i) = f(x_i) = i$ , como se puede observar en la gráfica 3.1.



Gráfica 3.1. Orden creciente de  $X_i$

En la tabla 3.1, se presenta el conjunto de posibles muestras bajo un orden tipo creciente.

### Caso 0

Conjunto de posibles muestras

$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$	$s_7$	$s_8$	$s_9$	$s_{10}$	
1	2	3	4	5	6	7	8	9	10	
11	12	13	14	15	16	17	18	19	20	
21	22	23	24	25	26	27	28	29	30	
31	32	33	34	35	36	37	38	39	40	
41	42	43	44	45	46	47	48	49	50	
51	52	53	54	55	56	57	58	59	60	
61	62	63	64	65	66	67	68	69	70	
71	72	73	74	75	76	77	78	79	80	
81	82	83	84	85	86	87	88	89	90	
91	92	93	94	95	96	97	98	99	100	
$\sum x_j = t_{s_i}$	460	470	480	490	500	510	520	530	540	550

Tabla 3.1. Orden creciente de  $x_i$  con  $k = 10$

En este caso la varianza es,

$$\begin{aligned} Var_{SY}(\hat{t}_\pi) &= k \sum_{j=1}^k (t_{s_j} - \bar{t})^2 = 10 \sum_{j=1}^{10} (t_{s_j} - \bar{t})^2 \\ &= 10 \left( (460 - 505)^2 + (470 - 505)^2 + \dots + (550 - 505)^2 \right) \\ &= 10(8,250) = 8.25 \cdot 10^4 \end{aligned}$$

$$\text{Donde } \bar{t} = \sum_{j=1}^k \frac{t_{s_j}}{k} = \sum_{j=1}^{10} \frac{t_{s_j}}{10} = 505$$

Como puede observarse la  $Var_{SY}(\hat{t}_\pi)$  es nueve veces menor que la varianza obtenida con un muestreo aleatorio simple.

Debido a que existió un decremento sustancial, solamente al ordenar a las unidades  $x_i$ , se presentan diez ordenamientos distintos adicionales.

Dentro de los estimadores calculados para los distintos ordenamientos utilizados están:

i. La varianza  $Var_{SY}(\hat{t}_\pi) = k \sum_{j=1}^k (t_{s_j} - \bar{t})^2$

ii. La entropía relativa que es calculada como

$$\sum_i^N q(x_i) \log \frac{q(x_i)}{(f(x_i))} = \sum_i^N i \log \frac{i}{(f(x_i))}.$$

iii. La medida delta que es calculada como  $\delta = 1 - \frac{(N-1)}{(N-k)} \left( \frac{SSW}{SST} \right)$

Para mostrar los distintos tipos de ordenamientos, se usará una tabla que muestre los valores  $x_i$  asignados a un orden  $q(x_i)$ . En la tabla de orden creciente se tendría:

$S_1$		...	$S_5$		...	$S_{10}$	
$q(x_1)=1$	$x_1=1$	...	$q(x_5)=5$	$x_5=5$	...	$q(x_{10})=10$	$x_{10}=10$
$q(x_{11})=11$	$x_{11}=11$	...	$q(x_{15})=15$	$x_{15}=15$	...	$q(x_{20})=20$	$x_{20}=20$
...	...	...	...	...	...	...	...
$q(x_{81})=81$	$x_{81}=81$	...	$q(x_{85})=85$	$x_{85}=85$	...	$q(x_{90})=90$	$x_{90}=90$
$q(x_{91})=91$	$x_{91}=91$	...	$q(x_{95})=95$	$x_{95}=95$	...	$q(x_{100})=100$	$x_{100}=100$
$\sum_{S_1} x_i = t_1$			$\sum_{S_5} x_i = t_5$			$\sum_{S_{10}} x_i = t_{10}$	

Para simplificar la notación en adelante no se mostrarán los valores de  $q(x_i)$ , éstos serán los mismos que la tabla anterior y sólo se modificarán los valores de  $x_i$  que sean asignados a un orden.

Analizando el primer ordenamiento Caso 1, se tiene,

Conjunto de posibles muestras									
$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$	$S_7$	$S_8$	$S_9$	$S_{10}$
$x_1=50$	$x_2=49$	$x_3=48$	$x_4=47$	$x_5=46$	$x_6=45$	$x_7=44$	$x_8=43$	$x_9=42$	$x_{10}=41$
$x_{11}=40$	$x_{12}=39$	$x_{13}=38$	$x_{14}=37$	$x_{15}=36$	$x_{16}=35$	$x_{17}=34$	$x_{18}=33$	$x_{19}=32$	$x_{20}=31$
$x_{21}=30$	$x_{22}=29$	$x_{23}=28$	$x_{24}=27$	$x_{25}=26$	$x_{26}=25$	$x_{27}=24$	$x_{28}=23$	$x_{29}=22$	$x_{30}=21$
$x_{31}=20$	$x_{32}=19$	$x_{33}=18$	$x_{34}=17$	$x_{35}=16$	$x_{36}=15$	$x_{37}=14$	$x_{38}=13$	$x_{39}=12$	$x_{40}=11$
$x_{41}=10$	$x_{42}=9$	$x_{43}=8$	$x_{44}=7$	$x_{45}=6$	$x_{46}=5$	$x_{47}=4$	$x_{48}=3$	$x_{49}=2$	$x_{50}=1$
$x_{51}=51$	$x_{52}=52$	$x_{53}=53$	$x_{54}=54$	$x_{55}=55$	$x_{56}=56$	$x_{57}=57$	$x_{58}=58$	$x_{59}=59$	$x_{60}=60$
$x_{61}=61$	$x_{62}=62$	$x_{63}=63$	$x_{64}=64$	$x_{65}=65$	$x_{66}=66$	$x_{67}=67$	$x_{68}=68$	$x_{69}=69$	$x_{70}=70$
$x_{71}=71$	$x_{72}=72$	$x_{73}=73$	$x_{74}=74$	$x_{75}=75$	$x_{76}=76$	$x_{77}=77$	$x_{78}=78$	$x_{79}=79$	$x_{80}=80$
$x_{81}=81$	$x_{82}=82$	$x_{83}=83$	$x_{84}=84$	$x_{85}=85$	$x_{86}=86$	$x_{87}=87$	$x_{88}=88$	$x_{89}=89$	$x_{90}=90$
$x_{91}=91$	$x_{92}=92$	$x_{93}=93$	$x_{94}=94$	$x_{95}=95$	$x_{96}=96$	$x_{97}=97$	$x_{98}=98$	$x_{99}=99$	$x_{100}=100$
505	505	505	505	505	505	505	505	505	505

1. En este caso la varianza es,

$$\begin{aligned}
 V_{SY}(t_{\pi}) &= k \sum_{j=1}^k (t_{s_j} - \bar{t})^2 = 10 \sum_{j=1}^{10} (t_{s_j} - \bar{t})^2 \\
 &= 10((505 - 505)^2 + (505 - 505)^2 + \dots + (505 - 505)^2) \\
 &= 10(0) = 0
 \end{aligned}$$

Donde  $\bar{t} = \sum_{i=1}^k t_{s_i} / k = 5050/10 = 505$  con  $t_{s_i} = \sum_{s_i} y_m$

2. La entropía relativa se calcula como,

$$\sum_i^N q(x_i) \log\left(\frac{q(x_i)}{f(x_i)}\right) = \left(1 \log \frac{1}{50}\right) + \left(2 \log \frac{2}{49}\right) + \dots + \left(100 \log \frac{1}{100}\right) = 1,154.44$$

3. La delta se calcula como,

$$\delta = 1 - \frac{(N-1)}{(N-k)} \left(\frac{SSW}{SST}\right) = 1 - \frac{(99)}{(90)}(1) = 1 - 1.1 = -0.1$$

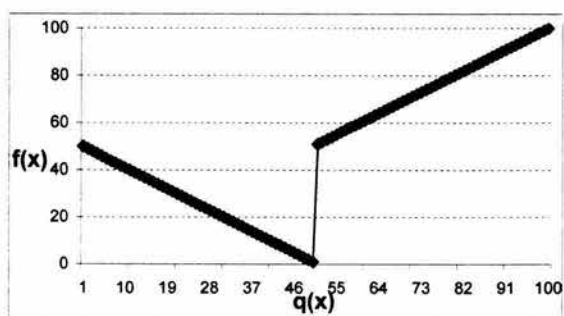
El caso que acabamos de analizar tiene varianza cero, por lo que se decidió generar otros casos con diferente orden, y a cada uno se le evaluará la varianza, la entropía relativa y la delta.

$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$	$S_7$	$S_8$	$S_9$	$S_{10}$
50	49	48	47	46	45	44	43	42	41
40	39	38	37	36	35	34	33	32	31
30	29	28	27	26	25	24	23	22	21
20	19	18	17	16	15	14	13	12	11
10	9	8	7	6	5	4	3	2	1
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100
<b>505</b>	<b>505</b>	<b>505</b>	<b>505</b>	<b>505</b>	<b>505</b>	<b>505</b>	<b>505</b>	<b>505</b>	<b>505</b>

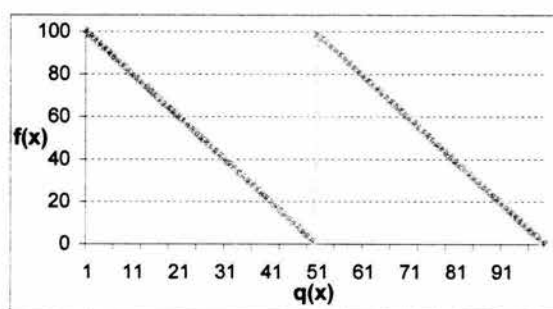
Tabla 3.2. Caso 1

$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$	$S_7$	$S_8$	$S_9$	$S_{10}$
100	98	96	94	92	90	88	86	84	82
80	78	76	74	72	70	68	66	64	62
60	58	56	54	52	50	48	46	44	42
40	38	36	34	32	30	28	26	24	22
20	18	16	14	12	10	8	6	4	2
98	97	95	93	91	89	87	85	83	81
79	77	75	73	71	69	67	65	63	61
59	57	55	53	51	49	47	45	43	41
39	37	35	33	31	29	27	25	23	21
19	17	15	13	11	9	7	5	3	1
<b>594</b>	<b>575</b>	<b>555</b>	<b>535</b>	<b>515</b>	<b>495</b>	<b>475</b>	<b>455</b>	<b>435</b>	<b>415</b>

Tabla 3.4. Caso 2



Gráfica 3.2. Caso 1



Gráfica 3.3. Caso 2

**CASO 1**

<b>Varianza</b>	<b>0</b>
<b>Entropía</b>	<b>1,665.51</b>
<b>Delta</b>	<b>-0.1</b>

Tabla 3.3. Estimadores caso 1

**CASO 2**

<b>Varianza</b>	<b>328,210</b>
<b>Entropía</b>	<b>5,563.40</b>
<b>Delta</b>	<b>-0.0566221</b>

Tabla 3.5. Estimadores caso 2

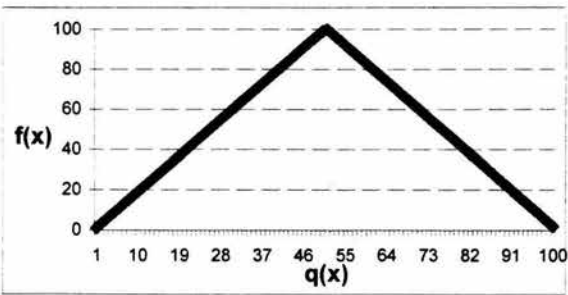


$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$	$s_7$	$s_8$	$s_9$	$s_{10}$
1	3	5	7	9	11	13	15	17	19
21	23	25	27	29	31	33	35	37	39
41	43	45	47	49	51	53	55	57	59
61	63	65	67	69	71	73	75	77	79
81	83	85	87	89	91	93	95	97	99
100	98	96	94	92	90	88	86	84	82
80	78	76	74	72	70	68	66	64	62
60	58	56	54	52	50	48	46	44	42
40	38	36	34	32	30	28	26	24	22
20	18	16	14	12	10	8	6	4	2
505	505	505	505	505	505	505	505	505	505

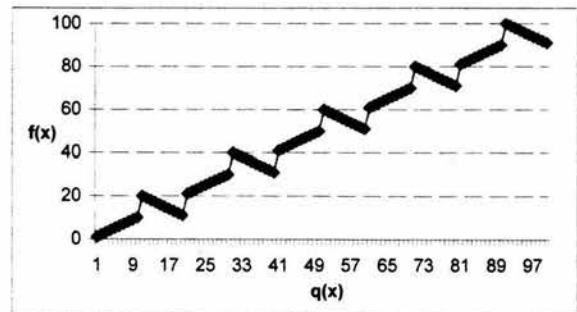
Tabla 3.6. Caso 3

$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$	$s_7$	$s_8$	$s_9$	$s_{10}$
1	2	3	4	5	6	7	8	9	10
20	19	18	17	16	15	14	13	12	11
21	22	23	24	25	26	27	28	29	30
40	39	38	37	36	35	34	33	32	31
41	42	43	44	45	46	47	48	49	50
60	59	58	57	56	55	54	53	52	51
61	62	63	64	65	66	67	68	69	70
80	79	78	77	76	75	74	73	72	71
81	82	83	84	85	86	87	88	89	90
100	99	98	97	96	95	94	93	92	91
505	505	505	505	505	505	505	505	505	505

Tabla 3.8. Caso 4



Gráfica 3.4. Caso 3



Gráfica 3.5. Caso 4

**CASO 3**

<b>Varianza</b>	<b>0</b>
<b>Entropía</b>	<b>2,138.42</b>
<b>Delta</b>	<b>-0.1</b>

Tabla 3.7. Estimadores caso 3

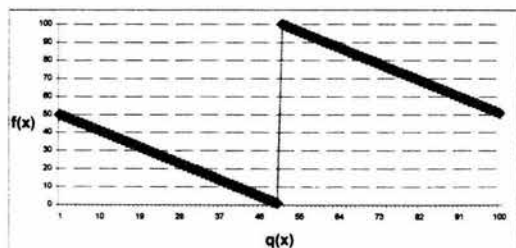
**CASO 4**

<b>Varianza</b>	<b>0</b>
<b>Entropía</b>	<b>32.36</b>
<b>Delta</b>	<b>-0.1</b>

Tabla 3.9. Estimadores caso 4

$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$	$S_7$	$S_8$	$S_9$	$S_{10}$
50	49	48	47	46	45	44	43	42	41
40	39	38	37	36	35	34	33	32	31
30	29	28	27	26	25	24	23	22	21
20	19	18	17	16	15	14	13	12	11
10	9	8	7	6	5	4	3	2	1
100	99	98	97	96	95	94	93	92	91
90	89	88	87	86	85	84	83	82	81
80	79	78	77	76	75	74	73	72	71
70	69	68	67	66	65	64	63	62	61
60	59	58	57	56	55	54	53	52	51
<b>550</b>	<b>540</b>	<b>530</b>	<b>520</b>	<b>510</b>	<b>500</b>	<b>490</b>	<b>480</b>	<b>470</b>	<b>460</b>

Tabla 3.10. Caso 5



Gráfica 3.6. Caso 5

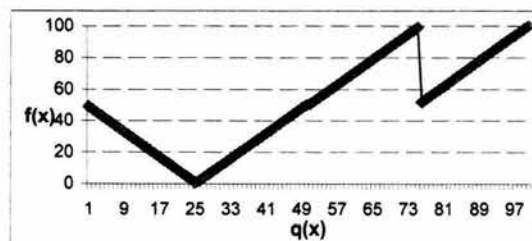
### CASO 5

<b>Varianza</b>	<b>82,500</b>
<b>Entropía</b>	<b>2,072.60</b>
<b>Delta</b>	<b>-0.08910891</b>

Tabla 3.11. Estimadores caso 5

$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$	$S_7$	$S_8$	$S_9$	$S_{10}$
49	47	45	43	41	39	37	35	33	31
29	27	25	23	21	19	17	15	13	11
9	7	5	3	1	2	4	6	8	10
12	14	16	18	20	22	24	26	28	30
32	34	36	38	40	42	44	46	48	50
51	53	55	57	59	61	63	65	67	69
71	73	75	77	79	81	83	85	87	89
91	93	95	97	99	52	54	56	58	60
62	64	66	68	70	72	74	76	78	80
82	84	86	88	90	92	94	96	98	100
<b>488</b>	<b>496</b>	<b>504</b>	<b>512</b>	<b>520</b>	<b>482</b>	<b>494</b>	<b>506</b>	<b>518</b>	<b>530</b>

Tabla 3.12. Caso 6



Gráfica 3.7. Caso 6

### CASO 6

<b>Varianza</b>	<b>20,900</b>
<b>Entropía</b>	<b>1,196.47</b>
<b>Delta</b>	<b>-0.09724092</b>

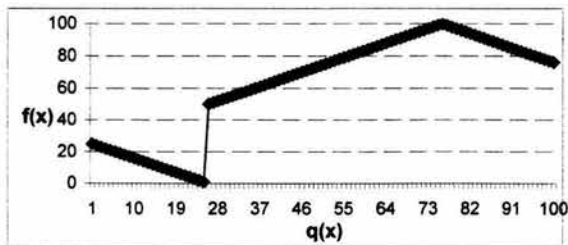
Tabla 3.13. Estimadores caso 6

$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$	$S_7$	$S_8$	$S_9$	$S_{10}$
25	24	23	22	21	20	19	18	17	16
15	14	13	12	11	10	9	8	7	6
5	4	3	2	1	50	51	52	53	54
55	56	57	58	59	60	61	62	63	64
65	66	67	68	69	70	71	72	73	74
75	76	77	78	79	80	81	82	83	84
85	86	87	88	89	90	91	92	93	94
95	96	97	98	99	100	99	98	97	96
95	94	93	92	91	90	89	88	87	86
85	84	83	82	81	80	79	78	77	76
600	600	600	600	600	650	650	650	650	650

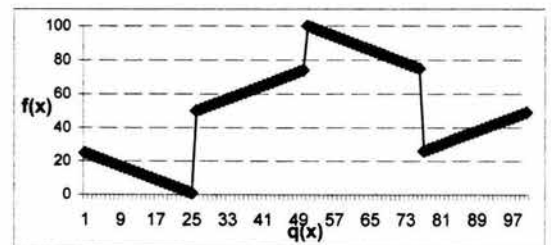
Tabla 3.14. Caso 7

$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$	$S_7$	$S_8$	$S_9$	$S_{10}$
25	24	23	22	21	20	19	18	17	16
15	14	13	12	11	10	9	8	7	6
5	4	3	2	1	50	51	52	53	54
55	56	57	58	59	60	61	62	63	64
65	66	67	68	69	70	71	72	73	74
100	99	98	97	96	95	94	93	92	91
90	89	88	87	86	85	84	83	82	81
80	79	78	77	76	75	26	27	28	29
30	31	32	33	34	35	36	37	38	39
40	41	42	43	44	45	46	47	48	49
505	503	501	499	497	545	497	499	501	503

Tabla 3.16. Caso 8



Gráfica 3.8. Caso 7



Gráfica 3.9. Caso 8

**CASO 7**

<b>Varianza</b>	62,500
<b>Entropía</b>	2,551.72
<b>Delta</b>	-0.09296495

Tabla 3.15. Estimadores caso 7

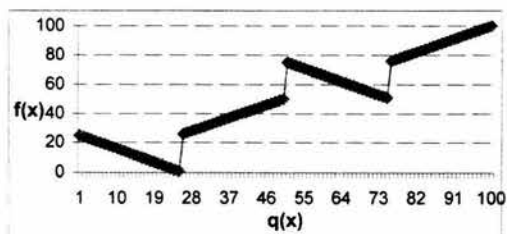
**CASO 8**

<b>Varianza</b>	18,400
<b>Entropía</b>	1,508.16
<b>Delta</b>	-0.09757096

Tabla 3.17. Estimadores caso 8

$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$	$S_7$	$S_8$	$S_9$	$S_{10}$
25	24	23	22	21	20	19	18	17	16
15	14	13	12	11	10	9	8	7	6
5	4	3	2	1	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
75	74	73	72	71	70	69	68	67	66
65	64	63	62	61	60	59	58	57	56
55	54	53	52	51	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100
<b>484</b>	<b>482</b>	<b>480</b>	<b>478</b>	<b>476</b>	<b>526</b>	<b>528</b>	<b>530</b>	<b>532</b>	<b>534</b>

Tabla 3.18. Caso 9



Gráfica 3.10. Caso 9

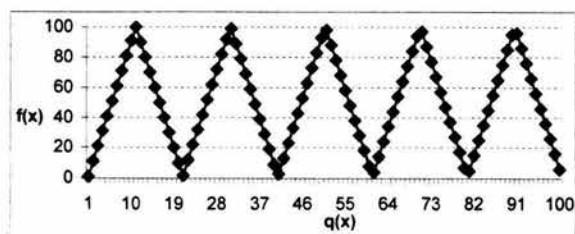
**CASO 9**

<b>Varianza</b>	<b>62,256</b>
<b>Entropía</b>	<b>453.45</b>
<b>Delta</b>	<b>-0.09164356</b>

Tabla 3.19. Estimadores caso 9

$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$	$S_7$	$S_8$	$S_9$	$S_{10}$
1	11	21	31	41	51	61	71	81	91
100	90	80	70	60	50	40	30	20	10
2	12	22	32	42	52	62	72	82	92
99	89	79	69	59	49	39	29	19	9
3	13	23	33	43	53	63	73	83	93
98	88	78	68	58	48	38	28	18	8
4	14	24	34	44	54	64	74	84	94
97	87	77	67	57	47	37	27	17	7
5	15	25	35	45	55	65	75	85	95
96	86	76	66	56	46	36	26	16	6
<b>505</b>	<b>505</b>	<b>505</b>	<b>505</b>	<b>505</b>	<b>505</b>	<b>505</b>	<b>505</b>	<b>505</b>	<b>505</b>

Tabla 3.20. Caso 10



Gráfica 3.11. Caso 10

**CASO 10**

<b>Varianza</b>	<b>0</b>
<b>Entropía</b>	<b>3,120.55</b>
<b>Delta</b>	<b>-0.1</b>

Tabla 3.21. Estimadores caso 10

	VARIANZA $Var(\hat{t}_\pi)$	ENTROPÍA RELATIVA	DELTA $\delta$
Caso1	0	1,154.44	-0.1
Caso3 <sup>1</sup>	0	2,285.75	-0.1
Caso4	0	22.43	-0.1
Caso10	0	973.12	-0.1
Caso8	18,400	1,151.78	-0.097571
Caso6	29,900	631.57	-0.097241
Caso9	62,256	314.31	-0.091644
Caso7	62,500	-667.61	-0.092965
Caso 0 <sup>2</sup>	82,500	0	-0.089
Caso5	82,500	1,436.61	-0.089109
Caso2	328,210	3,645.01	-0.056622
Caso MAS	757,500		

Tabla 3.22. Concentrado de valores estimados

Este primer ejemplo tiene varios resultados interesantes:

- i. El caso que se aproxima más a una curva (caso 3), fue uno de los cuatro que tuvo varianza 0 (varianza mínima), resultado esperado por las conclusiones del área de la Estereología.

<sup>1</sup> Caso 3, es el correspondiente al orden tipo curva.

<sup>2</sup> Caso 0, es el correspondiente al orden creciente.

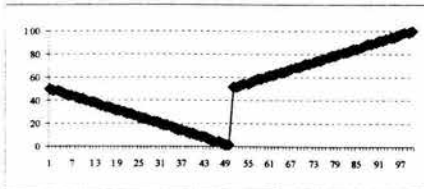
- ii. No existe necesariamente un orden único que minimice la varianza, para este ejemplo al menos cuatro casos tuvieron varianza 0.
- iii. El caso 3, tuvo la mayor entropía relativa en comparación con los otros tres casos donde la varianza fue cero.
- iv. El estimador delta  $\delta$  alcanzó su mínimo en los cuatro casos donde la varianza fue cero, definido como  $1 - \frac{(N-1)}{(N-k)} = 1 - \frac{(99)}{(90)} = -0.1$ .
- v. El valor máximo que podía alcanzar delta  $\delta$  era 1, sin embargo el caso con mayor delta fue el 2, registrando un valor de  $-0.056622$ , lo que significa que los datos tienen una heterogeneidad alta dentro de las 10 muestras.
- vi. Todos aquéllos que tuvieron varianza 0, registraron un  $SSB = 0$ , es decir hubo total homogeneidad entre las muestras, y el valor de  $\delta$  alcanzó su mínimo total.
- vii. Todos los casos con orden diferente al aleatorio tuvieron una mejor eficiencia que el muestreo aleatorio simple. El caso ordenado que mayor varianza reportó fue el caso 2, y aún la varianza del muestreo aleatorio simple fue 2.3 veces mayor que esta última. Aunque posiblemente existan tipos de orden donde la varianza bajo un diseño de muestreo aleatorio simple sea menor que la varianza de un muestreo sistemático.
- viii. En el capítulo 2, se hace la demostración que la entropía relativa en orden creciente (Caso 0) es menor a la del orden curva (Caso 3). Este resultado también queda demostrado en este primer ejemplo. Es importante destacar que sólo los casos 2 y 7, salieron del intervalo de la entropía formado por el orden creciente y orden curva.

	Valor mínimo	Valor máximo
		757,500 (Caso MAS)
<b>Varianza</b>	0 (Casos 1,3,4,10)	Dentro de los de muestreo sistemático  328,210 (Caso 2)
<b>Entropía</b>	0 (Caso 0)	3,645.01 (Caso 2)
<b>Delta</b>	-0.1 (Casos 1,3,4,10)	-0.056622 (Caso 2)

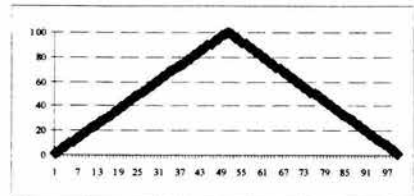
Tabla 3.23. Valores mínimos y máximos de los estimadores

**Casos donde la varianza fue cero**

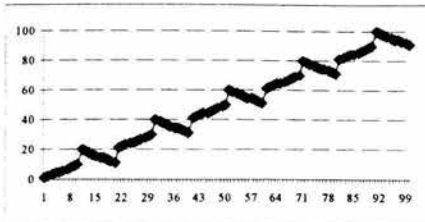
**Caso 1**



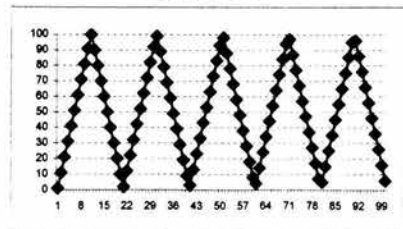
**Caso 3**



**Caso 4**



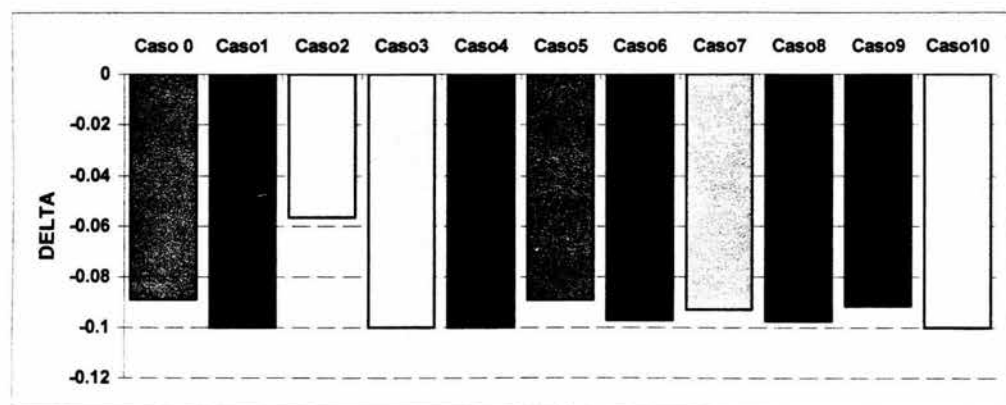
**Caso 10**



Las similitudes que tienen estos cuatro casos, muestran que tuvieron varianza cero con una  $n = 10$ :

- Líneas con pendiente igual pero con signo contrario
- Dichas líneas dividen a la población de tal manera que las unidades que se encuentran sobre cada línea es un total par. Ejemplo, Caso 1, 50, Caso 3, 50, Caso 4, 10, Caso 10, 10.

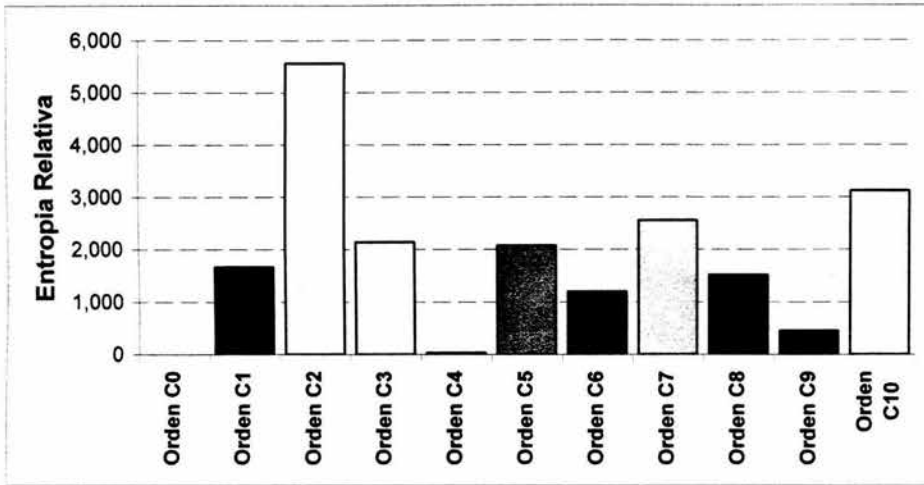
### DELTA



*Gráfica 3.12. Medida de homogeneidad, delta para los once casos.*

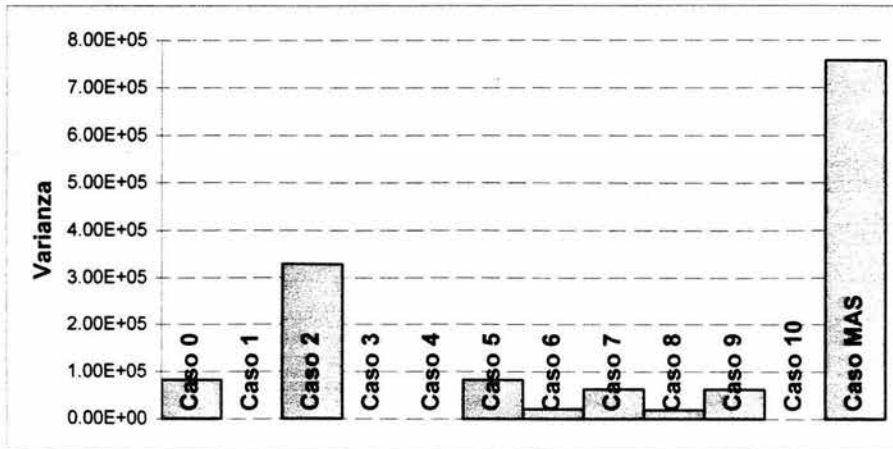


### ENTROPÍA RELATIVA



Gráfica 3.13. Entropía relativa para los once casos.

### VARIANZA



Gráfica 3.14. Varianza para los once casos.

## **3.2. ELECCIONES PRESIDENCIALES DE MÉXICO EN EL 2000 A NIVEL ESTATAL Y DISTRITO ELECTORAL.**

---

Este análisis está basado en los resultados oficiales de las elecciones presidenciales de México en el año 2000.

Se decidió estudiar esta información porque son datos reales, interesantes y complejos, y por lo tanto las conclusiones de este trabajo, toman un sentido más práctico.

Igual que los otros ejemplos que se incluyen en este trabajo, las unidades poblacionales de muestreo  $u = \{x, y, a, z, \dots, l\}$  se ordenan de distintas maneras (creciente, tipo curva y aleatorio), en este ejemplo las unidades poblacionales son las secciones electorales, tomando como referencia a la variable de referencia  $l$ , que en este ejemplo es la Lista Nominal. En el primer ejemplo los datos fueron generados y se evaluaron once tipos de orden distintos, lo anterior se logró gracias a que las unidades  $y_i$  eran igualmente espaciadas, la  $N = 100$  era "pequeña", el tamaño de muestra era  $n = 10$  y sólo había una variable.

En la práctica no suelen darse condiciones similares. En este ejemplo de elecciones presidenciales, la información es compleja. De manera similar lo que se busca es ver cuál de estos tres tipos de orden resulta ser más eficiente y bajo qué condiciones.

Las variables a considerar en este ejemplo son:

**Lista Nominal  $I$ .** En la lista nominal se encuentran todos aquellos ciudadanos que solicitaron su inscripción al Padrón Electoral y cuentan ya con su credencial para votar con fotografía vigente. De esta manera, el día de la jornada electoral, en las casillas electorales, se encontrarán las listas nominales con los nombres de todos aquellos ciudadanos, que acudieron oportunamente a tramitar su credencial para votar y tienen el derecho de ejercer libremente su voto.

**Total  $X$ .** Es la suma de todos los votos correspondiente a los grupos políticos participantes.

Las variables que indican el número de votos hacia los **grupos políticos  $Y$**  que participaron en las elecciones presidenciales del año 2000:

GRUPO POLÍTICO	DESCRIPCIÓN	GRUPO POLÍTICO	DESCRIPCIÓN
Alianza por el cambio (A_CAM)	Conformado por los partidos del Acción Nacional y el Verde Ecologista.	Alianza por México (A_MEX)	Conformado por el Partido de la Revolución Democrática, el Partido del Trabajo y otros.
PRI	Conformado por el partido Revolucionario Institucional.	PCD, PARM Y DSPPN	Los grupos políticos que también participaron en estas elecciones presidenciales, pero que en conjunto solo tuvieron un impacto del 2.55% en los resultados a nivel nacional.

*Tabla 3.2.1. Grupos políticos considerados en el análisis.*

Debido a que la relación entre Lista Nominal y el voto es muy débil o nula a nivel nacional, se decidió trabajar a nivel entidad federativa, eligiendo a tres de ellas para esta evaluación.

Los estados elegidos fueron:

Entidad Federativa	Grupo político que ganó las elecciones presidenciales a nivel entidad federativa
Guanajuato	Acción por el Cambio A_CAM
Sinaloa	Partido Revolucionario Institucional PRI
Michoacán	• Alianza por México A_MEX

En este ejemplo, se deseaba evaluar la eficiencia del muestreo sistemático bajo el orden curva y creciente con un estimador de razón, así como un muestreo aleatorio simple.

Defínase la razón como  $R = Y/X$ , donde,

- $X$  = Es el número total de votos efectuados en la población a nivel entidad federativa.
- $Y$  = Todos los votos a favor de un grupo político en específico a nivel entidad federativa.



La variable de referencia sobre la cual se ordenan los datos es la **Lista Nominal  $l$** .

También se decidió elegir varios tamaños de muestra  $n$ , estos fueron: 40, 80, 120, 160 y 200.

La hipótesis inicial fue que el orden en curva de las unidades poblacionales  $u$  sería más eficiente que el orden creciente para las estimaciones de razón.

### **Metodología**

Los pasos utilizados para la medición de la eficiencia del orden seleccionado fueron:

- i. Ordenar por medio de la variable de Lista Nominal  $l$  a las unidades poblacionales bajo el criterio definido (Curva , creciente  o aleatorio)
- ii. Obtención de todas las muestras sistemáticas dependiendo del tamaño de  $n$  (sólo cuando se tiene un orden tipo curva o creciente).
- iii. Cálculo del error cuadrático medio ECM dependiendo del tipo de orden.
- iv. Cálculo de la medida de homogeneidad,  $\delta$ . Dado que depende de  $k$  sólo se mide para los tipos de orden: curva y creciente.
- v. La entropía relativa se calcula sólo a nivel entidad federativa, es decir no se mide la entropía por cada muestra sistemática. No depende del tamaño de las muestras  $n$ . La entropía sólo se mide para los tipos de orden: curva y creciente.

Las variables utilizadas para el error cuadrático medio fueron:

1.  $X$  = El número total de votos efectuados en la muestra seleccionada.
2.  $Y$  = Todos los votos a favor de un partido en específico en la muestra seleccionada.
3. Se calculo el estimador de razón como el cociente entre  $Y$  y  $X$ ,  $R = Y/X$ , que es la razón de votantes por determinado grupo político.
4. El error cuadrático medio para los tipos de orden curva y creciente<sup>1</sup> se calculó como:

$$ECM(\hat{R}_{SY}) = \frac{k}{X^2} \sum_{i=1}^k (Y_i - RX_i)^2$$

donde  $k = N/n$

5. El error cuadrático medio para el tipo de orden aleatorio<sup>2</sup> se calculó como:

$$ECM(\hat{R}_{MAS}) = \frac{N^2}{X^2} \left(1 - \frac{n}{N}\right) \frac{1}{n} \sum_{i=1}^N \frac{(Y_i - \hat{R}X_i)^2}{N-1}$$

6. La entropía relativa fue calculada como:

$$D(q | f) = \sum_x q(x) \log \frac{q(x)}{f(x)}$$

---

<sup>1</sup> Se calculó bajo el muestreo sistemático.

<sup>2</sup> Se calculó bajo el muestreo aleatorio simple.

Donde  $q(x)$  es el orden que se le asigna a cada valor de la Lista Nominal  $l = f(x)$ .

Delta fue calculada como:

$$\delta = 1 - \left( \frac{N-1}{N-k} \right) \left( \frac{SSW}{SST} \right)$$

Las gráficas 3.2.1 a 3.2.9 presentan a la variable Lista Nominal  $l$ , ordenada bajos los tipos: curva, creciente y aleatorio, de los tres estados seleccionados. Enseguida se presentan cinco tablas por entidad federativa que contienen los errores cuadráticos medios dependiendo del tamaño de muestra para los tres tipos de orden.

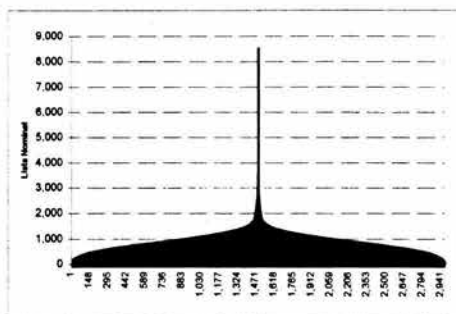
Por último se exponen unos concentrados y comentarios de los resultados obtenidos con respecto al error cuadrático medio.

#### Forma de ordenar a la variable Lista Nominal.

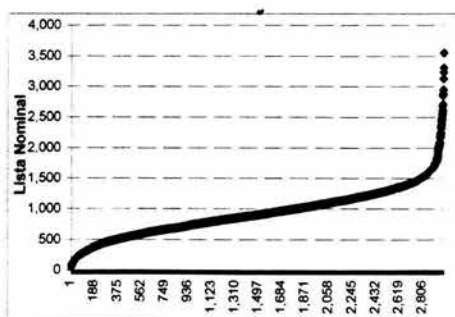
Orden Tipo Curva	
Orden $q(l)$	Lista Nominal $l$
1	49
2	52
3	71
4	73
5	81
6	98
7	104
...	...
2971	103
2972	81
2973	78
2974	73
2975	58
2976	51

Orden Tipo Creciente	
Orden $q(l)$	Lista Nominal $l$
1	49
2	51
3	52
4	58
5	71
6	73
7	73
...	...
2971	2949
2972	3131
2973	3234
2974	3307
2975	3556
2976	8528

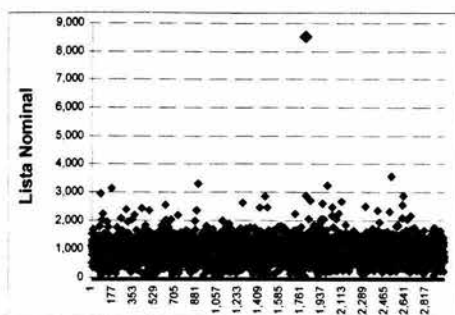
### Guanajuato



Gráfica 3.2.1. Lista Nominal ordenada en forma de curva a nivel estado (Guanajuato).



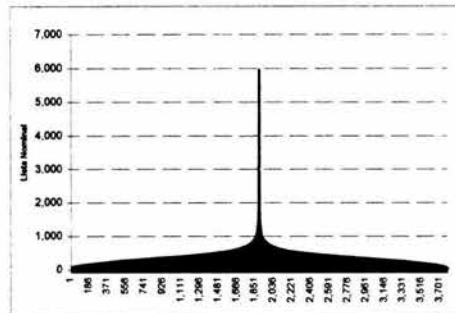
Gráfica 3.2.2. Lista Nominal ordenada en forma creciente a nivel estado (Guanajuato).



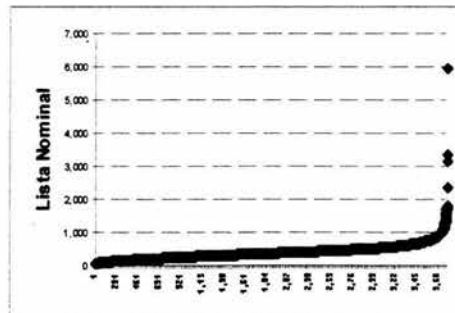
Gráfica 3.2.3. Lista Nominal ordenada en forma aleatoria a nivel estado (Guanajuato).



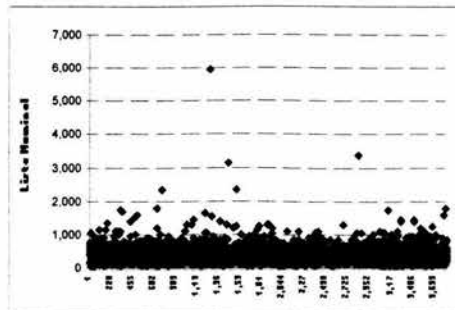
### Sinaloa



Gráfica 3.2.4. Lista Nominal ordenada en forma curva a nivel estado (Sinaloa).

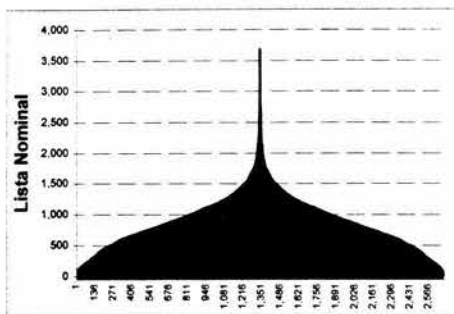


Gráfica 3.2.5. Lista Nominal ordenada en forma creciente a nivel estado (Sinaloa).

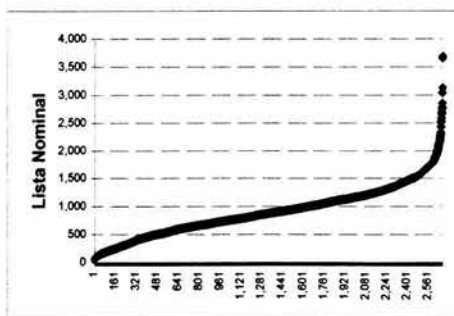


Gráfica 3.2.6. Lista Nominal ordenada en forma aleatoria a nivel estado (Sinaloa).

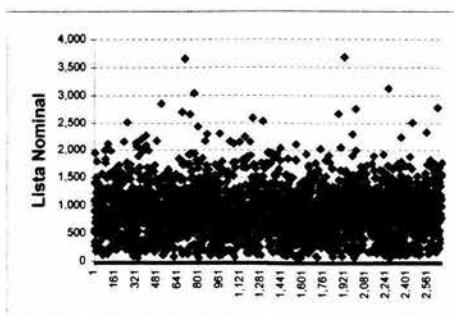
### **Michoacán**



*Gráfica 3.2.7. Lista Nominal ordenada en forma curva a nivel estado (Michoacán).*



*Gráfica 3.2.8. Lista Nominal ordenada en forma creciente a nivel estado (Michoacán)*



*Gráfica 3.2.9. Lista Nominal ordenada en forma aleatoria a nivel estado (Michoacán)*

## CÁLCULO DEL ECM PARA MUESTRAS DEL ESTADO DE MICHOCÁN BAJO DISTINTOS ÓRDENES



PARTIDO	CURVA	CRECIENTE	ALEATORIO
A_CAM	10.5959893	<b>10.1301941</b>	115.006275
A_MEX	<b>7.96084286</b>	9.81705301	56.0652536
PRI	<b>3.24666157</b>	3.99144128	62.5753598
DSPPN	0.00024686	<b>0.00023304</b>	0.00029399
PARM	<b>0.00002746</b>	0.00002894	0.00009787
PCD	<b>0.00005901</b>	0.00006746	0.00007215

*Tabla 3.2.2. Errores Cuadráticos Medios (ECM) multiplicados por 100 con una  $n = 40$ .*

PARTIDO	CURVA	CRECIENTE	ALEATORIO
A_CAM	5.11016811	<b>4.09457497</b>	56.6295614
A_MEX	3.82065223	<b>3.69680168</b>	27.6067608
PRI	<b>1.40882737</b>	1.48394726	30.8123638
DSPPN	0.00007661	<b>0.00004773</b>	0.00014476
PARM	0.00002600	<b>0.00001413</b>	0.00004819
PCD	<b>0.00001985</b>	0.00004050	0.00003553

*Tabla 3.2.3. Errores Cuadráticos Medios (ECM) multiplicados por 100 con una  $n = 80$ .*

<b>PARTIDO</b>	<b>CURVA</b>	<b>CRECIENTE</b>	<b>ALEATORIO</b>
A_CAM	<b>2.01869086</b>	3.56484962	37.1706569
A_MEX	<b>2.22221559</b>	3.25912826	18.1205966
PRI	<b>0.60101761</b>	0.86525426	20.2246985
DSPPN	<b>0.00007051</b>	0.00008841	0.00009502
PARM	<b>0.00001202</b>	0.00002011	0.00003163
PCD	0.00002381	<b>0.00001684</b>	0.00002332

*Tabla 3.2.4. Errores Cuadráticos Medios (ECM) multiplicados por 100 con una  $n = 120$ .*

<b>PARTIDO</b>	<b>CURVA</b>	<b>CRECIENTE</b>	<b>ALEATORIO</b>
A_CAM	<b>1.69610856</b>	2.05042911	27.4412047
A_MEX	1.76288781	<b>1.38763893</b>	13.3775145
PRI	<b>0.48956478</b>	0.64772855	14.9308659
DSPPN	0.00002806	<b>0.00001776</b>	0.00007015
PARM	0.00000982	<b>0.00000570</b>	0.00002335
PCD	<b>0.00001204</b>	0.00002296	0.00001722

*Tabla 3.2.5. Errores Cuadráticos Medios (ECM) multiplicados por 100 con una  $n = 160$ .*

<b>PARTIDO</b>	<b>CURVA</b>	<b>CRECIENTE</b>	<b>ALEATORIO</b>
A_CAM	<b>1.38587303</b>	1.54265565	21.6035334
A_MEX	<b>0.55986074</b>	1.91009105	10.5316652
PRI	<b>0.38871112</b>	0.68117944	11.7545663
DSPPN	<b>0.00002157</b>	0.00003056	0.00005522
PARM	0.00000519	<b>0.00000264</b>	0.00001838
PCD	0.00001325	<b>0.00000796</b>	0.00001355

*Tabla 3.2.6. Errores Cuadráticos Medios (ECM) multiplicados por 100 con una  $n = 200$ .*

## CÁLCULO DEL ECM PARA MUESTRAS DEL ESTADO DE GUANAJUATO BAJO DISTINTOS ÓRDENES



PARTIDO	CURVA	CRECIENTE	ALEATORIO
A_CAM	<b>0.000687</b>	0.000692	0.000978
A_MEX	<b>0.000096</b>	0.000116	0.000194
PRI	<b>0.000400</b>	0.000401	0.000442
DSPPN	0.000182	<b>0.000166</b>	0.000220
PARM	<b>0.000146</b>	0.000172	0.000204
PCD	0.000483	<b>0.000260</b>	0.000552

*Tabla 3.2.7. Errores Cuadráticos Medios (ECM) multiplicados por 100 con una  $n = 40$ .*

PARTIDO	CURVA	CRECIENTE	ALEATORIO
A_CAM	<b>3.41123384</b>	4.36894502	4.82284803
A_MEX	<b>0.57258390</b>	0.61171083	0.95904777
PRI	<b>1.55340951</b>	2.29439627	2.17958554
DSPPN	0.00013034	<b>0.00009440</b>	0.00010856
PARM	0.00012873	0.00012180	<b>0.00010066</b>
PCD	0.00020270	<b>0.00017103</b>	0.00027256

*Tabla 3.2.8. Errores Cuadráticos Medios (ECM) multiplicados por 100 con una  $n = 80$ .*

PARTIDO	CURVA	CRECIENTE	ALEATORIO
A_CAM	<b>1.80326835</b>	2.19659001	3.17082274
A_MEX	<b>0.31013988</b>	0.37888233	0.63053417
PRI	<b>0.96164754</b>	1.00933312	1.43298718
DSPPN	<b>0.00006475</b>	0.00009430	0.00007137
PARM	0.00008871	<b>0.00005992</b>	0.00006618
PCD	<b>0.00016243</b>	0.00017271	0.00017920

*Tabla 3.2.9. Errores Cuadráticos Medios (ECM) multiplicados por 100 con una  $n = 120$ .*

PARTIDO	CURVA	CRECIENTE	ALEATORIO
A_CAM	2.63581947	3.15508125	<b>2.34481009</b>
A_MEX	0.32935144	<b>0.32335685</b>	0.46627737
PRI	1.10441173	1.31046869	<b>1.05968799</b>
DSPPN	0.00007443	<b>0.00003145</b>	0.00005278
PARM	0.00006961	0.00005095	<b>0.00004894</b>
PCD	0.00013662	<b>0.00010039</b>	0.00013251

*Tabla 3.2.10. Errores Cuadráticos Medios (ECM) multiplicados por 100 con una  $n = 160$ .*

PARTIDO	CURVA	CRECIENTE	ALEATORIO
A_CAM	<b>0.59355242</b>	0.95447455	1.84920250
A_MEX	<b>0.15830072</b>	0.38194051	0.36772329
PRI	0.55456273	<b>0.35450067</b>	0.83570849
DSPPN	0.00003609	<b>0.00001599</b>	0.00004162
PARM	<b>0.00002572</b>	0.00004917	0.00003860
PCD	0.00013213	<b>0.00004818</b>	0.00010451

*Tabla 3.2.11. Errores Cuadráticos Medios (ECM) multiplicados por 100 con una  $n = 200$ .*

## CÁLCULO DEL ECM PARA MUESTRAS DEL ESTADO DE SINALOA BAJO DISTINTOS ÓRDENES



PARTIDO	CURVA	CRECIENTE	ALEATORIO
A_CAM	<b>5.672954760</b>	6.682522640	7.178540534
A_MEX	1.775122227	<b>1.611776668</b>	5.391972676
PRI	6.307063445	<b>5.364922677</b>	303.463347886
DSPPN	<b>0.00019147</b>	0.00021166	0.00034912
PARM	0.00002426	<b>0.00002245</b>	0.00021050
PCD	<b>0.00005877</b>	0.00007760	0.00006506

*Tabla 3.2.12. Errores Cuadráticos Medios (ECM) multiplicados por 100 con una  $n = 40$ .*

PARTIDO	CURVA	CRECIENTE	ALEATORIO
A_CAM	3.245405936	<b>3.117395414</b>	3.550872083
A_MEX	0.970364364	<b>0.714413037</b>	2.667144548
PRI	3.727347873	<b>2.635870034</b>	150.108441272
DSPPN	0.00007831	<b>0.00007625</b>	0.00017269
PARM	0.00001265	<b>0.00001098</b>	0.00010412
PCD	0.00003792	0.00003627	<b>0.00003218</b>

*Tabla 3.2.13. Errores Cuadráticos Medios (ECM) multiplicados por 100 con una  $n = 80$ .*

<b>PARTIDO</b>	<b>CURVA</b>	<b>CRECIENTE</b>	<b>ALEATORIO</b>
A_CAM	<b>2.252831285</b>	2.257502661	2.341649266
A_MEX	<b>0.545344166</b>	0.572085717	1.758868505
PRI	2.676461930	<b>1.542027436</b>	98.990139067
DSPPN	<b>0.00005312</b>	0.00006995	0.00011388
PARM	<b>0.00000749</b>	0.00000865	0.00006867
PCD	<b>0.00001443</b>	0.00002147	0.00002122

*Tabla 3.2.14. Errores Cuadráticos Medios (ECM) multiplicados por 100 con una  $n = 120$ .*

<b>PARTIDO</b>	<b>CURVA</b>	<b>CRECIENTE</b>	<b>ALEATORIO</b>
A_CAM	<b>1.448624268</b>	1.882381630	1.737037857
A_MEX	<b>0.295034324</b>	0.424007628	1.304730484
PRI	<b>1.317636612</b>	1.465994419	73.430987965
DSPPN	<b>0.00003998</b>	0.00005069	0.00008448
PARM	<b>0.00000504</b>	0.00000605	0.00005094
PCD	0.00002763	0.00001965	<b>0.00001574</b>

*Tabla 3.2.15. Errores Cuadráticos Medios (ECM) multiplicados por 100 con una  $n = 160$ .*

<b>PARTIDO</b>	<b>CURVA</b>	<b>CRECIENTE</b>	<b>ALEATORIO</b>
A_CAM	0.981660289	<b>0.684133107</b>	1.374271012
A_MEX	0.478085636	<b>0.359676678</b>	1.032247671
PRI	1.332865907	<b>0.617326052</b>	58.095497303
DSPPN	0.00003850	<b>0.00003669</b>	0.00006684
PARM	0.00000559	<b>0.00000307</b>	0.00004030
PCD	<b>0.00001044</b>	0.00001574	0.00001246

*Tabla 3.2.16. Errores Cuadráticos Medios (ECM) multiplicados por 100 con una  $n = 200$ .*



La siguiente tabla indica el número de grupos políticos que tuvieron el menor error cuadrático medio, con respecto a los tres tipos de orden<sup>3</sup>.

<i>n</i>	Michoacán			Guanajuato			Sinaloa		
	Curva	Creciente	Aleatorio	Curva	Creciente	Aleatorio	Curva	Creciente	Aleatorio
40	2	1	0	3	0	0	1	2	0
80	1	2	0	3	0	0	0	3	0
120	3	0	0	3	0	0	2	1	0
160	2	1	0	0	1	2	3	0	0
200	3	0	0	2	1	0	0	3	0
<b>Total</b>	<b>11</b>	<b>4</b>	<b>0</b>	<b>11</b>	<b>2</b>	<b>2</b>	<b>6</b>	<b>9</b>	<b>0</b>
<b>%</b>	<b>73%</b>	<b>26%</b>	<b>0%</b>	<b>73%</b>	<b>13%</b>	<b>13%</b>	<b>40%</b>	<b>60%</b>	<b>0%</b>

**Tabla 3.2.17. Número de Grupos políticos que tuvieron un menor error cuadrático medio con respecto a los tres tipos de orden (curva, creciente, aleatorio<sup>4</sup>).**

Es interesante hacer notar que en dos de los tres estados (Michoacán y Guanajuato), el 73% de los casos fue más eficiente el orden curva que el orden creciente.

También es importante resaltar que el muestreo aleatorio simple resultó ser el menos eficiente.

El que un tipo de orden sea en términos generales más eficiente que otro en una determinada población, no indica que para todas las variables sea así.

<sup>3</sup> Solo se consideran los grupos políticos A\_CAM, A\_MEX y PRI, no así los demás partidos políticos debido a su reducido número de votos.

<sup>4</sup> ALEATORIO, se refiere a que se obtuvo de un muestreo aleatorio simple.

Por ejemplo, la razón de votantes hacia el grupo político PRI, siempre tuvo una mejor eficiencia en Michoacán con un tipo de orden curva. Lo mismo ocurre para el grupo Alianza por México A\_MEX en Guanajuato.

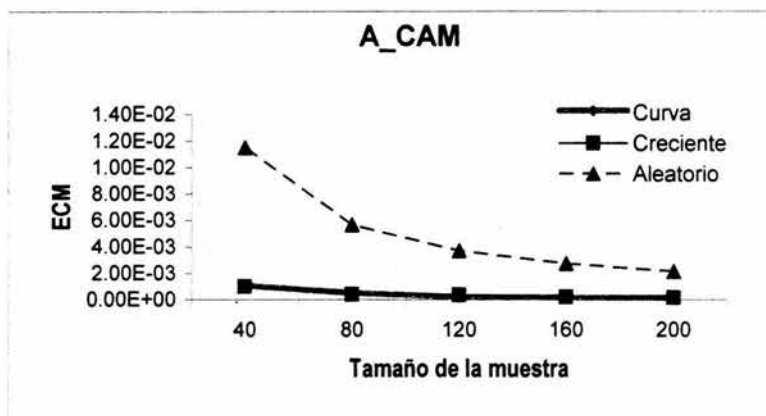
Como se puede observar en la tabla 3.2.17, el estado de Sinaloa no tuvo una tendencia importante hacia un tipo de orden, aunque en el 60% de los casos fue menor el error cuadrático medio en orden tipo creciente.

Otro resultado importante es que al cambiar el tamaño de la muestra, puede cambiar el tipo de orden más eficiente. Por lo que evaluar con una  $n$  específica no garantiza que los resultados obtenidos sean similares si se cambia la  $n$ .

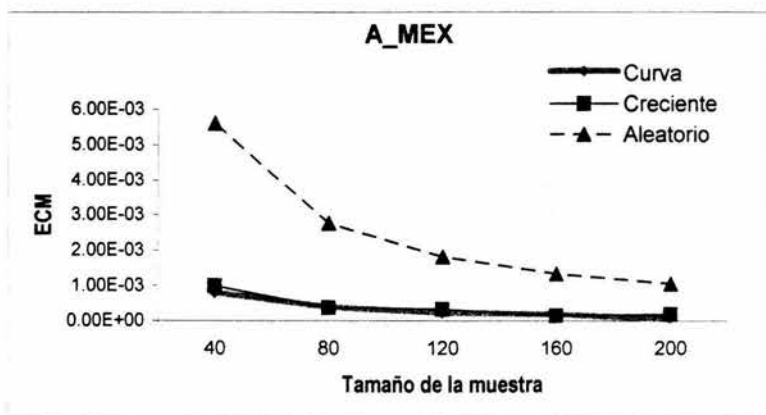
Otro punto interesante, es que la gráfica tipo curva de la variable Lista Nominal del estado de Sinaloa, es muy puntiaguda, es decir, a pesar del orden tipo curva, el histograma no tiene una forma parecida a la de una distribución normal.

A continuación se muestran algunas gráficas de los errores cuadráticos medios presentados anteriormente.

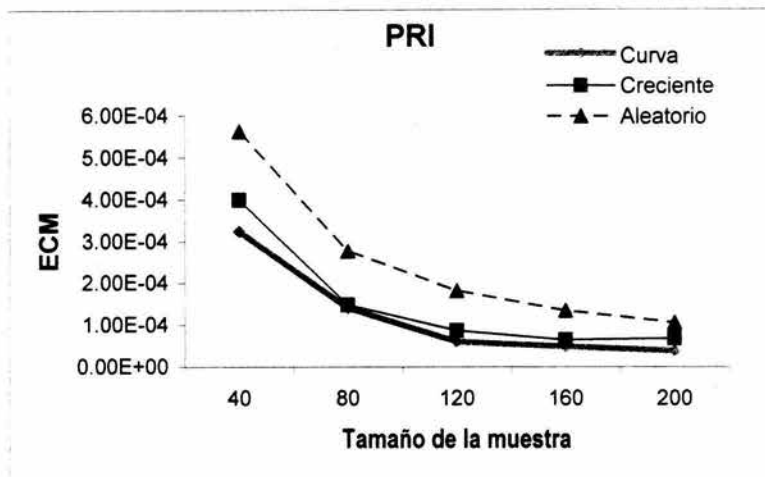
## MICHOACÁN



Gráfica 3.2.10. Errores Cuadráticos Medios (ECM) correspondientes a Michoacán del grupo político A\_CAM.

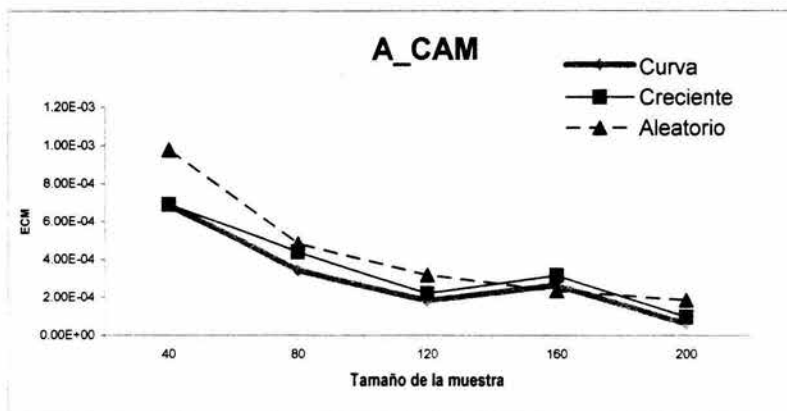


Gráfica 3.2.11. Errores Cuadráticos Medios (ECM) correspondientes a Michoacán del grupo político A\_MEX.

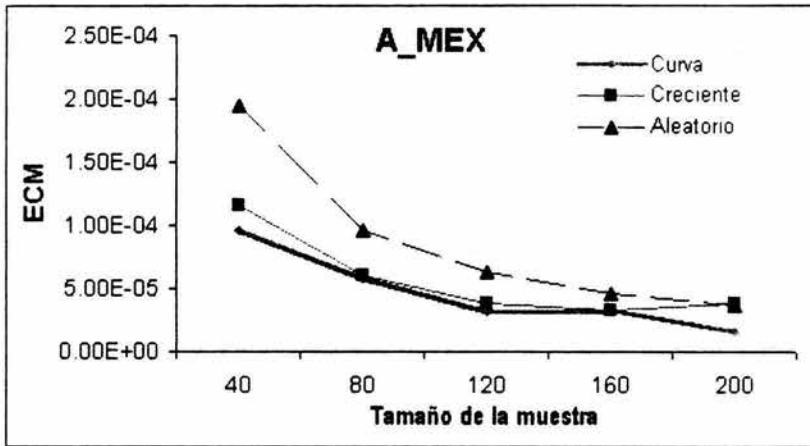


Gráfica 3.2.12. Errores Cuadráticos Medios (ECM) correspondientes a Michoacán del grupo político PRI.

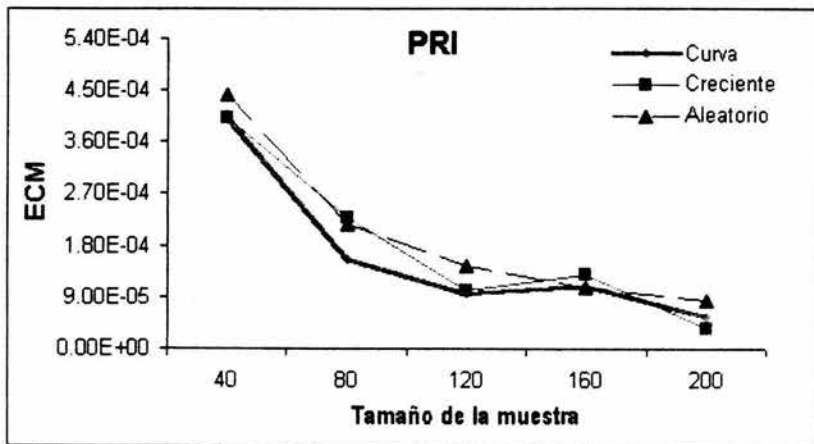
## GUANAJUATO



Gráfica 3.2.13. Errores Cuadráticos Medios (ECM) correspondientes a Guanajuato del grupo político A\_CAM.

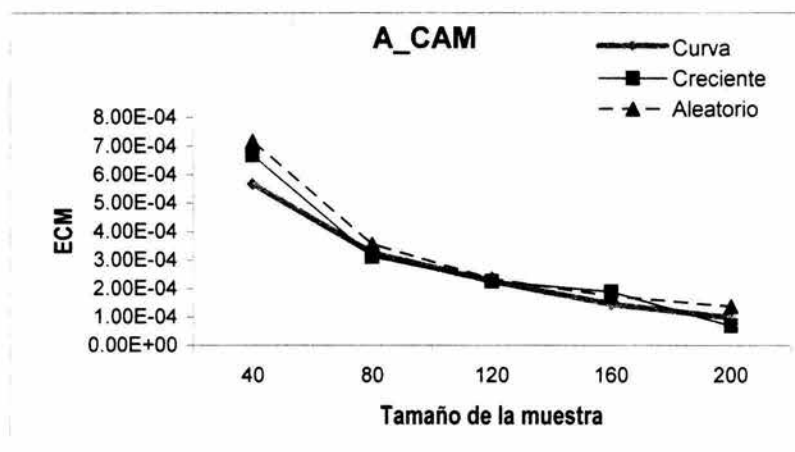


Gráfica 3.2.14. Errores Cuadráticos Medios (ECM) correspondientes a Guanajuato del grupo político A\_MEX.

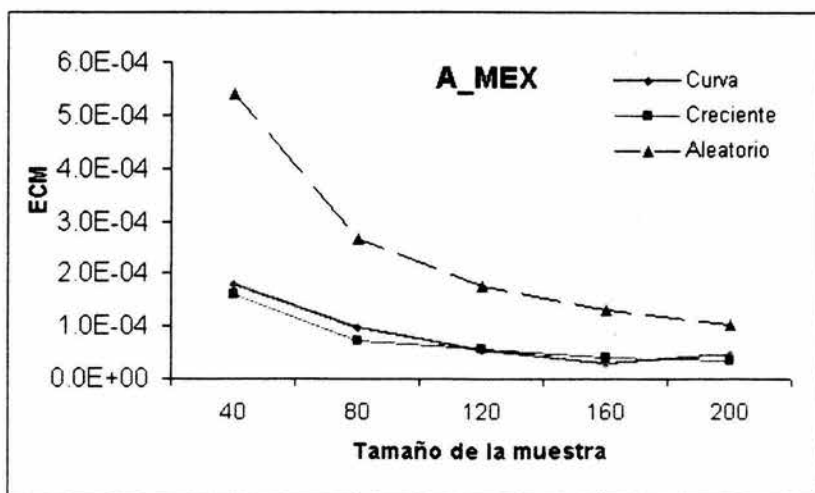


Gráfica 3.2.15. Errores Cuadráticos Medios (ECM) correspondientes a Guanajuato del grupo político PRI.

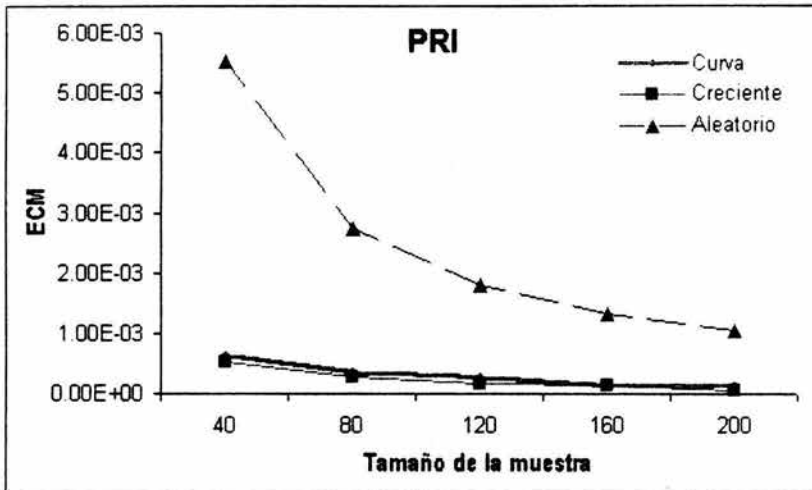
## SINALOA



Gráfica 3.2.16. Errores Cuadráticos Medios (ECM) correspondientes a Sinaloa del grupo político A\_CAM.



Gráfica 3.2.17. Errores Cuadráticos Medios (ECM) correspondientes a Sinaloa del grupo político A\_MEX.



**Gráfica 3.2.18. Errores Cuadráticos Medios (ECM) correspondientes a Sinaloa del grupo político PRI.**

En todas las gráficas es evidente el hecho de que el orden tipo curva o tipo creciente son mucho más eficientes que el aleatorio simple. Dicha diferencia es mayor entre la  $n$  es más pequeña. También es importante anotar que en algunos casos de orden tipo creciente y curva el ECM es menor cuando  $n = 120$  y no en  $n = 200$ , esto se debe a la manera de seleccionar las muestras y por los outliers.

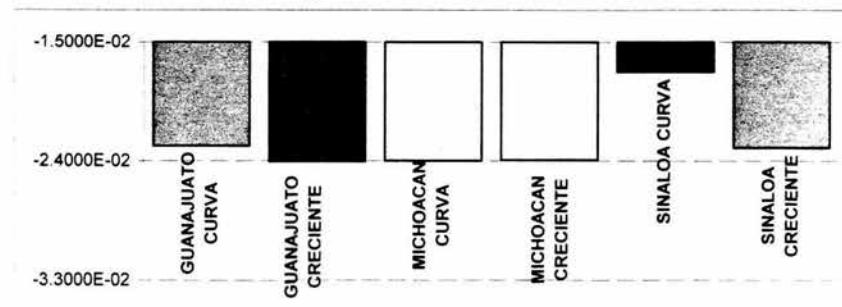
## DELTA, RESULTADOS A NIVEL ENTIDAD FEDERATIVA

A continuación se presentan los resultados de la medida de homogeneidad delta, para los diferentes tamaños de muestra a nivel entidad federativa.

( $n = 40$ )

		SSB	SSW	TOTAL	K	DELTA
GUANAJUATO	<b>CURVA</b>	1,212,731.68	510,364,750.03	511,577,481.70	74	-0.022864
	<b>CRECIENTE</b>	491,608.37	409,984,615.18	410,476,223.56	74	-0.024067
MICHUACAN	<b>CURVA</b>	611,623.14	500,958,485.65	501,570,108.79	66	-0.024002
	<b>CRECIENTE</b>	549,077.29	431,617,705.03	432,166,782.31	66	-0.023950
SINALOA	<b>CURVA</b>	1,682,954.27	211,527,897.72	213,210,851.98	94	-0.017275
	<b>CRECIENTE</b>	323,978.90	140,094,943.35	140,418,922.25	94	-0.023002

Tabla 3.2.18. Medida de homogeneidad delta para los tres estados con una  $n = 40$



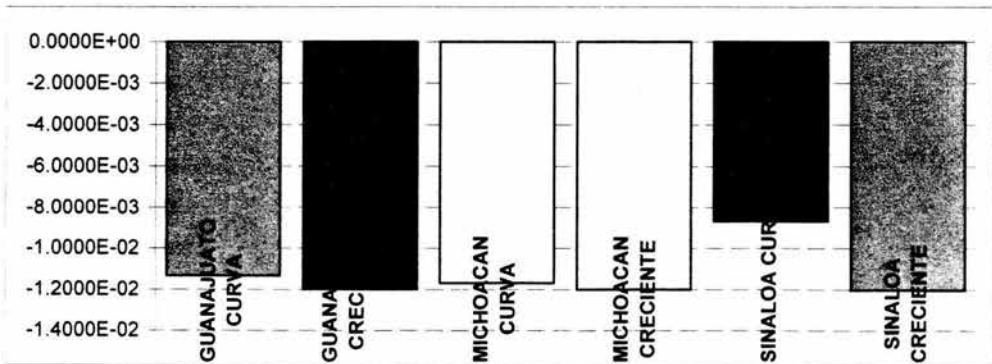
Gráfica 3.2.19. Medida de homogeneidad delta para los tres estados con una  $n = 40$



$(n = 80)$ 

		SSB	SSW	TOTAL	K	DELTA
GUANAJUATO	<b>CURVA</b>	452,564.30	476,831,173.65	477,283,737.95	36	-0.01134673
	<b>CRECIENTE</b>	96,212.93	319,231,865.66	319,328,078.60	36	-0.01200160
MICHOCACAN	<b>CURVA</b>	255,864.46	463,199,693.44	463,455,557.90	32	-0.01170381
	<b>CRECIENTE</b>	90,925.67	356,570,437.85	356,661,363.52	32	-0.01200460
SINALOA	<b>CURVA</b>	751,049.55	205,513,678.64	206,264,728.19	46	-0.00869677
	<b>CRECIENTE</b>	33,690.21	102,944,695.75	102,978,385.96	46	-0.01205184

Tabla 3.2.19. Medida de homogeneidad delta para los tres estados con una  $n = 80$

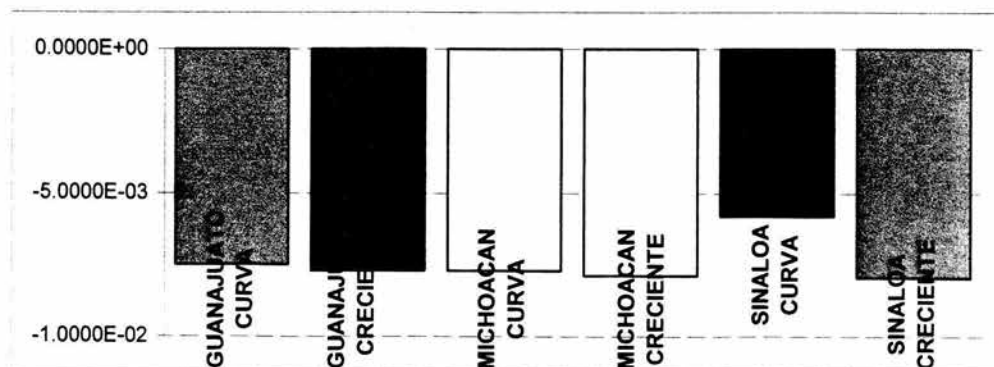


Gráfica 3.2.20. Medida de homogeneidad delta para los tres estados con una  $n = 80$

( $n = 120$ )

		SSB	SSW	TOTAL	K	DELTA
GUANAJUATO	<b>CURVA</b>	259,808.47	477,023,929.48	477,283,737.95	24	-0.007504489
	<b>CRECIENTE</b>	111,122.64	332,469,856.90	332,580,979.55	24	-0.007716408
MICHOACAN	<b>CURVA</b>	151,701.90	501,418,406.88	501,570,108.79	22	-0.007716510
	<b>CRECIENTE</b>	63,283.72	432,103,498.59	432,166,782.31	22	-0.007873782
SINALOA	<b>CURVA</b>	477,764.89	208,984,513.88	209,462,278.77	31	-0.005832825
	<b>CRECIENTE</b>	20,260.73	115,973,254.74	115,993,515.47	31	-0.007956193

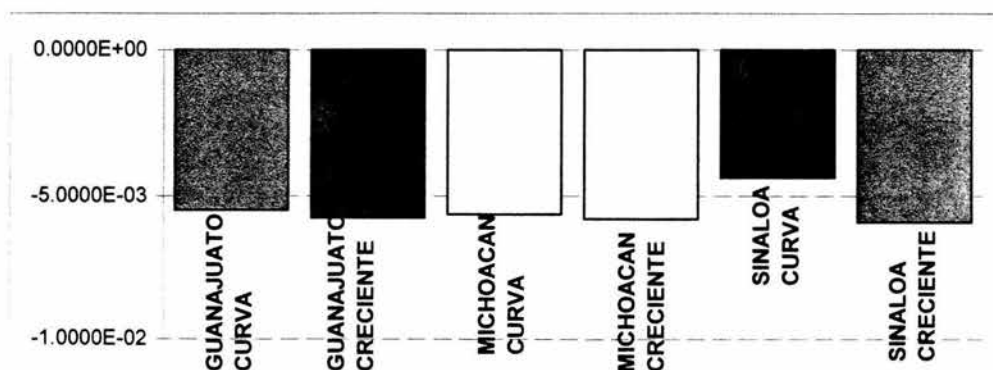
Tabla 3.2.20. Medida de homogeneidad delta para los tres estados con una  $n = 120$



Gráfica 3.2.21. Medida de homogeneidad delta para los tres estados con una  $n = 120$

$(n = 160)$ 

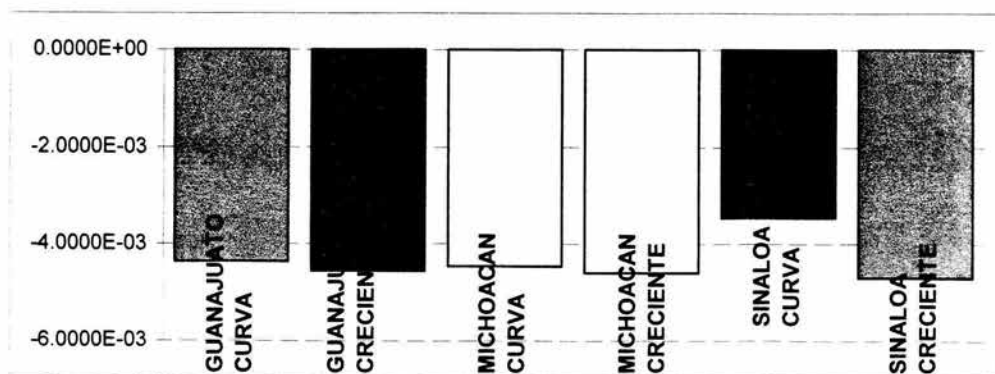
		SSB	SSW	TOTAL	K	DELTA
GUANAJUATO	<b>CURVA</b>	203,743.38	477,079,994.58	477,283,737.95	18	-0.005510486
	<b>CRECIENTE</b>	49,089.24	332,531,890.31	332,580,979.55	18	-0.005791425
MICHOACAN	<b>CURVA</b>	108,190.66	463,347,367.24	463,455,557.90	16	-0.005661407
	<b>CRECIENTE</b>	22,729.47	356,638,634.05	356,661,363.52	16	-0.005832122
SINALOA	<b>CURVA</b>	337,527.57	205,927,200.62	206,264,728.19	23	-0.004369635
	<b>CRECIENTE</b>	8,415.30	102,969,970.66	102,978,385.96	23	-0.005933649

Tabla 3.2.21. Medida de homogeneidad delta para los tres estados con una  $n = 160$ Gráfica 3.2.22. Medida de homogeneidad delta para los tres estados con una  $n = 160$

( $n = 200$ )

		<b>SSB</b>	<b>SSW</b>	<b>TOTAL</b>	<b>K</b>	<b>DELTA</b>
<b>GUANAJUATO</b>	<b>CURVA</b>	139,373.75	455,633,898.65	455,773,272.40	14	-0.004358965
	<b>CRECIENTE</b>	32,061.80	298,783,080.98	298,815,142.78	14	-0.004558391
<b>MICHOACAN</b>	<b>CURVA</b>	84,156.37	480,884,260.51	480,968,416.88	13	-0.004462793
	<b>CRECIENTE</b>	15,922.66	388,222,406.96	388,238,329.62	13	-0.004597375
<b>SINALOA</b>	<b>CURVA</b>	257,842.66	200,662,623.12	200,920,465.78	18	-0.003456554
	<b>CRECIENTE</b>	3,986.62	86,024,701.66	86,028,688.28	18	-0.004699391

**Tabla 3.2.22. Medida de homogeneidad delta para los tres estados con una  $n = 200$**



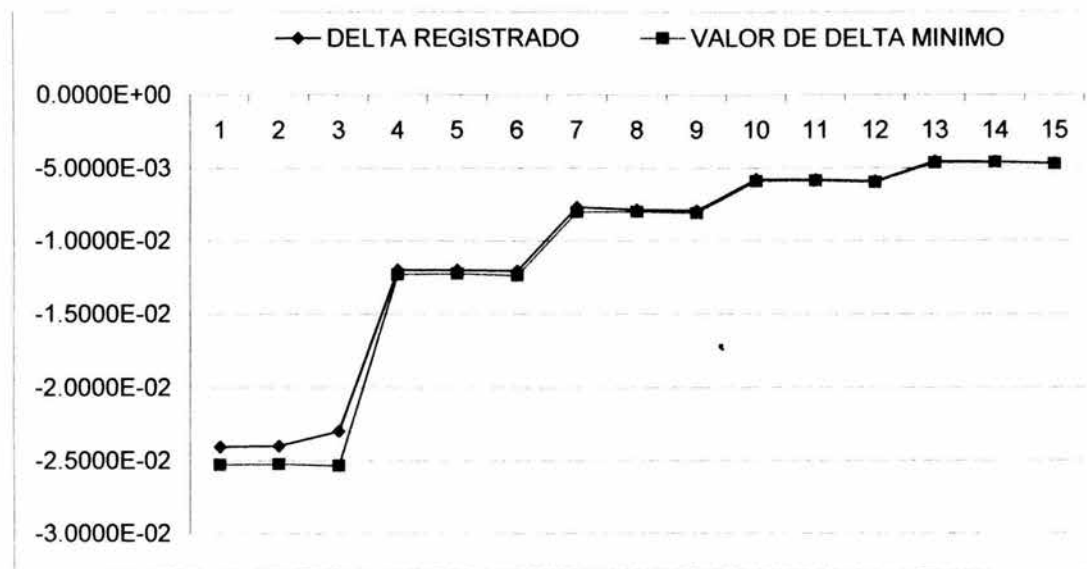
**Gráfica 3.2.23. Medida de homogeneidad delta para los tres estados con una  $n = 200$**

En la siguiente tabla se muestran los tipos de orden de cada entidad federativa donde  $\delta$  fue menor. Donde solo un caso de quince la delta fue menor en un orden curva. También se calculó el valor de  $\delta$  mínimo mediante la expresión  $\delta_{\min} = -(k-1)/(N-k)$ .

Tamaño de muestra	ORDEN CON EL $\delta$ MENOR	$\delta$ REGISTRADO	$\delta$ MÍNIMO	DIFERENCIA
n=40	GUANAJUATO CRECIENTE	-0.0240666	-0.0252945	0.0012279
	<b>MICHOACÁN CURVA</b>	-0.0240023	-0.0252525	0.0012502
	SINALOA CRECIENTE	-0.0230025	-0.0253682	0.0023658
n=80	GUANAJUATO CRECIENTE	-0.0120016	-0.0123066	0.0003050
	MICHOACÁN CRECIENTE	-0.0120046	-0.0122627	0.0002581
	SINALOA CRECIENTE	-0.0120518	-0.0123830	0.0003312
n=120	GUANAJUATO CRECIENTE	-0.0077164	-0.0080532	0.0003368
	MICHOACÁN CRECIENTE	-0.0078738	-0.0080214	0.0001476
	SINALOA CRECIENTE	-0.0079562	-0.0081323	0.0001761
n=160	GUANAJUATO CRECIENTE	-0.0057914	-0.0059399	0.0001485
	MICHOACÁN CRECIENTE	-0.0058321	-0.0058962	0.0000641
	SINALOA CRECIENTE	-0.0059336	-0.0060159	0.0000822
n=200	GUANAJUATO CRECIENTE	-0.0045584	-0.0046662	0.0001078
	MICHOACÁN CRECIENTE	-0.0045974	-0.0046386	0.0000412
	SINALOA CRECIENTE	-0.0046994	-0.0047460	0.0000466

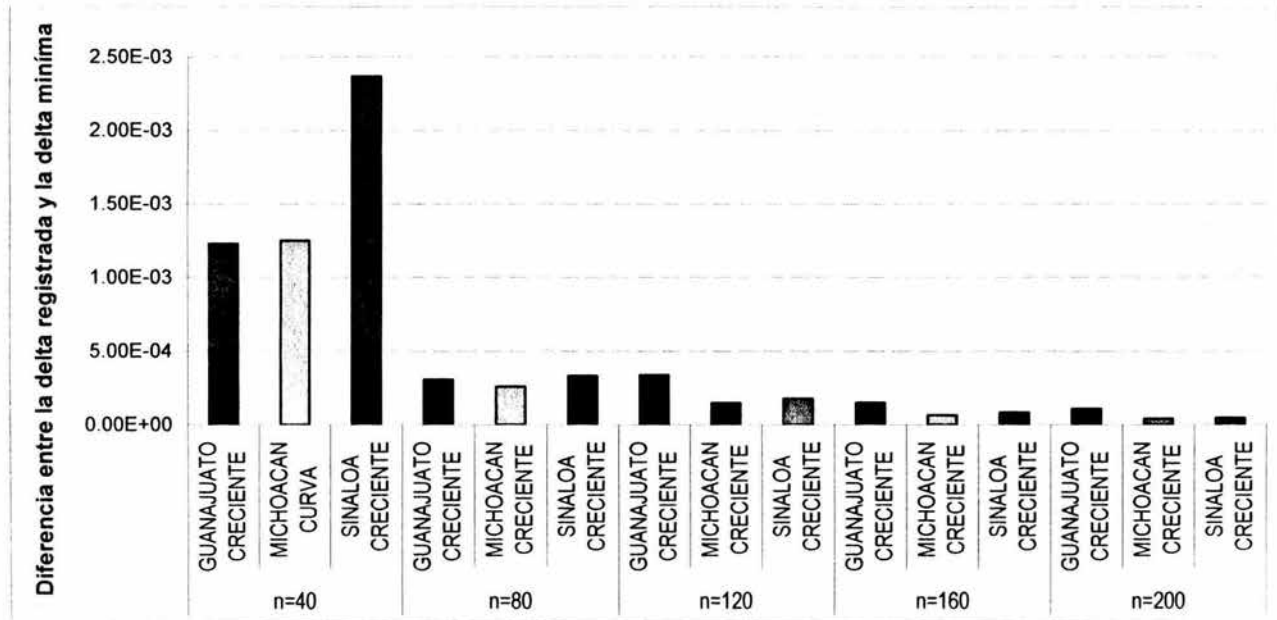
Tabla 3.2.23. Concentrado de los resultados de delta por entidad federativa y tamaño de muestra.

Como se puede observar en la mayoría de los casos, los valores de  $\delta$  registrados están muy próximos a su mínimo. Debido a que  $SSW \approx SST$ . Lo que habla de una alta heterogeneidad dentro de las muestras.



Gráfica 3.2.24. Valores de delta registrado contra sus valores mínimos posibles.

Se obtuvo la diferencia entre los valores  $\delta$  registrados y los valores  $\delta$  mínimos posibles, y como se puede observar, al ir incrementándose el tamaño de muestra, esta diferencia tiende a cero.



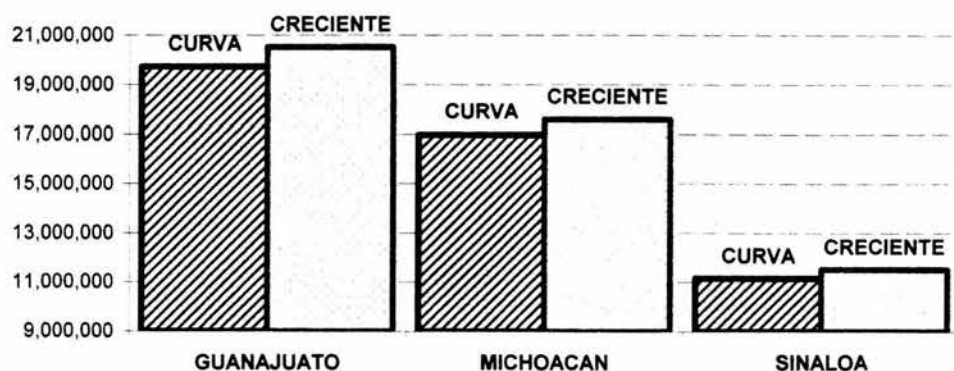
**Gráfica 3.2.25. Diferencia entre los valores de delta registrado contra sus valores mínimos posibles**

## **RESULTADOS A NIVEL ENTIDAD FEDERATIVA**

### **ENTROPÍA RELATIVA**

	<b>CURVA</b>	<b>CRECIENTE</b>
<b>GUANAJUATO</b>	20,500,807.67	19,715,782.87
<b>MICHOACAN</b>	17,576,522.69	16,977,070.83
<b>SINALOA</b>	11,519,532.88	11,147,248.12

*Tabla 3.2.23. Resultados de entropía relativa a nivel entidad federativa por los tipos de orden curva y creciente.*



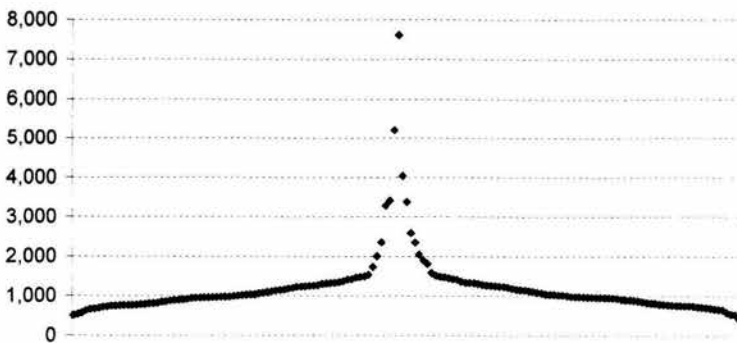
*Gráfica 3.2.24. Resultados de entropía relativa a nivel entidad federativa por los tipos de orden curva y creciente.*



### 3.2.1. ANÁLISIS DE AGUASCALIENTES A NIVEL DISTRITO (DISTRITO DOS)

Para este análisis se decidió tomar una muestra "pequeña" porque podría ser que la relación de Lista Nominal y el número de votos a favor de un partido sea más fuerte a nivel distrito que a nivel estado y así poder corroborar la hipótesis de que entre más parecidos sean los datos entre si, se favorecerá al tipo de orden curva de las unidades poblacionales  $\mu$ . Igual que los otros ejemplos se calculó el error cuadrático medio, en un muestreo sistemático, bajo un orden tipo curva y tipo creciente, así como la entropía relativa y la medida de homogeneidad  $\delta$ .

El orden a utilizar para este muestreo sistemático, es de curva, como puede observarse en la siguiente gráfica,



*Gráfica 3.2.1.1. Lista Nominal bajo el orden de curva del Distrito 2 del estado de Aguascalientes.*

Para el caso del PRI, se tienen los siguientes datos:

- $k = 2$
- $Y = 41,255$
- $X = 139,288$
- $R = \frac{Y}{X} = 0.2961$

Los valores de  $Y_i$  y de  $X_i$  son:

$$Y_1 = 21,132 \quad X_1 = 69,883$$

$$Y_2 = 20,123 \quad X_2 = 69,405$$

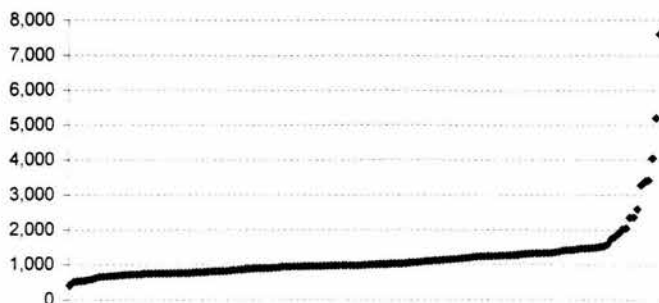
El cálculo del error cuadrático medio, se obtiene de la siguiente manera:

$$EMC(\hat{R}) = \frac{k}{X^2} \sum_{i=1}^k (y_i - x_i R)^2$$

Por lo tanto para este caso tenemos:

$$ECM(\hat{R}) = 3.8782E-05$$

Ahora, se hará el mismo ejercicio pero ahora con un orden creciente:



*Gráfica 3.2.1.2. Lista Nominal bajo un orden creciente del Distrito 2 del estado de Aguascalientes.*

Los valores de  $Y_i$  y de  $X_i$  son:

$$\begin{array}{ll} Y_1 = 18,882 & X_1 = 63,208 \\ Y_2 = 17,532 & X_2 = 62,856 \end{array}$$

En este caso de orden creciente tenemos:

$$ECM(\hat{R}) = 9.8056E-05$$

Concluimos que para el error cuadrático medio en un muestreo sistemático bajo un orden de curva es más eficiente en un  $5.9273E-05$  que aquél que ha sido ordenado de manera creciente, lo que significa una eficiencia del 2.5 mayor.

También se obtuvo el error cuadrático medio del aleatorio simple (Cochran (1977) expresión 2.39, p.31).

$$ECM(\hat{R}_{SRS}) = \frac{N^2}{X^2} \left(1 - \frac{n}{N}\right) \frac{1}{n} \sum_{i=1}^N \frac{(y_i - Rx_i)^2}{N-1}$$

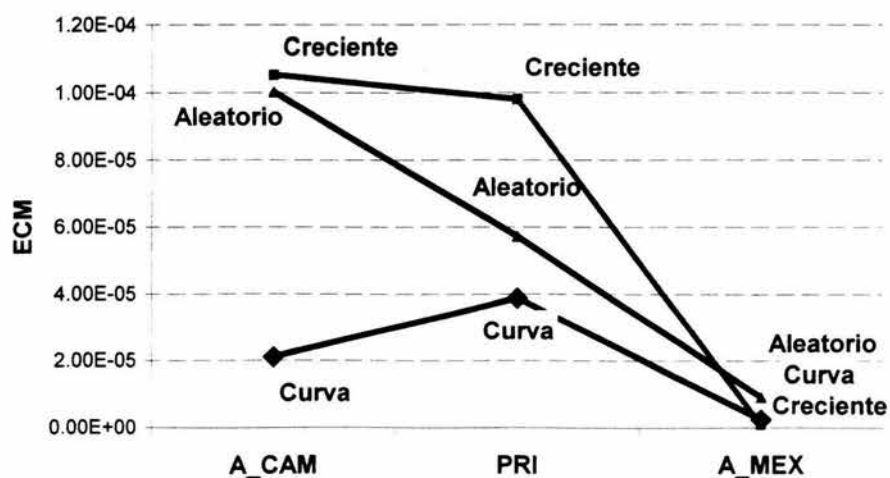
Los resultados fueron:

$$ECM(\hat{R}) = 5.7257E-05$$

Donde el error cuadrático medio fue menor al creciente pero mayor al del orden curva.

		ORDEN		
		Curva	Creciente	Aleatorio
PARTIDO	A_CAM	2.0957E-05	1.0540E-04	1.0022E-04
	PRI	3.8782E-05	9.8056E-05	5.7258E-05
	A_MEX	2.4294E-06	<b>9.3215E-07</b>	9.1315E-06

Tabla 3.2.1.1. Errores Cuadráticos Medios de los partidos A\_CAM, PRI y A\_MEX bajo los tipos de orden curva, creciente y aleatorio. Los ECM menores son señalados con letra **negrita**



Gráfica 3.2.1.3. Error Cuadrático Medio bajo los tres tipos de orden, aleatorio, creciente y curva.

Como se puede observar, para el caso del Distrito 2 del estado de Aguascalientes (con una  $n = 80$ ), en dos de los tres partidos fue menor el ECM para el orden tipo curva, sólo para el partido A\_MEX, fue mejor el orden creciente.

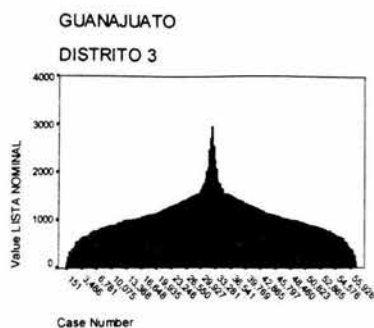
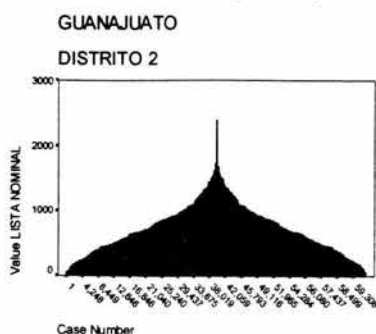
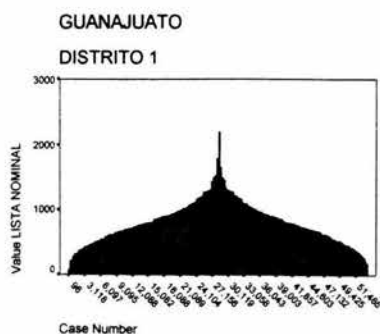
Aunque como se observa en la gráfica este último partido, tuvo un ECM menor a los otros dos y la diferencia entre los tres tipos de orden, es muy pequeña. En los siguientes ejemplos a nivel distrito (27), 18 fueron más eficientes con un orden curva.

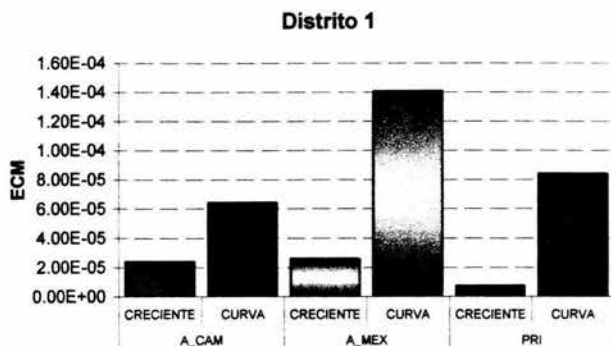
**GUANAJUATO** (Entidad Federativa 11)

En las elecciones ganó el **A\_CAM**

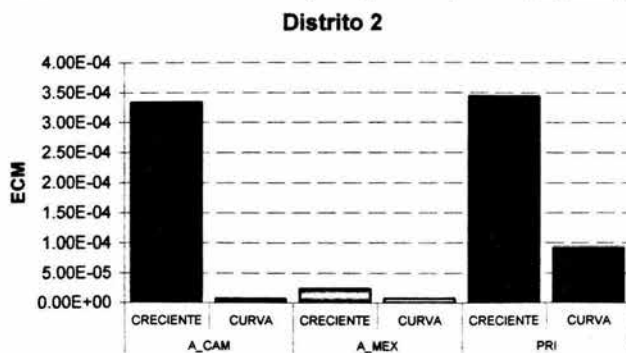
DISTRITO	PARTIDO	ORDEN		
		CURVA	CRECIENTE	ALEATORIO
1	A_CAM	6.404E-05	<b>2.384E-05</b>	8.657E-05
	A_MEX	1.406E-04	<b>2.593E-05</b>	1.892E-04
	PRI	8.386E-05	<b>7.358E-06</b>	6.248E-04
2	A_CAM	<b>6.990E-06</b>	3.338E-04	5.710E-04
	A_MEX	<b>7.243E-06</b>	2.367E-05	4.225E-05
	PRI	<b>9.181E-05</b>	3.436E-04	5.009E-04
3	A_CAM	1.733E-04	<b>1.607E-04</b>	6.384E-04
	A_MEX	3.444E-08	<b>2.655E-08</b>	2.301E-05
	PRI	2.047E-04	<b>1.236E-04</b>	4.668E-04

Tabla 3.2.1.2. Errores Cuadráticos Medios de los partidos de los primeros tres distritos de Guanajuato. Los ECM menores son señalados con letra negra ( $n = 80$ )

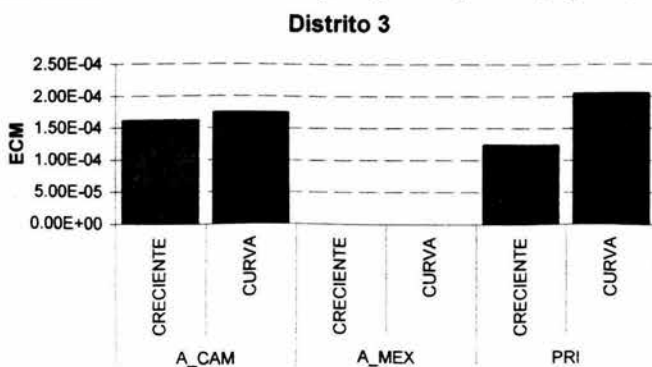




*Gráfica 3.2.1.4. ECM del Distrito 1 de Guanajuato para los partidos, A\_CAM, A\_MEX y PRI.*



*Gráfica 3.2.1.5. ECM del Distrito 2 de Guanajuato para los partidos, A\_CAM, A\_MEX y PRI.*

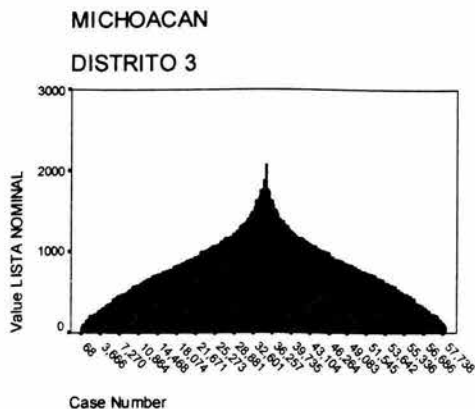
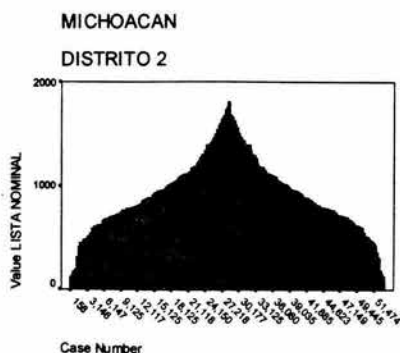
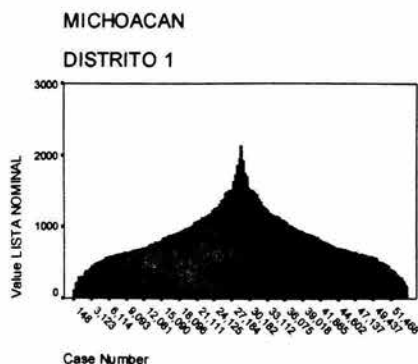


*Gráfica 3.2.1.6. ECM del Distrito 3 de Guanajuato para los partidos, A\_CAM, A\_MEX y PRI.*

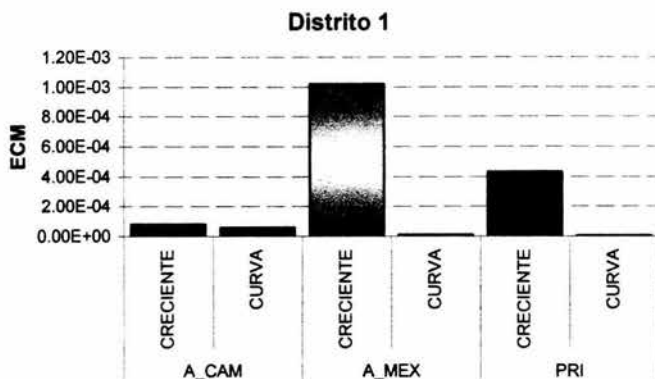
MICHOACÁN (Entidad Federativa 16)

DISTRITO	PARTIDO	CURVA	ORDEN	
			CRECIENTE	ALEATORIO
1	A CAM	<b>5.9130E-05</b>	7.9337E-05	1.483E-03
	A MEX	<b>1.0835E-05</b>	1.0183E-03	3.826E-04
	PRI	<b>2.7015E-06</b>	4.3138E-04	1.309E-03
2	A CAM	<b>1.0424E-04</b>	2.5029E-04	4.124E-04
	A MEX	<b>4.0200E-05</b>	1.8525E-04	2.337E-04
	PRI	1.5405E-05	<b>1.5704E-06</b>	1.704E-04
3	A CAM	<b>1.3530E-04</b>	7.3495E-04	4.913E-04
	A MEX	<b>6.2429E-05</b>	8.8297E-04	1.804E-04
	PRI	<b>3.4183E-07</b>	1.8916E-06	3.913E-04

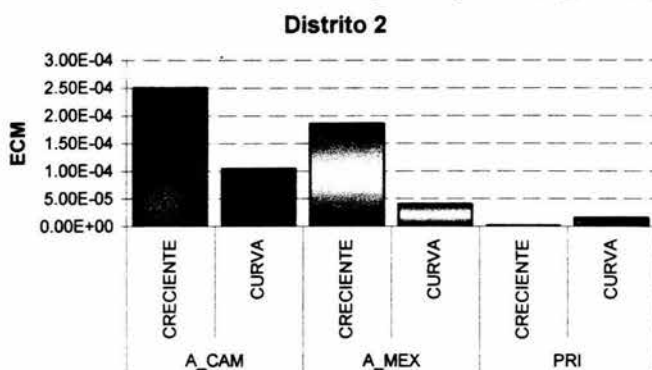
Tabla 3.2.1.3. Errores Cuadráticos Medios de los partidos de los primeros tres distritos de Michoacán. Los ECM menores son señalados con letra negra ( $n = 80$ )



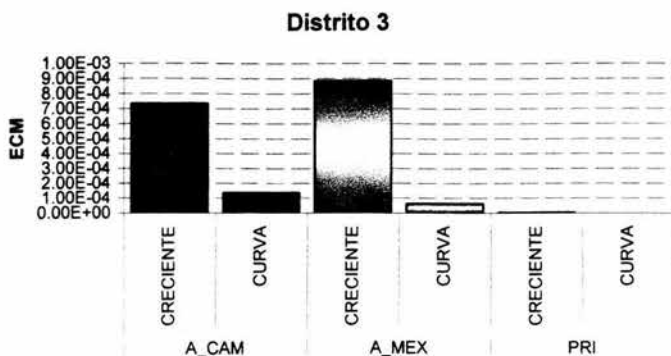




*Gráfica 3.2.1.7. ECM del Distrito 1 de Michoacán para los partidos, A\_CAM, A\_MEX y PRI.*



*Gráfica 3.2.1.8. ECM del Distrito 2 de Michoacán para los partidos, A\_CAM, A\_MEX y PRI.*

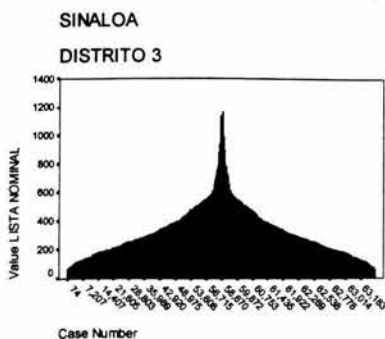
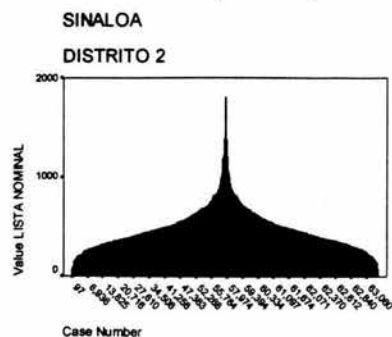
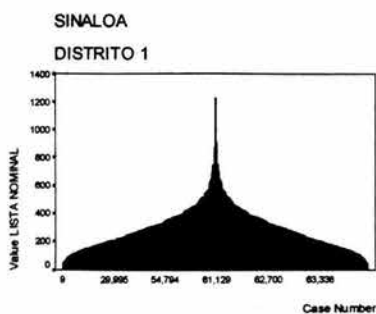


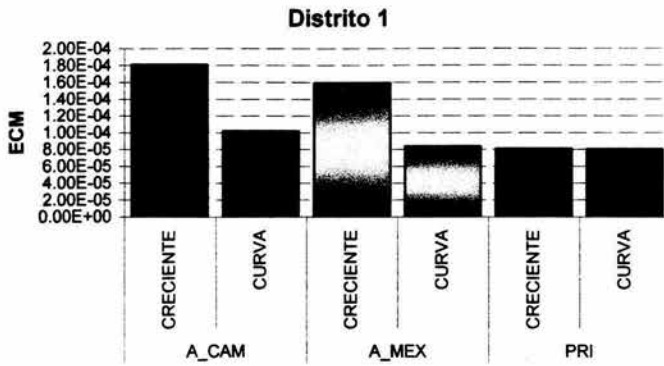
*Gráfica 3.2.1.9. ECM del Distrito 3 de Michoacán para los partidos, A\_CAM, A\_MEX y PRI.*

SINALOA (Entidad Federativa 25)

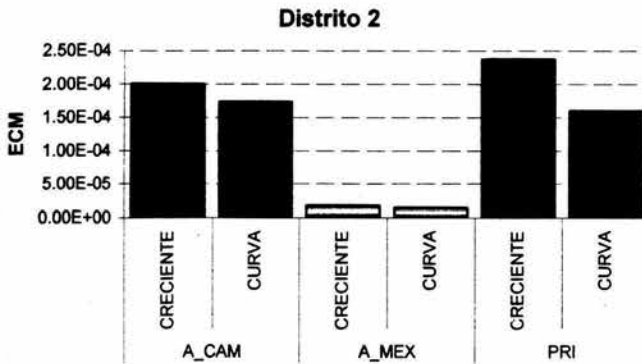
DISTRITO	PARTIDO	CURVA	Orden	
			CRECIENTE	ALEATORIO
1	A_CAM	<b>1.0161E-04</b>	1.8091E-04	1.608E-03
	A_MEX	<b>8.3275E-05</b>	1.5905E-04	2.647E-04
	PRI	<b>7.9873E-05</b>	8.0646E-05	6.334E-04
2	A_CAM	<b>1.7327E-04</b>	2.0013E-04	7.139E-05
	A_MEX	<b>1.4244E-05</b>	1.7518E-05	2.900E-05
	PRI	<b>1.5962E-04</b>	2.3728E-04	1.162E-04
3	A_CAM	2.8902E-04	<b>2.6940E-04</b>	2.724E-04
	A_MEX	<b>8.0336E-05</b>	1.2514E-04	5.375E-05
	PRI	2.0610E-04	<b>1.1585E-04</b>	2.924E-04

Tabla 3.2.1.4. Errores Cuadráticos Medios de los partidos de los primeros tres distritos de Sinaloa. Los ECM menores son señalados con letra negra ( $n = 80$ )

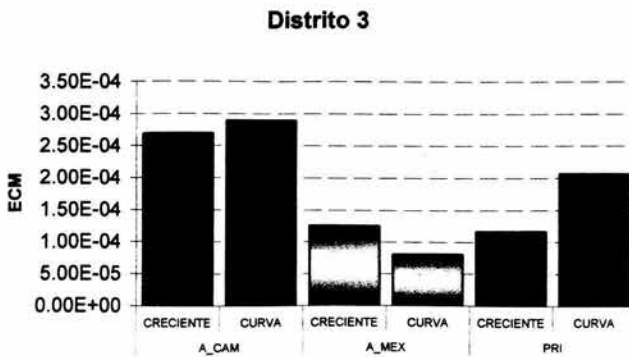




**Gráfica 3.2.1.10. ECM del Distrito 1 de Sinaloa para los partidos, A\_CAM, A\_MEX y PRI.**



**Gráfica 3.2.1.11. ECM del Distrito 2 de Sinaloa para los partidos, A\_CAM, A\_MEX y PRI.**



**Gráfica 3.2.1.12. ECM del Distrito 3 de Sinaloa para los partidos, A\_CAM, A\_MEX y PRI.**

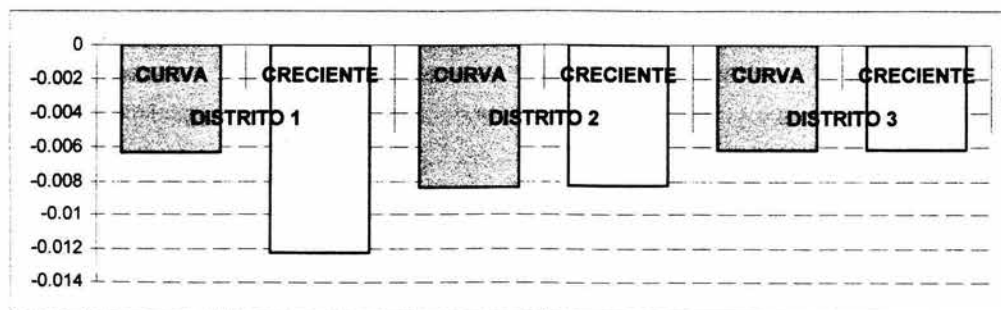
## RESULTADOS A NIVEL DISTRITO

### DELTA

#### GUANAJUATO ( $n = 80$ )<sup>1</sup>

DISTRITO	ORDEN	SSB	SSW	TOTAL	K	DELTA
1	CURVA	49.51	15,614,436.94	15,614,486.44	2	-0.00632592
	CRECIENTE	8,910.22	17,951,520.18	17,960,430.40	2	-0.01223643
2	CURVA	2,740.76	32,107,784.64	32,110,525.40	3	-0.00835274
	CRECIENTE	4,056.01	22,343,571.33	22,347,627.33	3	-0.00825579
3	CURVA	4,526.26	30,184,787.59	30,189,313.84	2	-0.00617824
	CRECIENTE	2,080.81	10,893,389.14	10,895,469.94	2	-0.00613693

Tabla 3.2.1.5. Delta de los primeros tres distritos de Guanajuato en orden tipo creciente y curva para una  $n=80$ .



Gráfica 3.2.1.13. Delta de los primeros tres distritos de Guanajuato en orden tipo creciente y curva para una  $n=80$ .

<sup>1</sup> SSW: Suma de cuadrados dentro de las muestras.

SSB: Suma de cuadrados entre las muestras.

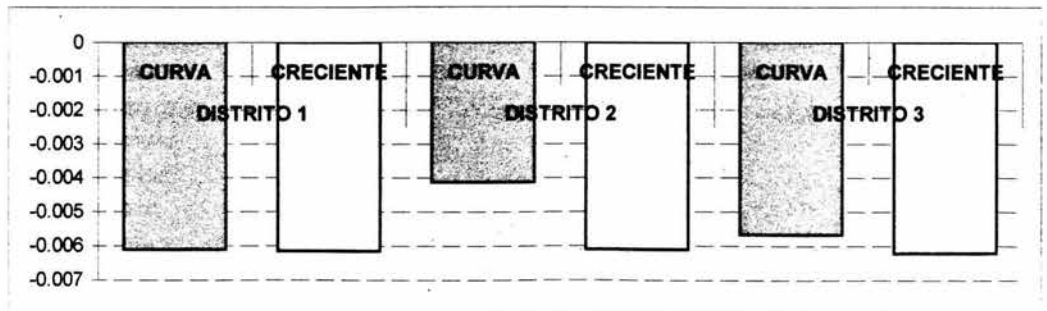
SST: Suma de cuadrados total.

$$SST = SSW + SSB$$

**MICHOACÁN  $n = 80$** 

DISTRITO	ORDEN	SSB	SSW	TOTAL	K	DELTA
1	<b>CURVA</b>	4,782.37	21,692,140.33	21,696,922.69	2	-0.00610730
	<b>CRECIENTE</b>	1,716.10	10,376,607.88	10,378,323.98	2	-0.00616271
2	<b>CURVA</b>	39,719.18	18,091,833.16	18,131,552.34	3	-0.00412464
	<b>CRECIENTE</b>	2,528.10	11,520,919.50	11,523,447.60	3	-0.00610834
3	<b>CURVA</b>	18,312.70	28,248,778.90	28,267,091.60	2	-0.00567717
	<b>CRECIENTE</b>	1,322.50	10,901,073.00	10,902,395.50	2	-0.00620704

**Tabla 3.2.1.6. Delta de los primeros tres distritos de Michoacán en orden tipo creciente y curva para una  $n=80$ .**

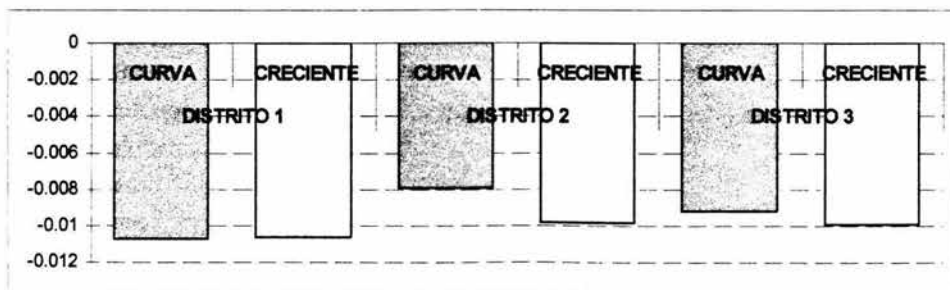


**Gráfica 3.2.1.14. Delta de los primeros tres distritos de Michoacán en orden tipo creciente y curva para una  $n=80$ .**

**SINALOA  $n = 80$**

DISTRITO	ORDEN	SSB	SSW	TOTAL	K	DELTA
1	CURVA	979.82	12,794,665.09	12,795,644.91	7	-0.01077250
	CRECIENTE	1,933.20	8,059,921.65	8,061,854.85	7	-0.01060751
2	CURVA	38,405.35	17,743,520.16	17,781,925.51	5	-0.00794491
	CRECIENTE	1,849.39	7,456,115.11	7,457,964.50	5	-0.00987610
3	CURVA	12,337.90	13,130,112.14	13,142,450.04	5	-0.00917829
	CRECIENTE	993.38	4,862,554.58	4,863,547.96	5	-0.00992026

*Tabla 3.2.1.7. Delta de los tres primeros distritos de Sinaloa en orden tipo creciente y curva para una  $n=80$ .*



*Gráfica 3.2.1.15. Delta de los primeros tres distritos de Sinaloa en orden tipo creciente y curva para una  $n=80$ .*

# *Capítulo 4*



### **3.3. MATRÍCULA DE LAS ESCUELAS PRIMARIAS DE GUERRERO Y DISTRITO FEDERAL (1999-2000)**

---

Este es el último ejemplo de este trabajo. Una de las razones para trabajar con los datos de las escuelas primarias de Guerrero y Distrito Federal, fue el que a diferencia del ejemplo de las elecciones del año 2000, que se ordenó con una variable, y se infirió sobre otras, sin que la relación entre éstas fuera muy robusta, en esta ocasión, se consideran dos variables, con una fuerte relación entre ellas.

Esta base de datos contiene 8,452 registros, cada uno de ellos se refiere a una escuela. Cada registro tiene información referente a la entidad federativa que pertenece, el tipo de población (rural y urbana), el tipo de escuela (Federal, particular, internado entre otras), la matrícula de alumnos, el número de docentes, y el número de grupos. Se tienen, 3,445 registros del Distrito Federal, y 5,007 de Guerrero.

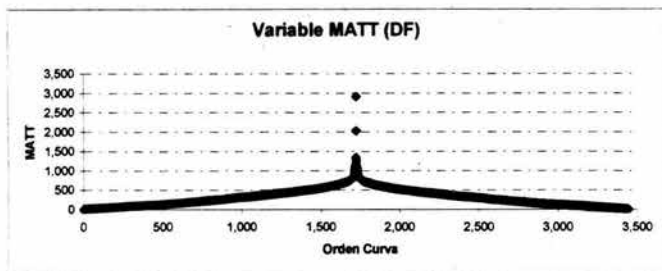
La hipótesis inicial, es que el número de alumnos tiene una relación lineal con el número de profesores (docentes). En este ejemplo, se decidió ordenar por la variable de matrícula  $x$  (número de alumnos) e inferir sobre la variable docentes  $y$  (número de maestros). Se trabajó con la clasificación de escuelas que manejan ambos estados, privadas, particulares, de conafe y otras.

Igual que los otros ejemplos, se trabaja con un muestreo sistemático con tipos de orden creciente y curva, además de un muestreo aleatorio simple (aleatorio). Para ver el comportamiento de la relación y cómo se ubican

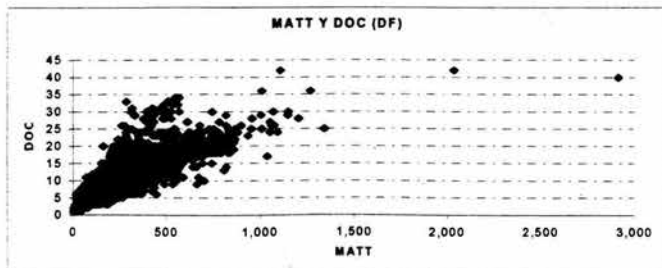


gráficamente los datos al ordenarse en forma aleatoria y creciente, se presentan las siguientes gráficas.

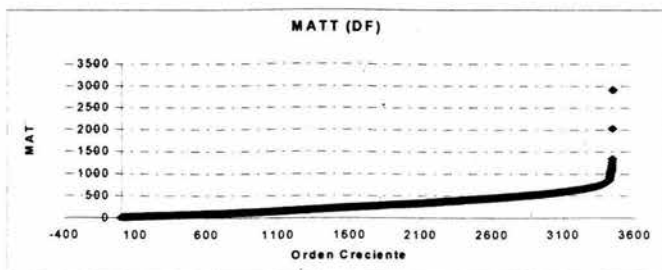
### Gráficas de Distrito Federal



Gráfica 3.3.1. Orden curva de la variable MATT

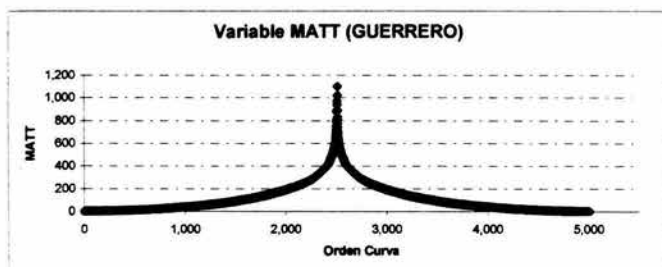


Gráfica 3.3.2. Relación de las variables Matricula y Docentes.  $R^2 = 0.7172$

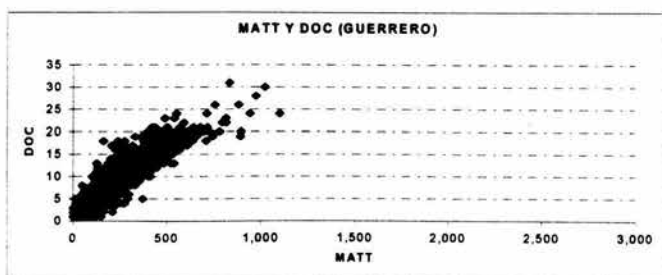


Gráfica 3.3.3. Orden creciente de la variable matricula MATT

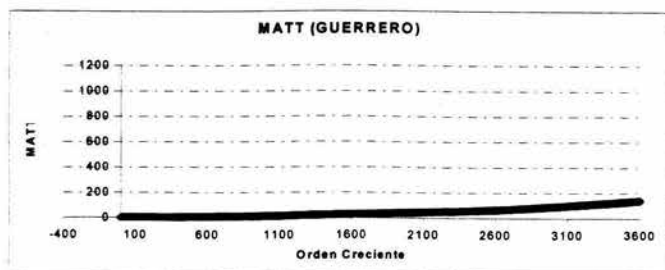
## Gráficas Guerrero



Gráfica 3.3.4. Orden curva de la variable matrícula MATT



Gráfica 3.3.5. Relación entre las variables matrícula MATT y docentes DOC.  $R^2 = .8904$



Gráfica 3.3.6. Orden creciente de la variable matrícula MATT

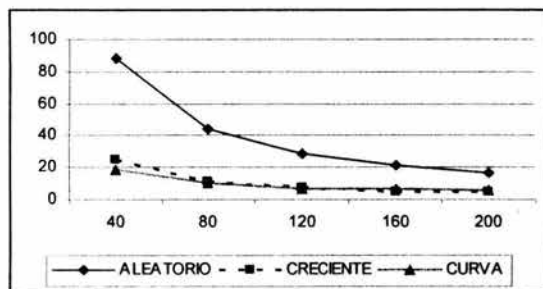
**Resultados de la varianza de la media de la variable Docentes  $V(\bar{y})$ .**

n	DISTRITO FEDERAL		
	CURVA	CRECIENTE	ALEATORIO
40	18.52933	24.43578	88.33247
80	10.08740	11.39254	43.64739
120	6.23398	6.98344	28.75237
160	6.05463	4.28249	21.30485
200	5.52274	4.25124	16.83635

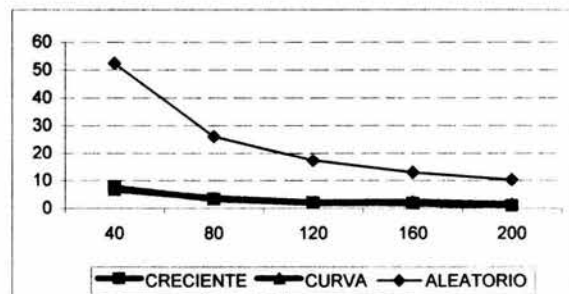
Tabla 3.3.1. Varianzas multiplicadas por 100 de Distrito Federal

n	GUERRERO		
	CURVA	CRECIENTE	ALEATORIO
40	6.65293	7.27226	52.43119
80	3.56782	3.25363	26.00448
120	2.03555	1.89880	17.19557
160	2.28695	1.71179	12.79112
200	1.43623	0.86986	10.14845

Tabla 3.3.2. Varianzas multiplicadas por 100 de Guerrero



Gráfica 3.3.7. Varianzas multiplicadas por 100 de Distrito Federal



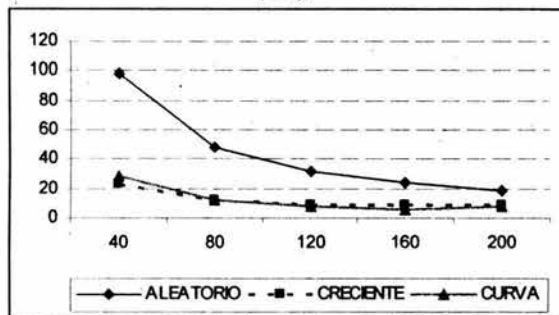
Gráfica 3.3.8. Varianzas multiplicadas por 100 de Guerrero

DISTRITO FEDERAL<sup>1</sup>

## MEDIA DOC PRIMARIA GENERAL FEDERAL

n	CURVA	CRECIENTE	ALEATORIO
40	28.65545	<b>22.88017</b>	97.95976
80	<b>11.75245</b>	12.44619	48.40449
120	<b>8.12453</b>	8.23553	31.88607
160	5.59311	<b>8.54679</b>	23.62685
200	<b>8.00628</b>	8.23830	18.67133

Tabla 3.3.3. Varianzas multiplicadas por 100 de PGF (DF)

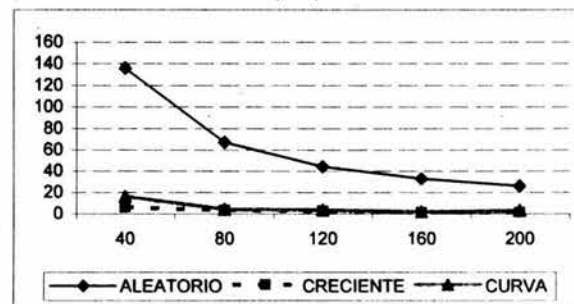


Gráfica 3.3.9. Varianzas multiplicadas por 100 de PGF (DF)

## MEDIA DOC PRIMARIA GENERAL PARTICULAR

	CURVA	CRECIENTE	ALEATORIO
	16.00505	<b>6.19387</b>	135.84810
	4.26557	<b>3.42104</b>	67.12611
	3.54465	<b>2.66623</b>	44.21878
	1.61103	<b>1.22482</b>	32.76512
	2.90061	<b>1.25360</b>	25.89292

Tabla 3.3.4. Varianzas multiplicadas por 100 de PGP (DF)



Gráfica 3.3.10. Varianzas multiplicadas por 100 de PGP (DF)

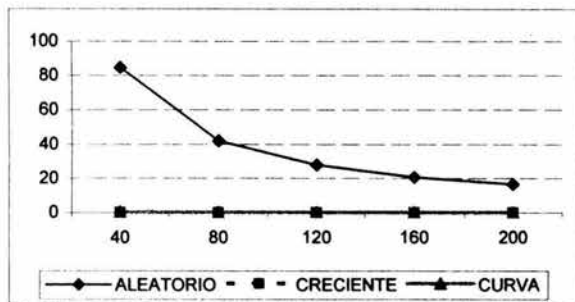
<sup>1</sup> Todos las varianzas calculadas han sido multiplicadas por 100.

GUERRERO<sup>2</sup>

MEDIA DOC PRIMARIA CONAFE FEDERAL

n	CURVA	CRECIENTE	ALEATORIO
40	<b>0.13416</b>	0.15338	84.64223
80	0.06162	<b>0.05070</b>	41.98029
120	0.04320	<b>0.02469</b>	27.75965
160	<b>0.01530</b>	0.02391	20.64933
200	0.01460	<b>0.01340</b>	16.38313

Tabla 3.3.5. Varianzas multiplicadas por 100 de PCF (GRO.)

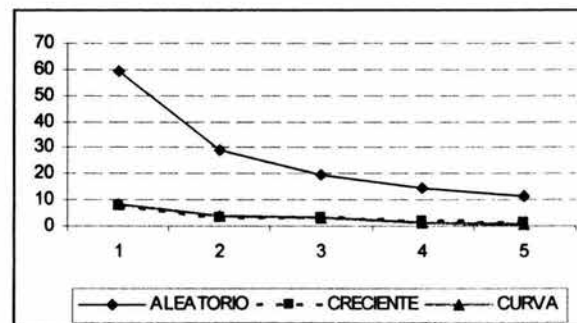


Gráfica 3.3.11. Varianzas multiplicadas por 100 de PCF (GRO.)

MEDIA DOC PRIMARIA GENERAL FED. TRANSF.

	CURVA	CRECIENTE	ALEATORIO
	8.10218	<b>7.53016</b>	59.06373
	3.66417	<b>3.18973</b>	29.29404
	3.16756	<b>2.91028</b>	19.37081
	<b>1.29557</b>	1.68080	14.40919
	<b>0.88957</b>	1.28657	11.43222

Tabla 3.3.6. Varianzas multiplicadas por 100 de PGFT (GRO.)



Gráfica 3.3.12. Varianzas multiplicadas por 100 de PGFT (GRO.)

<sup>2</sup> Todos las varianzas calculadas han sido multiplicadas por 100.

<i>n</i>	Distrito Federal			Guerrero		
	Curva	Creciente	Aleatorio	Curva	Creciente	Aleatorio
40	1	0	0	1	0	0
80	1	0	0	0	1	0
120	1	0	0	0	1	0
160	0	1	0	0	1	0
200	0	1	0	0	1	0
<b>Total</b>	<b>3</b>	<b>2</b>	<b>0</b>	<b>1</b>	<b>4</b>	<b>0</b>
<b>%</b>	<b>60%</b>	<b>40%</b>	<b>0%</b>	<b>20%</b>	<b>80%</b>	<b>0%</b>

Tabla 3.3.7. Concentrado de casos donde el ECM fue menor por estado

<i>n</i>	Distrito Federal			Primaria General Particular		
	Curva	Creciente	Aleatorio	Curva	Creciente	Aleatorio
40	0	1	0	0	1	0
80	1	0	0	0	1	0
120	1	0	0	0	1	0
160	0	1	0	0	1	0
200	0	1	0	0	1	0
<b>Total</b>	<b>2</b>	<b>3</b>	<b>0</b>	<b>0</b>	<b>5</b>	<b>0</b>
<b>%</b>	<b>40%</b>	<b>60%</b>	<b>0%</b>	<b>0%</b>	<b>100%</b>	<b>0%</b>

Tabla 3.3.8. Concentrado de casos donde el ECM fue menor por estado (DF) y tipo de escuela

Guerrero

<i>n</i>	Primaria Conafe Federal			Primaria General Fed. Transf.		
	Curva	Creciente	Aleatorio	Curva	Creciente	Aleatorio
40	1	0	0	0	1	0
80	0	1	0	0	1	0
120	0	1	0	0	1	0
160	1	0	0	1	0	0
200	0	1	0	1	0	0
Total	2	3	0	2	3	0
%	40%	60%	0%	40%	60%	0%

Tabla 3.3.9. Concentrado de casos donde el ECM fue menor por estado (GRO.) y tipo de escuela.

Como se puede observar en estos resultados, el orden tipo curva no tuvo una tendencia a ser el orden más eficiente, sin embargo, sigue siendo junto con el orden tipo creciente, mejor en todos los casos comparados con el aleatorio simple.

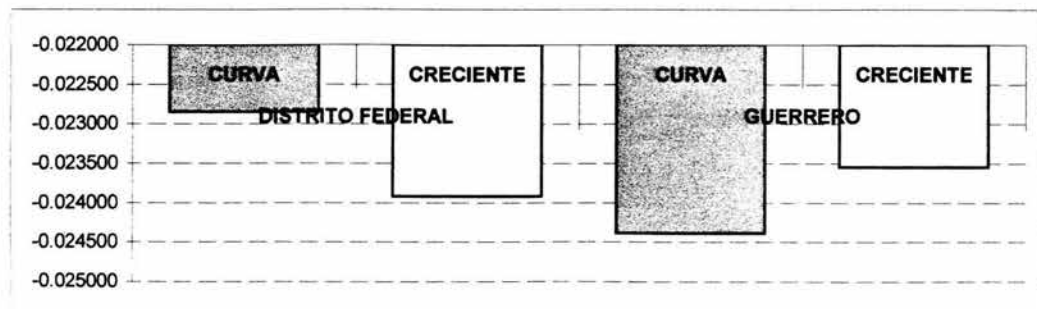
Por lo que podemos concluir que en este ejemplo, no se puede tener una afirmación sobre el tipo de orden más eficiente a utilizar (creciente o curva).

## DELTA, RESULTADOS A NIVEL ENTIDAD FEDERATIVA

( $n = 40$ )

		SSB	SSW	TOTAL	K	DELTA
<b>DISTRITO FEDERAL</b>	<b>CURVA</b>	386,769.04	158,658,880.91	159,045,649.95	86	-0.022849
	<b>CRECIENTE</b>	203,650.58	146,588,697.00	146,792,347.58	86	-0.023920
<b>GUERRERO</b>	<b>CURVA</b>	84,010.24	82,664,664.60	82,748,674.85	125	-0.024395
	<b>CRECIENTE</b>	143,291.23	77,642,369.95	77,785,661.18	125	-0.023547

Tabla 3.3.10. Medida de homogeneidad delta para los dos estados con una  $n = 40$



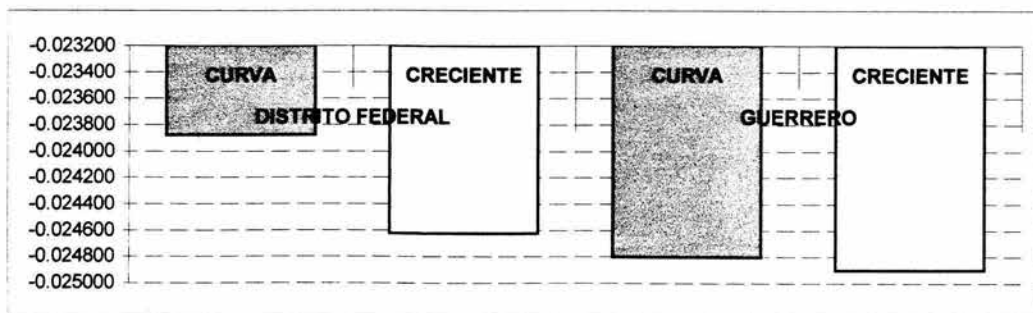
Gráfica 3.3.13. Medida de homogeneidad delta para los dos estados con una  $n = 40$



( $n = 80$ )

		<b>SSB</b>	<b>SSW</b>	<b>TOTAL</b>	<b>K</b>	<b>DELTA</b>
<b>DISTRITO FEDERAL</b>	<b>CURVA</b>	180,818.21	158,864,831.75	159,045,649.95	43	-0.023879
	<b>CRECIENTE</b>	60,083.66	146,732,263.93	146,792,347.58	43	-0.024625
<b>GUERRERO</b>	<b>CURVA</b>	34,055.89	82,239,684.07	82,273,739.97	62	-0.024803
	<b>CRECIENTE</b>	20,840.77	65,606,520.74	65,627,361.51	62	-0.024902

Tabla 3.3.11. Medida de homogeneidad delta para los dos estados con una  $n = 80$

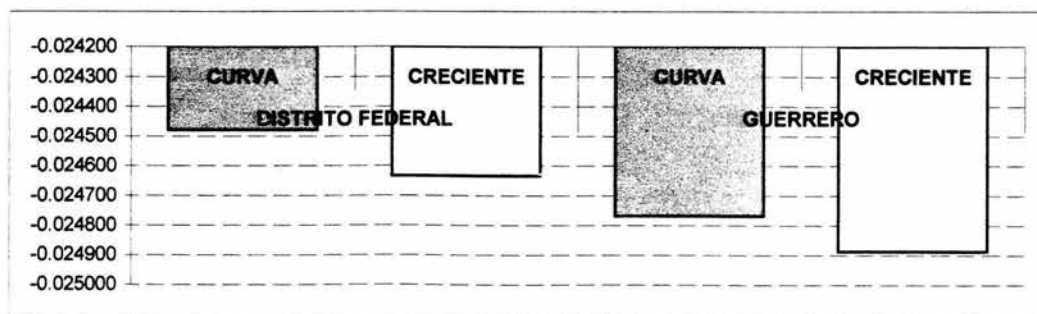


Gráfica 3.3.14. Medida de homogeneidad delta para los dos estados con una  $n = 80$

( $n = 120$ )

		SSB	SSW	TOTAL	K	DELTA
<b>DISTRITO FEDERAL</b>	<b>CURVA</b>	19,650.91	81,779,347.75	81,798,998.66	28	-0.024479
	<b>CRECIENTE</b>	10,707.05	120,705,794.75	120,716,501.80	28	-0.024634
<b>GUERRERO</b>	<b>CURVA</b>	19,650.91	81,779,347.75	81,798,998.66	41	-0.024769
	<b>CRECIENTE</b>	7,257.45	58,159,117.20	58,166,374.65	41	-0.024888

Tabla 3.3.12. Medida de homogeneidad delta para los dos estados con una  $n = 120$

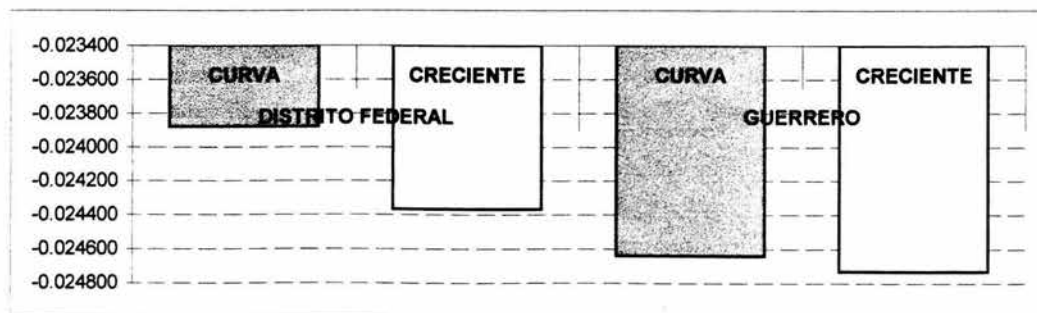


Gráfica 3.3.15. Medida de homogeneidad delta para los dos estados con una  $n = 120$

( $n = 160$ )

		SSB	SSW	TOTAL	K	DELTA
<b>DISTRITO FEDERAL</b>	<b>CURVA</b>	80,692.60	152,825,762.89	152,906,455.49	21	-0.023879
	<b>CRECIENTE</b>	6,237.66	120,710,264.14	120,716,501.80	21	-0.024367
<b>GUERRERO</b>	<b>CURVA</b>	13,965.23	82,259,774.73	82,273,739.97	31	-0.024640
	<b>CRECIENTE</b>	5,225.04	65,622,136.47	65,627,361.51	31	-0.024732

Tabla 3.3.13. Medida de homogeneidad delta para los dos estados con una  $n = 160$

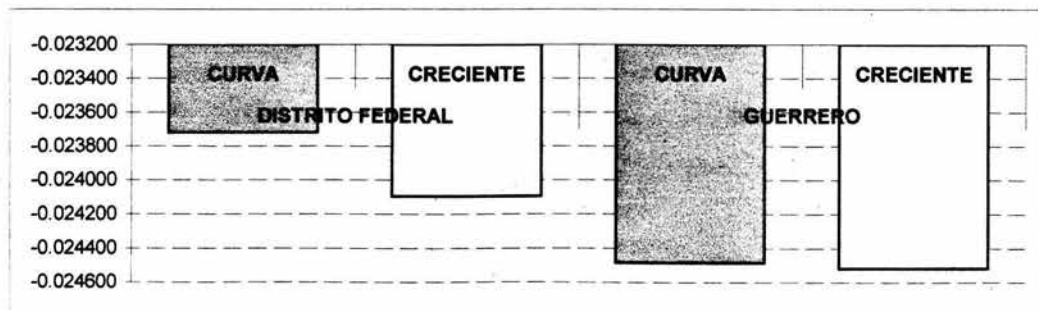


Gráfica 3.3.16. Medida de homogeneidad delta para los dos estados con una  $n = 160$

( $n = 200$ )

		SSB	SSW	TOTAL	K	DELTA
<b>DISTRITO FEDERAL</b>	<b>CURVA</b>	62,588.79	155,802,776.15	155,865,364.94	17	-0.023721
	<b>CRECIENTE</b>	4,672.82	130,486,882.89	130,491,555.71	17	-0.024096
<b>GUERRERO</b>	<b>CURVA</b>	10,575.61	82,738,099.24	82,748,674.85	25	-0.024484
	<b>CRECIENTE</b>	7,406.48	77,778,254.70	77,785,661.18	25	-0.024518

Tabla 3.3.14. Medida de homogeneidad delta para los dos estados con una  $n = 200$



Gráfica 3.3.17. Medida de homogeneidad delta para los dos estados con una  $n = 200$

La entropía para este ejemplo se muestra en la siguiente tabla.

<b>Entidad</b>	<b>Entropía Relativa</b>	<b>Tipo de Orden</b>
Distrito Federal	7,928,302.29	CURVA
Distrito Federal	7,577,663.84	CRECIENTE
Guerrero	4,589,167.29	CURVA
Guerrero	4,336,595.87	CRECIENTE

*Tabla 3.3.15. Concentrado de los resultados de la entropía relativa.*

## 4.1. CONCLUSIONES GENERALES

---

Dentro de las conclusiones obtenidas durante este trabajo sobresalen las siguientes:

### ***Ejemplo de los Números del 1 al 100***

1. No existe un orden único que minimice la varianza
2. Sólo en este ejemplo al menos cuatro distintos tipos de orden tienen varianza 0.
3. En los cuatro ejemplos de varianza cero, delta  $\delta$  fue la misma (-0.1).
4. En estos cuatro ejemplos, la entropía relativa se maximizó en el orden que es similar a la distribución normal.
5. En la medida en que la forma del histograma de los datos, se parezca menos a la normal, de igual manera, se debilita el supuesto de que el orden normal sea más eficiente que el orden creciente. Es decir que el  $ECM_{CURVA}$  sea menor que el  $ECM_{CRECIENTE}$ .



**Ejemplo de las Elecciones presidenciales de México en el 2000 a nivel estatal y distrito electoral**

1. No basta que el orden tipo curva de las unidades poblacionales  $\mu$  sea similar a la normal, sino que además la variable auxiliar  $x$  sobre la cual se ordena debe estar altamente relacionada sobre la variable de interés  $y$  sobre la cual se infiere.
2. A nivel entidad federativa, no existe una tendencia hacia que algún tipo de orden sea más eficiente.
3. A nivel distrito, existen casos, donde se cumple que el orden curva sea muy parecido a la normal, y se obtiene una mejor eficiencia.

**Ejemplo de la Matrícula de las escuelas primarias de Guerrero y Distrito Federal (1999-2000)**

1. A nivel entidad federativa, existe una tendencia a que el orden creciente sea más eficiente que el orden curva. Debido a que aún en orden curva, la forma del histograma de los datos no parece tener una distribución normal, la curva es muy plana.
2. Existen casos donde para un mismo conjunto de datos:
  - 2.1. Se tome una muestra de tamaño  $n$ , y el orden que tiene mayor eficiencia sea la curva.
  - 2.2. Incrementando el tamaño de la  $n$ , el orden que tiene mayor eficiencia es el creciente.

**Conclusiones Generales**

1. Para que el orden tipo curva sea más eficiente se deben de cumplir al menos las siguientes condiciones:
  - 1.1. Que al ordenar por la variable auxiliar  $x$  a las unidades poblacionales  $\mu$  se logre un histograma similar al de una distribución normal.
  - 1.2. La relación entre la variable auxiliar  $x$  y la variable de interés  $y$  sea alta, entre más alta sea la correlación, más se favorecerá al orden tipo curva.
2. Para poder cumplir con el punto 1.1, es indispensable conocer los valores de la variable  $x$  para toda la población y para cumplir con el punto 1.2, es necesario que esa variable  $x$  esté relacionada con la variable o variables de interés lo que en la práctica resulta complejo.
3. Si no se puede cumplir con el punto 1.1 pero sí con el punto 1.2 se recomienda ordenar a las unidades poblacionales  $\mu$  en orden creciente o curva, no en un orden aleatorio, para tener una mayor eficiencia.
4. Entre menor sea el tamaño de muestra mayor será el impacto en la eficiencia del orden elegido, dado que el tamaño de muestra en varias ocasiones es fijo por las variables de tiempo y costo, se sugiere analizar la condición de las variables y su relación, para elegir un orden que coadyuve a una mejor eficiencia.



# *Anexo A*



## ANEXO A

---

Demostración de que la entropía para los datos ordenados en curva es mayor o igual que aquella que sea para datos en orden creciente.

Para hacer la demostración supongamos que  $q(x) = x \quad \forall x$ , por lo tanto:

$$D(q|f) = \sum_i^n x_i \log x_i - \sum_i^n x_i \log f(x_i)$$

Como el primer término de la derecha será igual para un orden creciente como para un orden tipo curva, nos enfocaremos para la demostración al

segundo término  $\sum_i^n x_i \log f(x_i)$ .

La siguiente tabla muestra cómo se asignan los valores  $x_i$  a un orden creciente y a un orden tipo curva para  $n$  impar.

Orden	Valores asignados a un orden tipo curva	Valores asignados a un orden tipo creciente
$q(x_i) = x_i$	$f(x_i)$	$f(x_i)$
$x_1$	$f(x_1) = x_1$	$f(x_1) = x_1$
$x_2$	$f(x_2) = x_3$	$f(x_2) = x_2$

Orden	Valores asignados a un orden tipo	Valores asignados a un orden tipo
$q(x_i) = x_i$	curva $f(x_i)$	creciente $f(x_i)$
...	...	...
$x_{n+1/2}$	$f(x_{n+1/2}) = x_n$	$f(x_{n+1/2}) = x_{n+1/2}$
...	...	...
$x_{(n-1)}$	$f(x_{(n-1)}) = x_{(4)}$	$f(x_{(n-1)}) = x_{(n-1)}$
$x_n$	$f(x_n) = x_2$	$f(x_n) = x_n$

Supuestos de la demostración:

1.  $q(x) = x \quad \forall x$
2.  $x_i = i, 1 \leq i \leq n$
3.  $3 \leq n < \infty, n \in \mathbb{Z}$

Demostración:

$$\begin{aligned}
 \sum_{i=1}^n x_i f(x_i) & \underset{\text{Orden\_Curva}}{=} \sum_{i=1}^{\frac{n+1}{2}} x_i \log(2i-1) + \sum_{i=\frac{n+3}{2}}^n x_i \log(2(n-i+1)) \\
 & = \sum_{i=1}^{\frac{n+1}{2}} x_i \log\left(\frac{i(2i-1)}{i}\right) + \sum_{i=\frac{n+3}{2}}^n x_i \log\left(i \frac{2(n-i+1)}{i}\right) \\
 & = \sum_{i=1}^{\frac{n+1}{2}} x_i \left[ \log(i) + \log\left(\frac{2i-1}{i}\right) \right] + \sum_{i=\frac{n+3}{2}}^n x_i \left[ \log(i) + \log\left(\frac{2(n-i+1)}{i}\right) \right]
 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=1}^n x_i \log(i) + \sum_{i=1}^{\frac{n+1}{2}} x_i \log\left(\frac{2i-1}{i}\right) + \sum_{i=\frac{n+3}{2}}^n x_i \log\left(\frac{2(n-i+1)}{i}\right) \\
 &= \sum_{i=1}^n x_i \log(i) + A + B
 \end{aligned}$$

Donde

$$\begin{aligned}
 A &= \sum_{i=1}^{\frac{n+1}{2}} x_i \log\left(\frac{2i-1}{i}\right) \\
 B &= \sum_{i=\frac{n+3}{2}}^n x_i \log\left(\frac{2(n-i+1)}{i}\right)
 \end{aligned}$$

Desarrollando cada una de las expresiones, se tiene:

$$\begin{aligned}
 A &= \sum_{i=1}^{\frac{n+1}{2}} x_i \log\left(\frac{2i-1}{i}\right) = x_1 \log\left(\frac{2-1}{1}\right) + x_2 \log\left(\frac{4-1}{2}\right) + x_3 \log\left(\frac{6-1}{3}\right) \dots \\
 &+ x_{\frac{n+1}{2}} \log\left(\frac{2\left(\frac{n+1}{2}\right)-1}{\frac{n+1}{2}}\right) = 0 + x_2 \log\left(\frac{3}{2}\right) + x_3 \log\left(\frac{5}{3}\right) + \dots + x_{\frac{n+1}{2}} \log\left(\frac{2n}{n+1}\right)
 \end{aligned}$$

Mientras que

$$B = \sum_{i=\frac{n+3}{2}}^n x_i \log\left(\frac{2(n-i+1)}{i}\right) = x_{\frac{n+3}{2}} \log\left(\frac{2(n-1)}{n+3}\right) + x_{\frac{n+4}{2}} \log\left(\frac{2(n-2)}{n+4}\right) \\ + x_n \log\left(\frac{2}{n}\right)$$

$\therefore A+B$

$$= x_2 \log\left(\frac{3}{2}\right) + x_3 \log\left(\frac{5}{3}\right) + \dots + x_{\frac{n+1}{2}} \log\left(\frac{2n}{n+1}\right) + x_{\frac{n+3}{2}} \log\left(\frac{2(n-1)}{n+3}\right) \\ + x_{\frac{n+4}{2}} \log\left(\frac{2(n-2)}{n+4}\right) + x_n \log\left(\frac{2}{n}\right) \\ = x_2 \log\left(\frac{3}{2}\right) + x_3 \log\left(\frac{5}{3}\right) + \dots + x_{\frac{n+1}{2}} \log\left(\frac{2n}{n+1}\right) - x_{\frac{n+3}{2}} \log(2) \left(\frac{n+3}{n-1}\right) \\ - x_{\frac{n+4}{2}} \log(2) \left(\frac{n+4}{n-2}\right) - x_n \log\left(\frac{n}{2}\right)$$

Agrupando los elementos de la suma, se tiene:

$\therefore$

$$= \left( x_2 \log\left(\frac{3}{2}\right) - x_{\frac{n+3}{2}} \log(2) \left(\frac{n+3}{n-1}\right) \right) + \left( x_3 \log\left(\frac{5}{3}\right) - x_{\frac{n+4}{2}} \log(2) \left(\frac{n+4}{n-2}\right) \right) \\ + \dots + \left( x_{\frac{n+1}{2}} \log\left(\frac{2n}{n+1}\right) - x_n \log\left(\frac{n}{2}\right) \right) = C$$

Dado que

$$\left( x_2 \log\left(\frac{3}{2}\right) \leq x_{\frac{n+3}{2}} \log(2) \binom{n+3}{n-1} \right)$$

$$\left( x_3 \log\left(\frac{5}{3}\right) \leq x_{\frac{n+4}{2}} \log(2) \binom{n+4}{n-2} \right)$$

...

$$\left( x_{\frac{n+1}{2}} \log\left(\frac{2n}{n+1}\right) \leq x_n \log\left(\frac{n}{2}\right) \right)$$

Lo que implica que  $C \leq 0$

Por lo tanto, se puede concluir que la entropía relativa para este ejemplo para datos bajo un orden curva será mayor o igual que cuando se tiene un orden creciente.

$$\sum_{i=1}^n x_i \log f(x_i)_{Orden\ Curva} = \sum_{i=1}^n x_i \log(i) + C$$

$$\leq \sum_{i=1}^n x_i \log(i) = \sum_{i=1}^n x_i \log f(x_i)_{Orden\ Creciente}$$

∴

$$-\sum_{i=1}^n x_i \log f(x_i)_{Orden\ Creciente} \leq -\sum_{i=1}^n x_i \log f(x_i)_{Orden\ Curva}$$

∴

$$D(q|f)_{Orden\ Creciente} \leq D(q|f)_{Orden\ Curva}$$

La demostración es similar para  $n$  par. También la demostración se puede generalizar y eliminar el supuesto de que  $x_i = i$ ,  $1 \leq i \leq n$ .

## 5. BIBLIOGRAFÍA

---

Cochran, W. G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. *Ann. Math. Stat.*, **17**, 164-177.

Cochran W.G. (1977). *Sampling Techniques*, Tercera edición.

Deming, William Edwards. Some theory of sampling, New Cork: J Wiley, 1955.

Diccionario de la Lengua Española, Real Academia Española, Vigésima Segunda Edición, 2001.

Kish, L. (1965). Survey sampling. John Wiley & Sons, New York.

Madow W.G., y Madow, L.H. (1944). On the theory of sistematic sampling, II. *Ann Math. Stat*, **20**, 1-24.

Renyi, Alfred (1970). Probability theory, North-Holland series in applied mathematics and mechanics, v.10.

Särndal C.E., Bengt Swensson, Jan Wretman (1972). Model Assisted Survey Sampling, Springer Series in Statistics.

Páginas visitadas en septiembre 2003

<http://www.nap.edu/books/030904376X/html/181.html>

<http://www.maphysto.dk/oldpages/events/S-and-GT2000/>

<http://www.maths.uwa.edu.au/~adrian/papers.html>

<http://citeseer.ist.psu.edu/cachin97smooth.html>