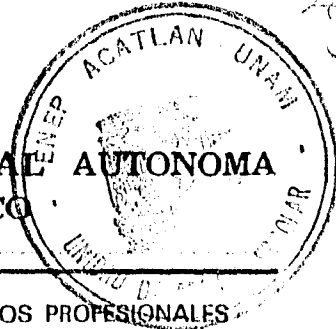




UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO



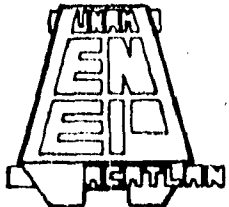
33  
2e

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES  
"ACATLAN"

EL "CRITERIO DE DALENIUS" PARA LA FORMACION  
DE ESTRATOS EN EL MUESTREO ALEATORIO  
ESTRATIFICADO PARA ESTIMAR EL TOTAL  
POBLACIONAL.  
EL CASO DE UNA RAMA DE ACTIVIDAD  
ECONOMICA EN MEXICO.

**T E S I S**

QUE PARA OBTENER EL TITULO DE  
A C T U A R I O  
P R E S E N T A  
MARITZA SANDOVAL BUSTOS



*Companado de en bichillo 8/26*

MEXICO, D. F.

1998

**TESIS CON  
FALLA DE ORIGEN**

**TESIS CON  
FALLA DE ORIGEN**



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**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.

*Dedicada a:*

*Manuel y Marfa Luisa  
por todo su amor.*

*Agradezco a:*

*Ricardo Apaicio Jiménez,  
Director de tesis y amigo, y  
a Blanca Elena del Pozo,  
asesora y amiga.*

*Jesús Fernández Morán,  
por introducirme al tema.*

## ÍNDICE

INTRODUCCIÓN.....	1
<b>CAPITULO I</b>	
<b>ALGUNOS CONCEPTOS BÁSICOS DE MUESTREO.....</b>	<b>4</b>
1.1. Introducción.....	4
1.2. Propiedades de los estimadores.....	8
1.3. Métodos de estimación.....	12
<b>CAPITULO II</b>	
<b>MUESTREO ALEATORIO SIMPLE.....</b>	<b>15</b>
2.1. El muestreo aleatorio simple (MAS).....	15
2.2. Un estimador para $Y$ y su varianza.....	18
2.3. Tamaño de muestra para estimar $Y$ .....	25
<b>CAPITULO III</b>	
<b>MUESTREO ALEATORIO ESTRATIFICADO.....</b>	<b>29</b>
3.1. El muestreo aleatorio estratificado (MAE).....	29
3.2. Un estimador para $Y$ y su varianza.....	31
3.3. Asignación del tamaño de muestra para estimar $Y$ .....	40
3.4. Una comparación de MAE y MAS.....	49

#### CAPITULO IV

<b>EL CRITERIO DE DALENIUS O REGLA <math>\text{cum}\sqrt{f(y)}</math> PARA ENCONTRAR FRONTERAS DE ESTRATOS.....</b>	<b>55</b>
4.1. Introducción.....	55
4.2. La variable de estratificación.....	56
4.3. El número de estratos.....	57
4.4. Fronteras de los estratos. Regla $\text{cum}\sqrt{f(y)}$ .....	60
4.4.1. Supuestos.....	60
4.4.2. Una aproximación a la solución exacta.....	71

#### CAPITULO V

<b>APLICACIÓN A UN PROBLEMA PRÁCTICO: " EL CASO DE UNA RAMA DE ACTIVIDAD ECONÓMICA EN MÉXICO".....</b>	<b>79</b>
5.1. Antecedentes de la información y de la aplicación.....	80
5.1. Estimación con muestreo aleatorio simple.....	84
5.2. Estimación con muestreo aleatorio estratificado.....	89
5.3. Comparación de resultados.....	111

<b>CONCLUSIONES.....</b>	<b>114</b>
--------------------------	------------

<b>BIBLIOGRAFÍA.....</b>	<b>117</b>
--------------------------	------------

#### APÉNDICES

<b>Apéndice A: Algunos parámetros de una función de distribución.....</b>	<b>118</b>
<b>Apéndice B: Cuadros de referencia.....</b>	<b>126</b>
<b>Apéndice C: Macro en Lotus-123 para estratificar.....</b>	<b>143</b>

## INTRODUCCIÓN

En el campo de la estadística, se investiga y se analiza el comportamiento de sucesos en diversas actividades, lo que implica obtener información de poblaciones relacionadas a ellos para conocer sus características. Una forma de recopilar información es el "censo". Éste investiga cada elemento de toda la población.

Para múltiples estudios, el censo es imposible de realizar debido a la variedad de recursos que requiere; por ello tiene un alcance limitado. Una alternativa al censo es el muestreo, el cual investiga algunos elementos de la población que constituyen una muestra.

La estadística trabaja con muestras para investigar y analizar sus características o propiedades, y extender resultados hacia la población con determinada medida de incertidumbre; dicho proceso se llama inferencia.

Del proceso de seleccionar muestras se encarga la teoría del muestreo, parte medular de la estadística. Su objetivo es desarrollar métodos de selección de muestras y de estimación, que permitan, con el mínimo de recursos, evaluar los estimadores con suficiente exactitud en condiciones específicas.

Se han desarrollado métodos eficientes de muestreo; los principales son: el no probabilístico y el probabilístico. Con el método de muestreo no son desconocidas las probabilidades de seleccionar las unidades y, por lo tanto, no podemos medir la confiabilidad de los resultados de la muestra.

Con el método de muestreo probabilístico, que en lo sucesivo se llamará únicamente muestreo, se puede calcular de antemano cuál es la probabilidad de obtener

cada una de las muestras que es posible seleccionar. La selección debe ser considerada como un evento aleatorio, lo que constituye la base de la teoría de la probabilidad. Este método ha reemplazado al no probabilístico para muchos propósitos prácticos.

Una situación óptima para aplicar el muestreo consiste en emplear recursos mínimos para obtener determinada información, o conseguir máxima información mediante recursos prefijados.

Con esta necesidad de optimizar, se usa el muestreo cuando la población es infinita, o tan grande que el censo exceda las posibilidades del investigador y, por grande que sea el riesgo de error, será inferior a la certeza de un costo prohibitivo.

También es práctico cuando la población es suficientemente uniforme o aceptablemente homogénea en sus unidades y cuando el riesgo de error sea despreciable, ya que una muestra de tal población da una buena generalización de ésta.

Es indispensable cuando medir o investigar las características de la población es destructivo, como ocurre al consumir un artículo para medir su calidad; o que los elementos investigados en la población no lleguen a inutilizarse, pero sí a disminuir su valor; lo mismo ocurriría en el caso de determinar una dosis letal.

Debido al interés de optimizar resultados y al tipo de población que se emplea, existen diferentes procedimientos de seleccionar una muestra. Entre los principales se pueden mencionar: el muestreo aleatorio simple, el muestreo aleatorio estratificado, el muestreo aleatorio por conglomerados, el muestreo aleatorio bietápico, el muestreo polietápico, el muestreo sistemático, el muestreo doble o bifásico, el muestreo múltiple o polifásico, las submuestras interpenetrantes, el muestreo repetido, así como diversas combinaciones de procedimientos.

El muestreo aleatorio simple, considerado como el mejor procedimiento para la introducción al muestreo, será usado para describir las principales etapas del diseño en el



El muestreo aleatorio simple, considerado como el mejor procedimiento para la introducción al muestreo, será usado para describir las principales etapas del diseño en el muestreo. Se exponen los conceptos básicos de ésta técnica para dirigir la investigación al muestreo estratificado, en que la población se puede dividir en subpoblaciones o estratos. Se llegará así a la parte fundamental de esta investigación, la cual consiste en elegir un criterio para formar los estratos. Se trata del criterio para encontrar las fronteras de los estratos, criterio de equipartición o "regla  $\text{cum}\sqrt{f}$ ", conocido también como, "criterio de Dalenius", en honor de Tore Dalenius quien lo propuso en 1950.

El objetivo de este trabajo es describir el "criterio de Dalenius" en la construcción de estratos para estimar el total poblacional, y compararlo con el muestreo aleatorio simple; tomando como ejemplo los datos del marco muestral de la "Encuesta sobre Formación de Capital Fijo" que se realiza en el Banco de México.

En resumen, la hipótesis de investigación es detallar la aplicación de los principios del "criterio de Dalenius" en la formación de estratos.

## CAPITULO I

### ALGUNOS CONCEPTOS BÁSICOS DE MUESTREO

#### 1.1. Introducción

Cuando el investigador aplica el muestreo, el primer paso es determinar la población o universo de los elementos de interés, así como las características a ser medidas.

**Población o universo ( $\Omega$ )**, se emplea en estadística, como cualquier colección finita o infinita de elementos o resultados de un experimento.

La población se divide en partes que son llamadas **unidades de muestreo ( $\omega$ )** o **unidades**. Estas son las unidades de análisis, deben cubrir la totalidad de la población y no se deben sobreponer, en el sentido de que todo elemento en la población pertenezca a una y sólo a una unidad.

$$\Omega = \{ \omega_1, \omega_2, \dots, \omega_N \},$$

donde:

$$\omega_1 \cap \omega_2 \cap \dots \cap \omega_N = \emptyset, \text{ y}$$

$$\omega_1 \cup \omega_2 \cup \dots \cup \omega_N = \Omega.$$

Su naturaleza se determina por los objetivos de la investigación. Esta lista de unidades de muestreo se llama **marco muestral** o espacio muestral y representa a la población.

Para definir a la población es necesario identificar el tipo de variables que se desean estudiar al realizar una secuencia de experimentos. De acuerdo con esto, una población puede ser de dos tipos: población discreta o población continua.

**Población discreta:** cuando la variable  $Y$  asociada a los elementos de la población tiene un conjunto finito o infinito numerable de datos posibles.

**Población continua:** cuando la variable  $Y$  asociada a los elementos de la población tiene un conjunto infinito no numerable de datos posibles.

Si se considera a la población como una colección de elementos que se asocian a una variable de escala de intervalo<sup>\*</sup>, el conjunto ordenado de las diferentes medidas de dichos elementos con la frecuencia de cada medida, definida de manera estática, constituye una distribución de frecuencias llamada **distribución poblacional**  $f(y)$ .

Una muestra consiste en una o más (en general  $n$ ) realizaciones de un experimento aleatorio de selección. Es como la población, un ente estático formado por  $n$  unidades (en este caso observadas). A la muestra le corresponde una distribución de frecuencias muestral (la distribución de la muestra). Al número  $n$  de unidades se le llama **tamaño de la muestra**.

---

\* Una variable con escala de intervalo está caracterizada por una unidad de medida común y constante que asigne un número real a todos los pares de objetos en un conjunto ordenado. En esta investigación siempre se tratará de variables de escala de intervalo.

La muestra se extrae de la población con el fin de estudiarla. A partir de los datos observados en la muestra, la "inferencia estadística" permite estimar con cierto margen de error las características que puede tener la población .

La **muestra** es entonces una parte de la población o un subconjunto de unidades, obtenidas con el fin de investigar sus propiedades y generalizarlas a la población o conjunto de procedencia.

El caso límite en que la muestra es igual a la población se llama **censo**, es la enumeración y anotación exhaustiva de ciertas características de una población.

En cuanto a los censos y recuentos, aparte de que todo censo da un valor particular de la población en el tiempo, la gran extensión y dificultad de las operaciones exhaustivas cuando se extienden a una población muy grande, hace tomar a menudo la alternativa de emplear métodos que generalicen a la población, como el de muestreo.

El **muestreo** es el proceso de selección y análisis de una muestra. "Es un método de seleccionar una fracción de la población, de tal manera que, la muestra seleccionada puede ser generalizada a la población entera."<sup>1</sup>

De la fracción de la población, existe el coeficiente  $\frac{n}{N} = f$  llamado **fracción de muestreo**, y el inverso de  $f$ ,  $\frac{N}{n}$  llamado **factor de expansión**, Inflación o elevación. Este término se usa algunas veces para expandir de valores muestrales a valores poblacionales.

---

<sup>1</sup> SUKHATME, Pandurang V.: *Sampling theory of survey with applications*, The Indian Society of Agricultural, New Delhi, India, 1953 p. 1.

También cabe mencionar al factor<sup>2</sup>  $\left(1 - \frac{n}{N}\right)$  ó  $(1-f)$  que se denomina **corrección para poblaciones finitas** (CPF). Frecuentemente, la teoría supone a la población de tamaño infinito; en el caso especial de una población finita, este factor es muy próximo a 1 cuando  $N$  es grande con respecto a  $n$ , es decir, cuando la fracción de muestreo es pequeña<sup>3</sup>. Si esto último sucede, la CPF puede ser omitida. La CPF aparece como factor de corrección de la porción principal de los términos de la varianza, y su raíz para el error estándar; en una población infinita el factor desaparece de la fórmula de varianza; de ahí su nombre<sup>4</sup>.

Existen actualmente una gran variedad de planes para seleccionar una muestra. Por cada plan, se pueden hacer estimaciones del tamaño de la muestra partiendo de un conocimiento del nivel de precisión deseado. Los costos relativos y el tiempo involucrado para cada plan son también comparados antes de tomar una decisión.<sup>5</sup>

El principio a seguir en la selección de un método de muestreo es, en realidad, el de obtener el resultado deseado con la precisión requerida a costo mínimo, o con la máxima precisión al costo dado, haciendo el uso más eficaz de los recursos disponibles.<sup>6</sup>

La precisión puede medirse considerando la dispersión de la distribución del estimador, es decir la varianza o raíz cuadrada, el error de muestreo o error estándar del

---

<sup>2</sup> Se considerará indistinto el factor  $\left(1 - \frac{n}{N}\right)$  ó  $\left(1 - \frac{n}{N-1}\right)$

<sup>3</sup> AZORIN, Francisco: *Curso de muestreo y aplicaciones*, México, Aguilar, 1971, p. 55.

<sup>4</sup> KISH, Leslie: *Muestreo de encuestas*, México, Trillas, 1979, p. 68.

<sup>5</sup> COCHRAN, William G.: *Técnicas de muestreo*, México, Continental, 1981, p. 28.

<sup>6</sup> SUKHATMÉ, P. V., *op. cit.*, p. 3.

estimador. La precisión tiene una relación inversa con la varianza, así, cuando la precisión es máxima, la varianza es mínima.<sup>7</sup>

"El establecimiento de la precisión deseada puede ser hecho indicando la cantidad de error que estamos dispuestos a tolerar en los estimadores muestrales. Esta cantidad se determina a la luz de los usos que tendrán los resultados de la muestra."<sup>8</sup>

## 1.2. Propiedades de los estimadores

En el espacio muestral puede considerarse una distribución probabilística  $n$  dimensional, constituida por las unidades  $\{\omega_1, \omega_2, \dots, \omega_n\}$  con los valores asociados  $\gamma(\omega_i) = p_i$ , correspondientes a todas las muestras posibles con las restricciones que impone el diseño y las probabilidades respectivas.

Considerando al muestreo como un experimento aleatorio, si para cada posible muestra se calcula el valor numérico de una cierta función  $\hat{\theta}$  de  $n$  datos, entonces esta función se llama estadístico o estimador. Cuando se toma una muestra en particular, la función toma un valor  $\hat{\theta} = \hat{\theta}(y_1, \dots, y_n)$ , al que se le llama estimación.

Estos valores asociados de las probabilidades de obtener las muestras respectivas constituyen una distribución probabilística, la llamada distribución muestral del estadístico  $\hat{\theta}$ , la cual es una variable aleatoria.

---

<sup>7</sup> AZCARRA, F. T., M. H. I. ET AL.  
<sup>8</sup> FORTINER, M. E. ET AL. 1968.

Se trata de estimar mediante  $\hat{\theta}$  una característica poblacional (parámetro poblacional) o un valor verdadero  $\theta$ . En tal caso se dice que  $\hat{\theta}$  es un estimador de  $\theta$ . Para cada muestra se tiene un valor numérico o estimación. El estimador es pues, una variable aleatoria en el muestreo cuyos valores numéricos son las estimaciones posibles.

El objetivo del proceso de estimación es encontrar buenos estimadores de los parámetros poblacionales, tales que reúnan ciertas propiedades que permitan reducir los errores de aproximación a los parámetros poblacionales.

Para que un estimador sea considerado bueno, debe tener las propiedades de insesgabilidad, eficiencia, consistencia y suficiencia.

### a) Insesgabilidad

Un estadístico  $\hat{\theta}$  es un estimador insesgado del parámetro  $\theta$  si y sólo si

$$E(\hat{\theta}) = \theta$$

En general,

$$E(\hat{\theta}) = \theta + b_1(\theta),$$

donde  $b_1(\theta)$  es el sesgo de  $\hat{\theta}$  para estimar  $\theta$ , es decir, es la desviación de  $E(\hat{\theta})$  a  $\theta$ .

Cuando  $b_1(\theta) = 0$  entonces  $\hat{\theta}$  es insesgado. O bien si  $b_1(\theta) \rightarrow 0$  cuando  $n \rightarrow \infty$

entonces el estimador  $\hat{\theta}$  es asintóticamente insesgado.

### b) Eficiencia

Si se tiene que elegir uno de varios estimadores insesgados, por lo general se toma aquél cuya distribución muestral tiene la menor varianza. Se dice que  $\hat{\theta}$  es el estimador más eficiente si es insesgado y de varianza mínima. La varianza de un estimador satisface la desigualdad de *Cramer-Rao*:

$$\sigma^2(\hat{\theta}) \leq \frac{1}{n \cdot E \left( \frac{\partial \ln f(y)}{\partial \theta} \right)^2}.$$

La varianza del estadístico  $\hat{\theta}$  queda denotada por  $\sigma^2(\hat{\theta})$ , donde  $f(y)$  es la función de densidad<sup>9</sup> de la variable  $Y$  y  $n$  es el tamaño de la muestra.

Si  $\hat{\theta}_1$  y  $\hat{\theta}_2$  son dos estimadores insesgados de un parámetro  $\theta$ , y

$$\sigma^2(\hat{\theta}_1) < \sigma^2(\hat{\theta}_2),$$

se dice que  $\hat{\theta}_1$  es relativamente más eficiente que  $\hat{\theta}_2$ .

Asimismo, se utiliza la razón, comparada con la unidad,

$$Ef = \frac{\sigma^2(\hat{\theta}_1)}{\sigma^2(\hat{\theta}_2)}.$$

¿Qué tan pequeña puede ser la varianza de un estimador? La desigualdad de *Cramer-Rao*, proporciona una cota inferior para  $\sigma^2(\hat{\theta})$ .

---

<sup>9</sup> La función de densidad es la distribución probabilística en una población continua.



$$\sigma^2(\hat{\theta}) \geq \frac{\sigma^2}{n}$$

$\hat{\theta}$  puede ser la media, el total, o algún otro estadístico muestral.

### c) Consistencia

El estadístico  $\hat{\theta}$  es un estimador consistente del parámetro  $\theta$  si y sólo si para alguna constante positiva  $\alpha$ :

$$p(|\hat{\theta} - \theta| \geq \alpha) = 0,$$

o en forma equivalente, si y sólo si:

$$\lim_{n \rightarrow \infty} p(|\hat{\theta} - \theta| < \alpha) = 1.$$

La consistencia es una propiedad asintótica; esto es porque un estimador puede ser consistente sin ser insesgado. Un estimador sesgado puede ser consistente sólo si es asintóticamente insesgado.

En la práctica, se puede juzgar si un estimador es consistente mediante el uso de condiciones suficientes (a pesar de no ser condiciones necesarias) que son una consecuencia inmediata del teorema de Chebyshev, las cuales se mencionan a continuación:

El estadístico  $\hat{\theta}$  es un estimador consistente del parámetro  $\theta$  si:

i)  $\hat{\theta}$  es insesgado,

ii)  $\sigma^2(\hat{\theta}) \rightarrow 0$  cuando  $n \rightarrow \infty$ .

#### d) Suficiencia

Se dice que un estimador  $\hat{\theta}$  es suficiente si utiliza toda la información de una muestra relevante para la estimación del parámetro  $\theta$  de la población; es decir, si todo el conocimiento que se obtiene acerca de  $\theta$  está dado solamente por la especificación de los valores del estadístico  $\hat{\theta}$  de la muestra.

En otras palabras, el estadístico  $\hat{\theta}$  es un estimador suficiente del parámetro  $\theta$  si y sólo si para cada valor de  $\hat{\theta}$ , la distribución condicional de la muestra  $y_1, y_2, \dots, y_n$  dado  $\hat{\theta}$  es independiente de  $\theta$ .

Se puede decir también que el estadístico  $\hat{\theta}$  es un estimador suficiente del parámetro  $\theta$  si y sólo si la densidad conjunta o distribución de probabilidad de la muestra se puede factorizar de manera que:

$$f(y_1, y_2, \dots, y_n, \hat{\theta}) = g(\hat{\theta}, \theta) \cdot h(y_1, y_2, \dots, y_n),$$

donde  $g(\hat{\theta}, \theta)$  depende sólo de  $\hat{\theta}$  y  $\theta$ , y  $h(y_1, y_2, \dots, y_n)$  no depende de  $\theta$ .

### 1.3. Métodos de estimación

Los métodos de estimación pueden ser clasificados en **estimación puntual** y **estimación por intervalos** (usualmente se usa el concepto de intervalos de confianza).

Para la estimación puntual hay métodos tales como el método de máxima verosimilitud, el método de mínimos cuadrados y el método de momentos.

En estudios de muestreo, el principal método de estimación es por intervalos.

Una estimación por intervalo de un parámetro  $\theta$  es con un intervalo de la forma  $\hat{\theta}_1 < \theta < \hat{\theta}_2$  donde  $\hat{\theta}_1$  y  $\hat{\theta}_2$  dependen del valor que tome  $\hat{\theta}$  en una muestra dada y de la distribución muestral de  $\hat{\theta}$ .

Se usa la distribución muestral de  $\hat{\theta}$  para elegir  $\hat{\theta}_1$  y  $\hat{\theta}_2$  tal que para cualquier nivel de confianza  $\alpha$ :

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha \quad \text{y} \quad 0 < \alpha < 1.$$

Este intervalo  $\hat{\theta}_1 < \theta < \hat{\theta}_2$  determinado en relación a una muestra particular, recibe el nombre de intervalo de confianza; la fracción  $(1 - \alpha)$  se conoce como coeficiente o **nivel de confianza**, y los extremos  $\hat{\theta}_1$  y  $\hat{\theta}_2$  reciben el nombre de **límites de confianza inferior y superior**, respectivamente.

Por lo general, cualquiera que sea el método de selección, una estimación diferirá con un error estándar de lo que se obtendría de un censo.

El **error muestral** es la diferencia entre un estimador muestral y el parámetro poblacional:

$$e = \hat{\theta} - \theta.$$

Este concepto se relaciona con las ideas de insesgamiento y valor esperado. En particular, si  $\hat{\theta}$  es un estimador insesgado de  $\theta$ :

$$E(\hat{\theta}) = \theta \quad \text{y} \quad e = \hat{\theta} - E(\hat{\theta}).$$

La desviación estándar de la distribución muestral del estimador es llamado **error estándar**; éste es la raíz cuadrada de la varianza de la distribución muestral, la cual es:

$$\sigma(\hat{\theta}) = \sqrt{\sigma^2(\hat{\theta})}$$

Diferentes diseños muestrales darán lugar a diferentes errores estándar y un principio de muestreo consiste en escoger el diseño con el error más pequeño.

## CAPITULO II

### MUESTREO ALEATORIO SIMPLE

#### 2.1. El muestreo aleatorio simple (MAS)

"El diseño muestral tiene dos aspectos: un proceso de selección que establece las reglas y operaciones con las que algunos miembros de la población son incluidos en la muestra; y un proceso de estimación para calcular los estadísticos muestrales, los cuales son estimadores muestrales de los valores poblacionales"<sup>1</sup>.

El muestreo aleatorio simple (MAS) es un método que consta en seleccionar  $n$  unidades, pertenecientes a una población ( $\Omega$ ) de tamaño  $N$ , de tal manera que cada una de las combinaciones de  $N$  en  $n$ ,  $\binom{N}{n}$ , que es el número total de posibles muestras, tiene igual oportunidad de ser seleccionada. La muestra seleccionada de este modo se conoce como muestra aleatoria simple.

La selección en el MAS puede ser de dos formas, con reemplazo o sin reemplazo. Considérese una población de tamaño  $N$  y una muestra determinada como una colección de  $n$  unidades específicas de esta población. En la selección con reemplazo, la extracción de  $n$  unidades de la población tiene la misma probabilidad de ser elegida.

---

<sup>1</sup> KISH, Leslie: *Muestreo de encuestas*, México, Trillas, 1979, p. 24.

Formalmente coincide con el muestreo de poblaciones infinitas, ya que al devolver a la población cada elemento extraído de la misma, la población es inagotable y el resultado de la extracción de cada elemento es independiente de los anteriores a él.

La segunda forma es la selección sin reemplazo o irrestrictamente aleatoria, donde todas las unidades de la población tienen también la misma probabilidad de ser extraídas; pero si la población es finita, la probabilidad de elegir un elemento dependerá de los elegidos anteriormente para formar parte de la muestra y que dejaron de pertenecer a los seleccionables. Así, en la primera extracción, la probabilidad de que se seleccione una de estas  $n$  unidades es  $\frac{n}{N}$ ; en la segunda extracción, la probabilidad de que se seleccione

una de las restantes  $(n-1)$  unidades especificadas es  $\left(\frac{n-1}{N-1}\right)$  y así sucesivamente.

Por lo tanto, la probabilidad de que se extraigan las  $n$  unidades especificadas es:

$$\begin{aligned} \frac{n}{N} \cdot \frac{(n-1)}{(N-1)} \cdot \frac{(n-2)}{(N-2)} \cdots \frac{1}{(N-n+1)} &= \\ &= \frac{n!(N-n)!}{N!} \\ &= \frac{1}{{}^N C_n} \end{aligned}$$

El nombre de MAS, que es un tipo de muestreo equiprobabilístico, se ha aplicado tanto a este tipo de selección como al anterior. Para fines de esta investigación, la selección sin reemplazo será mencionada como MAS.

Para hacer una selección aleatoria de muestras se usan las tablas de números aleatorios o programas que generen números aleatorios. "La selección se realiza unidad por unidad, primero se numeran las unidades de 1 a  $N$ ; posteriormente se extrae una serie de  $n$  números aleatorios entre 1 y  $N$ . En cada extracción el proceso debe otorgar la misma oportunidad de selección a todos y cada uno de los números que no hayan sido escogidos. Las unidades que llevan estos  $n$  números constituirán la muestra"<sup>2</sup>.

En general, los dos aspectos del diseño muestral se pueden desglosar de la siguiente manera:

- El proceso de selección:
  - Determinar el procedimiento de muestreo que se usará dependiendo de las condiciones de la población y de los fines de la estimación.
  - Determinar el tamaño de la muestra por seleccionar.
  - Seleccionar la muestra; para el caso del MAS, la selección es aleatoria.
- El proceso de estimación:
  - Calcular los estadísticos o estimadores de los parámetros poblacionales que interesa estimar.
  - Calcular el error estándar de esos estadísticos para evaluar su precisión, ya que está en función del error.
  - Construir intervalos de confianza para los parámetros poblacionales.

---

<sup>2</sup> COCHRAN, William G.: *Técnicas de muestreo*, México, Continental, 1981, p. 41.

## 2.2. Un estimador para $Y$ y su varianza

Una vez seleccionado el procedimiento muestral, se procede a estimar los parámetros poblacionales, como la media, el total, proporciones, varianza, etc.

Sean  $\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$  una población de tamaño  $N$ ,  $\{\omega_1, \omega_2, \dots, \omega_n\}$  una muestra de tamaño  $n$ , y  $Y(\omega_i) = y_i$  una variable asociada a los elementos de la población. El total poblacional\*, la media poblacional, el total muestral y la media muestral, para la variable  $Y$  son definidos como:

<p>Total poblacional</p> $Y = \sum_{i=1}^N y_i = y_1 + \dots + y_i + \dots + y_N$
<p>Media poblacional</p> $\bar{Y} = \frac{1}{N} \sum_{i=1}^N y_i = \frac{1}{N} (y_1 + \dots + y_i + \dots + y_N)$
<p>Total muestral</p> $y = \sum_{i=1}^n y_i = y_1 + \dots + y_i + \dots + y_n$
<p>Media muestral</p> $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} (y_1 + \dots + y_i + \dots + y_n)$

\* Advertir la diferencia de la notación de la variable  $Y$  y del estimador del total  $Y$ .



Empezando por el caso del estimador de la media poblacional: se dice que un buen estimador de la media poblacional  $\bar{Y}$  es la media muestral  $\bar{y}$  porque reúne las condiciones de un buen estimador.

La media muestral  $\bar{y}$  es un estimador insesgado y consistente de  $\bar{Y}$ . La prueba de eficiencia y suficiencia no se consideran en esta investigación debido a que ya han sido demostradas y porque no es objetivo de este trabajo. Sin embargo, se puede consultar en Yamane<sup>3</sup>.

Para demostrar la insesgabilidad de  $\bar{y}$ , el valor esperado de la media muestral, debe ser  $E(\hat{Y}) = \bar{Y}$ . La media de todas las posibles medias muestrales es:

$$E(\hat{Y}) = E(\bar{y}) = \sum_{G} \bar{y} p(\bar{y}),$$

donde:

$$p(\bar{y}) = \frac{1}{{}_N C_n}, y$$

$$\bar{y} = \frac{1}{n} (y_1 + y_2 + \dots + y_n)$$

Entonces<sup>4</sup>:

$$\begin{aligned} E(\bar{y}) &= \frac{\sum_{G} (y_1 + y_2 + \dots + y_n)}{{}_N C_n} \\ &= \frac{{}_{N-1} C_{n-1} (y_1 + y_2 + \dots + y_N)}{{}_N C_n} \end{aligned}$$

<sup>3</sup> YAMANE, Taro: *Elementary Sampling Theory*, Prentice-Hall, 1967, pp. 42 y sig.

<sup>4</sup> COCHRAN, W. G., 1981, *op. cit.*, p. 45.



Si se descompone el término cuadrático:

$$\sigma^2(\bar{y}) = \frac{1}{n^2} E \left[ \sum_{i=1}^n (y_i - \bar{Y})^2 + 2 \sum_{i \neq j} (y_i - \bar{Y})(y_j - \bar{Y}) \right],$$

y se utiliza la definición de valores esperados para el primer término del lado derecho,

$$E(y_i - \bar{Y})^2 = \sum_{i=1}^N \frac{1}{N} (y_i - \bar{Y})^2 = \sigma^2,$$

lo mismo para el segundo término,

$$E(y_i - \bar{Y})(y_j - \bar{Y}) = \frac{1}{N} \left( \frac{1}{N-1} \right) \sum_{i \neq j} (y_i - \bar{Y})(y_j - \bar{Y}),$$

donde la probabilidad de seleccionar  $(y_i - \bar{Y})(y_j - \bar{Y})$  sobre el  $i$ -ésimo o  $j$ -ésimo

experimento es  $\left( \frac{1}{N} \right) \left( \frac{1}{N-1} \right)$ . La suma puede ser combinada como sigue:

$$\begin{aligned} \sum_{i \neq j} (y_i - \bar{Y})(y_j - \bar{Y}) &= \left[ \sum_{i=1}^N (y_i - \bar{Y}) \right]^2 - \sum_{i=1}^N (y_i - \bar{Y})^2 \\ &= - \sum_{i=1}^N (y_i - \bar{Y})^2. \end{aligned}$$

Si se combinan estos resultados y se sustituyen términos, obtenemos:

$$\begin{aligned} \sigma^2(\bar{y}) &= \frac{1}{n^2} \sum_{i=1}^n \sigma^2 + \frac{1}{n^2} \frac{1}{N} \frac{1}{N-1} \sum_{i \neq j} \left[ - \sum_{i=1}^N (y_i - \bar{Y})^2 \right] \\ &= \frac{1}{n^2} (n \sigma^2) - \frac{1}{n^2} \frac{1}{N} \sum_{i \neq j} \sigma^2 \end{aligned}$$

$$= \frac{\sigma^2}{n} - \frac{1}{n^2} \frac{1}{N} n(n-1) \sigma^2$$

$$= \frac{N-n}{N} \frac{\sigma^2}{n}.$$

La varianza de la media  $\sigma^2(\bar{y})$  puede verse también como:

$$\sigma^2(\bar{y}) = (1 - f) \frac{1}{n} \sigma^2.$$

Y el error estándar de  $\bar{y}$  es:

$$\sigma(\bar{y}) = \sqrt{\sigma^2(\bar{y}) (1 - f)} = \frac{\sigma}{\sqrt{n}} \sqrt{(1 - f)}.$$

El estimador  $\bar{y}$  es un estimador consistente del parámetro poblacional  $\bar{Y}$ . De las condiciones suficientes para la consistencia de un estimador:

- i)  $\bar{y}$  es un estimador insesgado de  $\bar{Y}$ , y
- ii)  $\sigma^2(\bar{y}) \rightarrow 0$  cuando  $n \rightarrow \infty$ .

La primera condición de consistencia se cumple. Para la segunda:

$$\lim_{n \rightarrow \infty} \sigma^2(\bar{y}) = \lim_{n \rightarrow \infty} \frac{\sigma^2}{n} (1 - f),$$

ya que:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \rightarrow 0,$$

entonces:

$$\lim_{n \rightarrow \infty} \sigma^2(\bar{y}) \rightarrow 0.$$

---

\* Para una población infinita, la varianza del estimador de la media es:  $\sigma^2/n$ .

$\bar{y}$  es un estimador insesgado de  $\bar{Y}$  y  $\sigma^2(\bar{y}) \rightarrow 0$  cuando  $n \rightarrow \infty$ . Por lo tanto,  $\bar{y}$  es un estimador consistente del parámetro poblacional  $\bar{Y}$ .

Ahora, si bien la media muestral sirve para estimar la media poblacional, el total de la muestra es casi siempre mucho menor que el total poblacional y no sirve como estimador.

Para estimar el parámetro poblacional del total, un buen estimador es el producto del llamado factor de expansión por el total muestral:  $\hat{Y} = \frac{N}{n} \sum_{i=1}^n y_i$ ; o bien el producto del tamaño de la población por la media muestral:  $\hat{Y} = N\bar{y}$ .

$\hat{Y}$  es insesgado, consistente, suficiente y eficiente, las dos primeras propiedades serán demostradas.

$\hat{Y} = N\bar{y}$  es un estimador insesgado y consistente del total de la población  $Y$ , es decir,

$$E(\hat{Y}) = Y,$$

si:

$$E(\hat{Y}) = E(N\bar{y}).$$

Por la propiedad de valor esperado:

$$E(N\bar{y}) = N E(\bar{y}) \quad \text{y}$$

$$E(\bar{y}) = \bar{Y}.$$

entonces:

$$N E(\bar{y}) = N \bar{Y}$$

$$= N \left( \frac{1}{N} \sum_{i=1}^N y_i \right) = \sum_{i=1}^N y_i = Y.$$

Por lo tanto  $E(\hat{Y}) = Y$ , lo que prueba la insesgabilidad de  $\hat{Y} = N\bar{y}$ .

Como un estimador insesgado de la población total, La varianza de  $\hat{Y} = N\bar{y}$  es:

$$\begin{aligned} \sigma^2(\hat{Y}) &= \sigma^2(N\bar{y}) = N^2 \sigma^2(\bar{y}) = \\ &= \frac{N\sigma^2}{n} (N - n), \\ \sigma^2(\hat{Y}) &= \frac{N^2 \sigma^2}{n} \frac{N - n}{N} = \frac{N^2 \sigma^2}{n} (1 - f) \end{aligned}$$

El error estándar de  $\hat{Y}$  es:

$$\sigma(\hat{Y}) = \frac{N\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N}} = \frac{N\sigma}{\sqrt{n}} \sqrt{1-f}.$$

La consistencia de  $\hat{Y} = N\bar{y}$ , queda probada de la siguiente manera: primero,  $\hat{Y} = N\bar{y}$  es un estimador insesgado; enseguida, la segunda condición:

$$\lim_{n \rightarrow \infty} \sigma^2(\hat{Y}) = \lim_{n \rightarrow \infty} \frac{1}{n} N^2 \sigma^2 (1 - f),$$

ya que  $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$ , entonces:

$$N^2 \lim_{n \rightarrow \infty} \frac{1}{n} \sigma^2 (1 - f) = N^2 \lim_{n \rightarrow \infty} \sigma^2(\bar{y}) = N^2 \cdot 0 = 0.$$

Así pues,  $\hat{Y} = N\bar{y}$  también es un estimador consistente de  $Y$ .

### 2.3. Tamaño de muestra para estimar $Y$

"El diseño de la investigación se basa en los fines, limitaciones y recursos relativos a la misma. Para calcular cuál va a ser el tamaño de la muestra o número de elementos a seleccionar, se identifican primero la característica o características a estimar ( como la media, el total, etc.) y el grado de confianza y precisión requeridos, de modo que los resultados no sean en exceso costosos o imprecisos."<sup>6</sup>

Para medir una característica de la población, en el MAS debe indicarse lo que es esperado de la muestra, ya sea en términos de los límites de error deseado, o de alguna decisión o acción que se va a tomar cuando los resultados de la muestra sean conocidos.

Se debe encontrar una ecuación que permita al tamaño de muestra  $n$ , cumplir con la precisión deseada. La ecuación variará de acuerdo a la precisión y al tipo de muestreo que se están considerando. Contendrá, como parámetros, ciertas propiedades desconocidas de la población. Estos deben ser estimados con el fin de dar resultados específicos.

Se trata de una estimación por intervalos; así, el intervalo de confianza para una variable de estudio  $Z$  con el nivel de confianza  $(1 - \alpha)$  es:

$$P(-k < Z < k) = 1 - \alpha,$$

---

<sup>6</sup> AZORIN, F., 1971, *op. cit.*, p 96.

donde  $k$  es la desviación correspondiente al nivel de confianza. Se tiene que tomar un cierto valor, obtenido como desviación normal cuando  $N \geq 30$  o como desviación  $t$  de Student cuando  $N < 30$ .

Por el teorema del límite central, la variable  $Z$  se distribuye asintóticamente como una distribución normal estandarizada:

$$Z = \frac{\hat{\theta} - E(\hat{\theta})}{\sigma(\hat{\theta})}$$

Si el estimador  $\hat{\theta}$  es insesgado de  $\theta$ , entonces:

$$Z = \frac{\hat{\theta} - \theta}{\sigma(\hat{\theta})}$$

Se sustituye la variable  $Z$  en el intervalo de confianza:

$$P\left(-k < \frac{\hat{\theta} - \theta}{\sigma(\hat{\theta})} < k\right) = 1 - \alpha,$$

se llega así a una relación de la precisión y el nivel de confianza:

$$P(-k\sigma(\hat{\theta}) < \hat{\theta} - \theta < k\sigma(\hat{\theta})) = 1 - \alpha,$$

y en términos absolutos queda:

$$P(|\hat{\theta} - \theta| < k\sigma(\hat{\theta})) = 1 - \alpha.$$

Entonces, al nivel de confianza  $(1 - \alpha)$ , el valor que  $|\hat{\theta} - \theta|$  puede tomar es:

$$\epsilon = |\hat{\theta} - \theta| = k\sigma(\hat{\theta})$$



Se parte de esta ecuación para obtener el tamaño de la muestra  $n$ , en función del error  $\epsilon$  y del nivel de confianza  $(1 - \alpha)$ . Esto es, se expresa el error absoluto o semi-intervalo confidencial, como: producto del error de muestreo o error estándar del estimador  $\sigma(\hat{\theta})$  multiplicado por una desviación  $k$  que corresponde al nivel de confianza  $(1 - \alpha)$  en una distribución Normal (si  $N \geq 30$ ).

Nuevamente, se empieza por el caso de la media poblacional. Para estimar la media poblacional  $\bar{Y}$  mediante la media muestral  $\bar{y}$  con un error admisible  $\epsilon$  y un nivel de confianza  $(1 - \alpha)$ , se sustituye  $\sigma(\hat{\theta}) = \sigma(\bar{y})$  en la igualdad anterior:

$$\epsilon = k\sigma(\bar{y}) = k\sqrt{\frac{\sigma^2}{n} \left( \frac{N-n}{N} \right)}.$$

Al elevar la ecuación al cuadrado,

$$\epsilon^2 = k^2 \left( \frac{\sigma^2}{n} \frac{N-n}{N} \right),$$

y, al despejar para encontrar  $n$ :

$$n = \frac{Nk^2\sigma^2}{N\epsilon^2 + k^2\sigma^2}.$$

Se obtiene así, la ecuación para encontrar el tamaño de muestra para estimar  $\bar{Y}$ .

Para estimar el total poblacional  $Y$ , la ecuación fundamental que relaciona el error admisible  $\epsilon$  con el nivel de confianza  $(1 - \alpha)$  y su respectiva desviación  $k$  es:

$$P(-k\sigma(\bar{y}) < N\bar{y} - Y < k\sigma(\bar{y})) = 1 - \alpha,$$

$$P(|N\bar{y} - Y| < k\sigma(\hat{Y})) = 1 - \alpha,$$

$$\varepsilon = |N\bar{y} - Y| = k\sigma(\hat{Y})$$

Elevando al cuadrado y sustituyendo  $\sigma^2(\hat{Y}) = N(N - n) \frac{\sigma^2}{n}$  en la ecuación anterior,

$$\varepsilon^2 = k^2\sigma^2(\hat{Y}) = k^2N(N - n) \frac{\sigma^2}{n}.$$

Por lo que:

$$n = \frac{N^2k^2\sigma^2}{\varepsilon^2 + Nk^2\sigma^2},$$

fórmula que permite calcular el tamaño de la muestra para estimar  $Y$  a partir de:

- La desviación  $k$  correspondiente al nivel de confianza  $(1 - \alpha)$ , en una distribución Normal (si  $N \geq 30$ ), y el error admisible  $\varepsilon$ . Éstos se fijan con arreglo a los objetivos y limitaciones de la investigación.
- El tamaño de la población  $N$ , que se supone conocida.
- La varianza poblacional  $\sigma^2$ , puede calcularse a partir del conocimiento de la población en alguna fecha anterior, de resultados relativos a poblaciones altamente correlacionadas a esta o de una muestra piloto previamente seleccionada.

## CAPITULO III

### MUESTREO ALEATORIO ESTRATIFICADO

#### 3.1. El muestreo aleatorio estratificado (MAE)

En el MAS, la población  $\Omega$  de tamaño  $N$  se explica por una muestra de tamaño  $n$  seleccionada de tal manera que infiera las características poblacionales a un nivel de confianza y precisión determinados.

En el muestreo estratificado, la población  $\Omega$  de tamaño  $N$  se divide en  $r$  subpoblaciones ( $\Omega_h, h = \overline{1, r}$ ) llamadas estratos, tal que

$$\Omega = \Omega_1 \cup \Omega_2 \cup \dots \cup \Omega_r,$$

donde:

$$\Omega_h \neq \emptyset \text{ y } \Omega_h \cap \Omega_l = \emptyset \text{ si } h \neq l \text{ y } h, l = \overline{1, r}.$$

Se denota al tamaño poblacional del estrato  $\Omega_h$  como  $N_h$ , donde

$$N_h \geq 1 \text{ y } N = N_1 + N_2 + \dots + N_r.$$

Para cada estrato  $\Omega_h$ , se seleccionan  $n_h$  elementos que conformen la muestra de  $\Omega_h$ , donde

$$n_h > 0 \quad \forall h \text{ y } n = n_1 + n_2 + \dots + n_r.$$

Así, la muestra  $n$  está constituida por elementos de cada estrato. Además las  $r$  muestras son independientes. Si la selección de muestra en cada estrato se hace con MAS, entonces se llama muestreo aleatorio estratificado (MAE).

El objetivo principal del MAE es formar estratos de forma tal que se llegue a un grado de homogeneidad dentro de ellos y heterogeneidad entre ellos, es decir, que se agrupen los elementos con características comunes.

Al hacer homogéneos los estratos, la estratificación puede conseguir un aumento en la precisión, es decir, una disminución en la varianza de los estimadores de las características de toda la población. También puede recurrirse al MAE cuando se requiera información con cierta precisión en algunos estratos, aprovechando sus particularidades, tomándolos como dominio de estudio o para aplicar diferentes métodos de selección de muestra en cada uno. Otras situaciones para el uso del MAE son: cuando los costos o la información *a priori* difiere entre los estratos; o por la distribución física, cuando los elementos están geográficamente muy dispersos.

Al igual que en MAS, el MAE consiste de dos etapas, las cuales consisten de:

- El proceso de selección:
  - Determinar el procedimiento de muestreo para estratificar dependiendo de las condiciones de la población y de los fines de las estimaciones.
  - Decidir qué criterio tomar para estratificar a la población.
  - Seleccionar la muestra; para este caso, la selección es aleatoria.

- El proceso de estimación:
  - Calcular los estadísticos o estimadores de los parámetros poblacionales de cada estrato y ponderarlos adecuadamente para formar una estimación combinada de la población completa.
  - Calcular el error estándar de esos estadísticos para cada estrato y sumar en una estimación adecuada para el total.
  - Construir intervalos de confianza para los parámetros poblacionales.
  - Encontrar el tamaño de muestra por estrato a ser seleccionada.

### 3.2. Un estimador para $Y$ y su varianza

Sea  $\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$  una población de tamaño  $N$  dividida en  $r$  estratos:

$\Omega_1 = \{\omega_{11}, \dots, \omega_{1h}, \dots, \omega_{1N_1}\}, \dots, \Omega_h = \{\omega_{h1}, \dots, \omega_{hh}, \dots, \omega_{hN_h}\}, \dots, \Omega_r = \{\omega_{r1}, \dots, \omega_{rh}, \dots, \omega_{rN_r}\};$  y

una muestra  $\{\omega_{h1}, \dots, \omega_{hh}, \dots, \omega_{hh_n}\}$  para cada estrato respectivamente, y la variable

$Y(\omega_i) = y_i$  asociada a los elementos de la muestra.

La media y el total del estrato  $\Omega_h$  son definidos como:

$Y_h = \sum_{i=1}^{N_h} y_{hi}$	Total poblacional del estrato $h$
$\bar{Y}_h = \frac{\sum_{i=1}^{N_h} y_{hi}}{N_h}$	Media poblacional del estrato $h$
$y_h = \sum_{i=1}^{n_h} y_{hi}$	Total muestral del estrato $h$
$\bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$	Media muestral del estrato $h$

Se definen también:

$\frac{N_h}{N} = W_h$	Ponderador relativo de la población del estrato $h$
$\frac{n_h}{n} = w_h$	Ponderador relativo de la muestra del estrato $h$
$\frac{n_h}{N_h} = f_h$	Fracción de muestreo del estrato $h$

La media y el total para toda la población y la muestra, es decir, la combinación de los  $r$  estratos, son definidos como:

Total de la población estratificada	
$Y = \sum_{h=1}^r \sum_{i=1}^{N_h} y_{hi} \quad \text{ó} \quad Y = \sum_{h=1}^r Y_h$	
Media ponderada de la población estratificada	
$\bar{Y} = \frac{1}{N} \sum_{h=1}^r \sum_{i=1}^{N_h} y_{hi} \quad \text{ó} \quad \bar{Y} = \sum_{h=1}^r W_h Y_h$	

<p>Total de la muestra estratificada</p> $y = \sum_{h=1}^r \sum_{i=1}^{n_h} y_{hi} \quad \text{ó} \quad y = \sum_{h=1}^r y_h \quad \text{ó} \quad y = \sum_{h=1}^r n_h \bar{y}_h$
<p>Media ponderada de la muestra estratificada</p> $\bar{y} = \frac{1}{n} \sum_{h=1}^r \sum_{i=1}^{n_h} y_{hi} \quad \text{ó} \quad \bar{y} = \frac{1}{n} \sum_{h=1}^r y_h \quad \text{ó} \quad \bar{y} = \sum_{h=1}^r w_h \bar{y}_h$

Se empezará con el caso del estimador de  $\bar{Y}$ . Un estimador para  $\bar{Y}$  es la suma ponderada de los estimadores de cada  $\bar{Y}_h$ . Ya que los estratos son independientes, el estimador de la media por estrato es análogo al estimador de la media poblacional en MAS:

$$\bar{Y} = \sum_{h=1}^r W_h \bar{Y}_h = W_1 \bar{Y}_1 + W_2 \bar{Y}_2 + \dots + W_r \bar{Y}_r .$$

El estimador de la media por estrato es:

$$\hat{\bar{Y}}_h = \bar{y}_h .$$

y ya que:

$$E(\bar{y}) = \bar{Y} .$$

Se puede concluir que:

$$E(\bar{y}_h) = \bar{Y}_h .$$



Entonces, la media muestral por estrato es un estimador insesgado de la media poblacional por estrato.

Si se define:

$$\bar{y}_{est} = \sum_{h=1}^r W_h \bar{y}_h = W_1 \bar{y}_1 + W_2 \bar{y}_2 + \dots + W_r \bar{y}_r ,$$

donde  $\bar{y}_{est}$  es un estimador insesgado de  $\bar{Y}$ ,

$$\hat{Y} = \bar{y}_{est} ,$$

y en términos de valor esperado queda:

$$\begin{aligned} E(\hat{Y}) &= E(\bar{y}_{est}) \\ &= E\left(\sum_{h=1}^r W_h \bar{y}_h\right), \end{aligned}$$

por las propiedades de la esperanza:

$$\begin{aligned} E(\bar{y}_{est}) &= \sum_{h=1}^r W_h E(\bar{y}_h) \\ &= \sum_{h=1}^r W_h \bar{Y}_h = \bar{Y} . \end{aligned}$$

Por lo tanto, si cada estrato es independiente, entonces:

$$E(\bar{y}_{est}) = \bar{Y} ,$$

$\hat{Y}_h = \bar{y}_h$  es un estimador insesgado de  $\bar{Y}_h$ , y  $\hat{Y} = \bar{y}_{est}$  es un estimador insesgado de  $\bar{Y}$ .

Como un estimador insesgado de la media poblacional  $\bar{Y}$ , la varianza de  $\bar{y}_{est}$  es:

$$\sigma^2(\bar{y}_{est}) = \sigma^2\left(\sum_{h=1}^r W_h \bar{y}_h\right).$$

Por las propiedades de la varianza:

$$\sigma^2(\bar{y}_{est}) = \sum_{h=1}^r W_h^2 \sigma^2(\bar{y}_h),$$

y haciendo uso de la definición de  $\sigma^2(\bar{y})$  del el MAS en la ecuación anterior:

$$\sigma^2(\bar{y}_{est}) = \sum_{h=1}^r W_h^2 \frac{\sigma_h^2}{n_h} (1 - f_h),$$

y desarrollando los términos de  $W_h$  y  $f_h$ , para obtener  $\sigma^2(\bar{y}_{est})$  con la CPF fijada:

$$\sigma^2(\bar{y}_{est}) = \frac{1}{N^2} \sum_{h=1}^r N_h (N_h - n_h) \frac{\sigma_h^2}{n_h}$$

De la ecuación anterior, el error estándar de  $\bar{y}_{est}$  es:

$$\sigma(\bar{y}_{est}) = \frac{1}{N} \sqrt{\sum_{h=1}^r N_h (N_h - n_h) \frac{\sigma_h^2}{n_h}}$$

El estimador  $\bar{y}_{est}$  es un estimador consistente del parámetro poblacional  $\bar{Y}$  si cumple con las condiciones suficientes para la consistencia de un estimador:

i)  $\bar{y}_{est}$  es insesgado de  $\bar{Y}$ ,

ii)  $\sigma^2(\bar{y}_{est}) \rightarrow 0$  cuando  $n \rightarrow \infty$ .

La primer condición se cumple; para la segunda:

$$\sigma^2(\bar{y}_{est}) \rightarrow 0 \text{ cuando } n_h \rightarrow \infty \quad \forall n_h,$$

entonces:

$$\lim_{n_h \rightarrow \infty} \sigma^2(\bar{y}_{est}) = \lim_{n_h \rightarrow \infty} W_h^2 \frac{\sigma_h^2}{n_h} (1 - f_h)$$

Ya que:

$$\lim_{n_h \rightarrow \infty} \frac{1}{n_h} = 0,$$

Entonces:

$$\lim_{n_h \rightarrow \infty} \sigma^2(\bar{y}_{est}) = 0.$$

Lo mismo para  $\bar{y}_h$ , por lo tanto:  $\bar{y}_{est}$  es un estimador consistente del parámetro  $\bar{Y}$  y  $\bar{y}_h$  es un estimador consistente del parámetro  $\bar{Y}_h$ .

Ahora bien, el estimador del total poblacional de  $Y$ ,

$$Y = \sum_{h=1}^r \sum_{i=1}^{N_h} y_{hi} = \sum_{h=1}^{N_h} Y_h = Y_1 + Y_2 + \dots + Y_r,$$

es la suma de los estimadores de  $Y_h$ ,  $\forall h = \overline{1, r}$ , ya que los estratos son independientes.

$$\hat{Y} = \sum_{h=1}^r \hat{Y}_h = \hat{Y}_1 + \hat{Y}_2 + \dots + \hat{Y}_r$$

De la independencia por estrato y en analogía a la población de MAS con cada  $\Omega_h$ ,

$$\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} y_{hi} = \frac{Y_h}{N_h}.$$

y se tiene que:

$$Y_h = N_h \bar{Y}_h.$$

El estimador insesgado y consistente de  $\bar{Y}_h$  es  $\hat{Y}_h = \bar{y}_h$ . Entonces:

$$\hat{Y}_h = N_h \hat{\bar{Y}}_h = N_h \bar{y}_h,$$

dado que:

$$E(\hat{Y}_h) = E[N_h \bar{y}_h]$$

$$E(\hat{Y}_h) = N_h E[\bar{y}_h] = N_h Y_h.$$

Por lo tanto:

$$\hat{Y} = \sum_{h=1}^L \hat{Y}_h = \sum_{h=1}^L N_h \bar{y}_h = N \sum_{h=1}^L W_h \bar{y}_h, \text{ ó}$$

$$\hat{Y} = N \bar{y}_{est}.$$

$\hat{Y} = N \bar{y}_{est}$  es un estimador insesgado de  $Y$ , donde:

$$E(\hat{Y}) = N E(\bar{y}_{est}),$$

dado que  $\bar{y}_{est}$  es un estimador insesgado de  $\bar{Y}$ , entonces:

$$E(\hat{Y}) = N \bar{Y} = Y.$$

La varianza de  $\hat{Y} = N \bar{y}_{est}$  es:

$$\sigma^2(\hat{Y}) = \sigma^2(N \bar{y}_{est}) = N^2 \sigma^2(\bar{y}_{est}).$$

Dado que :

$$\sigma^2(\bar{y}_{est}) = \sum_{h=1}^L W_h^2 \frac{\sigma_h^2}{n_h} (1 - f_h)$$

Sustituyendo la ecuación anterior en  $\sigma^2(\hat{Y})$ , se tiene:

$$\sigma^2(\hat{Y}) = N^2 \sum_{h=1}^r W_h^2 \frac{\sigma_h^2}{n_h} (1 - f_h) ,$$

si se desarrollan los términos  $W_h$  y  $f_h$ , y se eliminan los términos comunes, la ecuación anterior es:

$$\sigma^2(\hat{Y}) = \sum_{h=1}^r N_h (N_h - n_h) \frac{\sigma_h^2}{n_h} .$$

El error estándar de  $\hat{Y}$  es:

$$\sigma(\hat{Y}) = \sqrt{\sum_{h=1}^r N_h (N_h - n_h) \frac{\sigma_h^2}{n_h}} .$$

El estimador  $\hat{Y}$  es consistente del parámetro poblacional  $Y$  si cumple con las condiciones suficientes para la consistencia del estimador:

- i)  $\hat{Y}$  es insesgado de  $Y$ ,
- ii)  $\sigma^2(\hat{Y}) \rightarrow 0$  cuando  $n \rightarrow \infty$ .

La primer condición se cumple; por lo que se refiere a la segunda, se tiene:

$$\sigma^2(\hat{Y}) = N^2 \sigma^2(\bar{y}_{est}) ,$$

$$\lim_{n \rightarrow \infty} \sigma^2(\hat{Y}) = \lim_{n \rightarrow \infty} N^2 \sigma^2(\bar{y}_{est}) = N^2 \lim_{n \rightarrow \infty} \sigma^2(\bar{y}_{est}) .$$

Se demostró anteriormente que:

$$\lim_{n \rightarrow \infty} \sigma^2(\bar{y}_{est}) = 0 ,$$

e implica que:

$$\lim_{n \rightarrow \infty} \sigma^2(\hat{Y}) = 0.$$

Por lo tanto, se cumple la segunda condición y  $\hat{Y} = N\bar{y}_{cv}$  es un estimador consistente del parámetro  $Y$ .

### 3.3. Asignación del tamaño de muestra para estimar $Y$

Para estimar el total poblacional  $Y$  mediante el total muestral  $\hat{Y} = N\bar{y}_{cv}$ , con un error admisible  $\epsilon$  y un nivel de confianza  $(1 - \alpha)$ , se recurre nuevamente al intervalo de confianza del estimador:

$$P(-k\sigma(\hat{Y}) < \hat{Y} - Y < k\sigma(\hat{Y})) = 1 - \alpha,$$

$$P(|\hat{Y} - Y| < k\sigma(\hat{Y})) = 1 - \alpha.$$

El error admisible es:

$$\epsilon = |\hat{Y} - Y| = k\sigma(\hat{Y}),$$

$$\epsilon^2 = k^2 \sigma^2(\hat{Y}).$$

Además,

$$\sigma^2(\hat{Y}) = \sum_{h=1}^L N_h (N_h - n_h) \frac{\sigma_h^2}{n_h}.$$

Sustituyendo en  $e^2$ :

$$e^2 = k^2 \sum_{h=1}^k N_h \left( \frac{N_h}{n_h} \right) \frac{\sigma_h^2}{n_h}$$

En el MAE no sólo hay que calcular el tamaño total de muestra  $n$ , también se presenta la situación de repartir la muestra en los  $k$  estratos. A esta repartición se le llama asignación del tamaño de muestra.

Hay dos procedimientos comunes de asignación de la muestra: asignación proporcional y asignación óptima.

La asignación proporcional es el reparto de la muestra de manera que cada uno de los estratos proporcionalmente al tamaño de estos en total.

$$n_h = n \frac{N_h}{N}$$

donde

$$N = \sum_{h=1}^k N_h$$

Tanto  $n$  es el tamaño total de muestra y  $n_h$  el tamaño de muestra en el estrato  $h$ .

Entonces

$$\frac{n_h}{N_h} = \frac{n}{N}$$

Entonces

$$n_h = N_h \frac{n}{N}$$

Este es el método de asignación proporcional. El tamaño de muestra en cada uno de los estratos es proporcional al tamaño de cada uno de ellos.

Entonces

$$n_h = n \frac{N_h}{N}$$

Sustituyendo en  $\epsilon^2$ :

$$\epsilon^2 = k^2 \sum_{h=1}^r N_h (N_h - n_h) \frac{\sigma_h^2}{n_h}.$$

En el MAE no sólo hay que calcular el tamaño total de muestra  $n$ , también se presenta la situación de repartir la muestra en los  $r$  estratos. A esta repartición se le llama asignación del tamaño de muestra.

Hay dos procedimientos comunes de asignación de muestra: asignación proporcional y asignación óptima.

La **asignación proporcional** es el reparto de la muestra de tamaño  $n$  entre los estratos proporcionalmente al tamaño de éstos; es decir,

$$n_h \propto N_h,$$

o bien:

$$n_h = a N_h \quad \text{y} \quad n = a N.$$

Donde  $a$  es la proporción entre la muestra y la población. Al igualar estas dos ecuaciones:

$$\frac{n_h}{n} = \frac{N_h}{N} \quad \Rightarrow \quad w_h = W_h,$$

entonces:

$$n_h = \frac{n}{N} N_h.$$

Cuando la asignación es proporcional, el estimador de  $Y$ ,  $\hat{Y} = N \bar{y}_{est}$ , se calcula con la ecuación:

$$\hat{Y} = N \sum_{h=1}^r W_h \bar{y}_h.$$



o bien, se sustituye  $W_h = w_h$ ,

$$\begin{aligned}\hat{Y}_{prop} &= N \sum_{h=1}^r w_h \bar{y}_h \\ &= \frac{N}{n} \sum_{h=1}^r n_h \bar{y}_h = \frac{N}{n} y,\end{aligned}$$

donde  $\hat{Y}_{prop}$  denota al estimador de  $Y$  en el muestreo estratificado con asignación proporcional.

Se dice que  $\hat{Y}_{prop}$  es un estimador autoponderado de  $Y$ , donde todas las unidades tienen la misma probabilidad de ser seleccionadas y el total poblacional se estima multiplicando el total muestral  $\sum_{h=1}^r n_h \bar{y}_h$  por el factor de expansión  $\frac{N}{n}$ .

La varianza de  $\hat{Y}_{prop}$  se obtiene de:

$$\sigma^2(\hat{Y}) = \sum_{h=1}^r N_h (N_h - n_h) \frac{\sigma_h^2}{n_h} \quad \text{ó} \quad \sigma^2(\hat{Y}) = \sum_{h=1}^r N_h^2 \frac{\sigma_h^2}{n_h} (1 - f_h),$$

sustituyendo  $n_h = \frac{n}{N} N_h$ , se tiene que:

$$\sigma^2(\hat{Y}_{prop}) = \sum_{h=1}^r N_h^2 \frac{\sigma_h^2}{\frac{n}{N} N_h} (1 - f_h).$$

si se sustituye también  $f_h = f = \frac{n}{N}$ :

$$\begin{aligned}\sigma^2(\hat{Y}_{prop}) &= \frac{N}{n} \sum_{h=1}^r N_h \sigma_h^2 \left( \frac{N - n}{N} \right) \\ &= \sum_{h=1}^r N_h \sigma_h^2 \left( \frac{N - n}{n} \right)\end{aligned}$$

$$= \left( \frac{N}{n} - 1 \right) \sum_{h=1}^r N_h \sigma_h^2.$$

De:

$$\varepsilon^2 = k^2 \sigma^2(\hat{Y}),$$

y sustituyendo  $\sigma^2(\hat{Y})$  por  $\sigma^2(\hat{Y}_{prop})$ , se obtiene:

$$\varepsilon^2 = k^2 \left( \frac{N}{n} - 1 \right) \sum_{h=1}^r N_h \sigma_h^2,$$

si se despeja para calcular  $n$ :

$$\frac{\varepsilon^2}{k^2} = \frac{N}{n} \sum_{h=1}^r N_h \sigma_h^2 - \sum_{h=1}^r N_h \sigma_h^2,$$

$$\frac{N}{n} \sum_{h=1}^r N_h \sigma_h^2 = \frac{\varepsilon^2}{k^2} + \sum_{h=1}^r N_h \sigma_h^2,$$

$$\frac{1}{n} = \frac{\frac{\varepsilon^2}{k^2} + \sum_{h=1}^r N_h \sigma_h^2}{N \sum_{h=1}^r N_h \sigma_h^2}.$$

Por lo tanto, el tamaño de muestra total y su asignación proporcional se calculan con la fórmula:

$$n = \frac{N \sum_{h=1}^r N_h \sigma_h^2}{\frac{\varepsilon^2}{k^2} + \sum_{h=1}^r N_h \sigma_h^2} \quad \text{y} \quad n_h = \frac{n}{N} N_h.$$

La asignación óptima, llamada también asignación de Neyman o de Chuprow-Neyman, es aquella donde el tamaño de la muestra para cada estrato  $n_h$  es proporcional al producto del tamaño de este estrato  $N_h$  por su variabilidad  $\sigma_h^2$ ,

$$n_h \propto N_h \sigma_h.$$

Se trata de establecer qué valores de  $n_h$  hacen mínima la expresión:

$$\sigma^2(\hat{y}) = \sum_{h=1}^r N_h (N_h - n_h) \frac{\sigma_h^2}{n_h},$$

sujeto a la condición:

$$\sum_{h=1}^r n_h = n.$$

Se puede aplicar el método de los multiplicadores de Lagrange<sup>1</sup> a la expresión

$\sigma^2(\hat{y})$  condicionada a  $\sum_{h=1}^r n_h = n$ :

$$\begin{aligned} \Psi &= \sum_{h=1}^r N_h (N_h - n_h) \frac{\sigma_h^2}{n_h} + \lambda \left( \sum_{h=1}^r n_h - n \right), \\ \Psi &= \frac{N_1(N_1 - n_1)}{n_1} \sigma_1^2 + \frac{N_2(N_2 - n_2)}{n_2} \sigma_2^2 + \dots + \\ &+ \frac{N_r(N_r - n_r)}{n_r} \sigma_r^2 + \lambda (n_1 + n_2 + \dots + n_r - n). \end{aligned}$$

Se deriva  $\Psi$  con respecto a  $n_1, n_2, \dots, n_r$  y se iguala a 0, lo que permite obtener el sistema de ecuaciones:

---

<sup>1</sup> AZORIN, Francisco: *Curso de muestreo y aplicaciones*, México, Aguilar, pp. 117 a 119.

MUESTREO ALEATORIO ESTRATIFICADO

$$\frac{\partial P}{\partial n_1} = - \frac{N_1'}{n_1^2} \sigma_1^2 + \lambda = 0$$

...

$$\frac{\partial P}{\partial n_r} = - \frac{N_r'}{n_r^2} \sigma_r^2 + \lambda = 0$$

Al despejar  $\lambda$ :

$$\frac{N_1' \sigma_1^2}{n_1^2} = \lambda \quad \frac{N_r' \sigma_r^2}{n_r^2} = \lambda$$

o bien entonces que:

$$\frac{N_1' \sigma_1^2}{n_1^2} = \frac{N_r' \sigma_r^2}{n_r^2}$$

$$\frac{n_1}{n_r} = \sqrt{\frac{N_1' \sigma_1^2}{N_r' \sigma_r^2}}$$

Por lo tanto, las muestras se reparten en función de la raíz cuadrada de la proporción de la muestra y la desviación estándar.

o bien entonces que:

Por lo tanto, las muestras se reparten en función de la raíz cuadrada de la proporción de la muestra y la desviación estándar.

o bien entonces que:

$$\frac{\partial \Psi}{\partial n_1} = - \frac{N_1^2}{n_1^2} \sigma_1^2 + \lambda = 0 ,$$

...

$$\frac{\partial \Psi}{\partial n_r} = - \frac{N_r^2}{n_r^2} \sigma_r^2 + \lambda = 0 .$$

Al despejar  $\lambda$ :

$$\frac{N_1 \sigma_1}{n_1} = \sqrt{\lambda} \quad \dots \quad \frac{N_r \sigma_r}{n_r} = \sqrt{\lambda} ,$$

lo cual implica que:

$$\frac{N_1 \sigma_1}{n_1} = \frac{N_2 \sigma_2}{n_2} = \dots = \frac{N_r \sigma_r}{n_r} , \quad \text{ó}$$

$$\frac{n_1}{N_1 \sigma_1} = \frac{n_2}{N_2 \sigma_2} = \dots = \frac{n_r}{N_r \sigma_r} = \frac{n}{\sum_{h=1}^r N_h \sigma_h}$$

Ya que son igualdades y la ecuación de condición es  $\sum_{h=1}^r n_h = n$ , entonces $\sum_{h=1}^r N_h \sigma_h$  también es válido.

Por lo tanto, la asignación óptima de la muestra en los estratos está dada por:

$$n_h = \frac{n N_h \sigma_h}{\sum_{h=1}^r N_h \sigma_h}$$

La varianza de  $\hat{Y} = N \bar{y}_{st}$  es:

$$\sigma^2(\hat{Y}) = \sum_{h=1}^r N_h (N_h - n_h) \frac{\sigma_h^2}{n_h} = \sum_{h=1}^r N_h^2 \frac{\sigma_h^2}{n_h} (1 - f_h) =$$

$$= \sum_{h=1}^r \frac{N_h^2}{n_h} \sigma_h^2 - \sum_{h=1}^r N_h \sigma_h^2.$$

Cuando la asignación es óptima, entonces se sustituye  $n_h = \frac{n N_h \sigma_h}{\sum_{h=1}^r N_h \sigma_h}$  en la ecuación anterior:

$$\begin{aligned} \sigma^2(\hat{Y}_{op}) &= \sum_{h=1}^r \frac{N_h^2}{\frac{n N_h \sigma_h}{\sum_{h=1}^r N_h \sigma_h}} \sigma_h^2 - \sum_{h=1}^r N_h \sigma_h^2 = \\ &= \sum_{h=1}^r \frac{N_h^2 \left( \sum_{h=1}^r N_h \sigma_h \right)}{n N_h \sigma_h} \sigma_h^2 - \sum_{h=1}^r N_h \sigma_h^2. \end{aligned}$$

Al eliminar el término cuadrado de  $N_h$  y  $\sigma_h^2$  con los respectivos términos del denominador para el primer factor del lado derecho de la ecuación se obtiene:

$$\sigma^2(\hat{Y}_{op}) = \frac{\sum_{h=1}^r N_h \sigma_h \left( \sum_{h=1}^r N_h \sigma_h \right)}{n} - \sum_{h=1}^r N_h \sigma_h^2,$$

y si se asocian los términos de suma, queda la ecuación para calcular la varianza de  $\hat{Y}_{op}$ :

$$\sigma^2(\hat{Y}_{op}) = \frac{\left( \sum_{h=1}^r N_h \sigma_h \right)^2}{n} - \sum_{h=1}^r N_h \sigma_h^2,$$

para poner a  $\sigma^2(\hat{Y}_{op})$  en términos de la CPF, se sustituye  $n_h = \frac{n N_h \sigma_h}{\sum_{h=1}^r N_h \sigma_h}$  en la

ecuación de varianza de  $\hat{Y}$  que contiene este término:

$$\sigma^2(\hat{Y}) = \sum_{h=1}^r N_h^2 \frac{\sigma_h^2}{n_h} (1 - f_h)$$

sustituyendo:

$$\sigma^2(\hat{Y}_{op}) = \sum_{h=1}^r N_h^2 \frac{\frac{\sigma_h^2}{n N_h \sigma_h}}{\sum_{h=1}^r N_h \sigma_h} (1 - f_h)$$

Al cambiar de lugar algunos términos:

$$\sigma^2(\hat{Y}_{op}) = \sum_{h=1}^r N_h^2 \frac{\sigma_h^2 \sum_{h=1}^r N_h \sigma_h}{n N_h \sigma_h} (1 - f_h);$$

y después de eliminar términos comunes, se llega a que:

$$\sigma^2(\hat{Y}_{op}) = \frac{1}{n} \sum_{h=1}^r N_h \sigma_h \sum_{h=1}^r N_h \sigma_h (1 - f_h).$$

No se pueden combinar los términos de suma en un término cuadrado porque la fracción de muestreo también depende de la suma. Por lo tanto,  $\sigma^2(\hat{Y}_{op})$  también queda expresado por la ecuación anterior.

Ahora, para calcular el tamaño de muestra se recurre a la ecuación que relaciona a la precisión y el nivel de confianza:

$$e^2 = k^2 \sigma^2(\hat{Y}),$$

y sustituyendo  $\sigma^2(\hat{Y})$  por  $\sigma^2(\hat{Y}_{op})$ :

$$e^2 = k^2 \sigma^2(\hat{Y}_{op}).$$

Si se sustituye para la varianza de los términos desarrollados anteriormente

$$E^2 = k^2 \left[ \frac{\left( \sum_{h=1}^r N_h \sigma_h \right)^2}{n} - \sum_{h=1}^r N_h \sigma_h^2 \right]$$

y se despeja el lado derecho de la ecuación para  $n$ :

$$\frac{1}{n} = \frac{\frac{E^2}{k^2} + \sum_{h=1}^r N_h \sigma_h^2}{\left( \sum_{h=1}^r N_h \sigma_h \right)^2}$$

Se puede concluir que el tamaño de muestra total y su asignación óptima se calculan como:

$$n = \frac{\left( \sum_{h=1}^r N_h \sigma_h \right)^2}{\frac{E^2}{k^2} + \sum_{h=1}^r N_h \sigma_h^2} \quad \text{y} \quad n_h = \frac{n N_h \sigma_h}{\sum_{h=1}^r N_h \sigma_h}$$

En general, para calcular  $n$  se necesita conocer:

- El nivel de confianza  $(1 - \alpha)$  y el error máximo admisible  $e$ .
- Los tamaños de los estratos  $N_1, N_2, \dots, N_r$  que se desean muestrear.
- La varianza de cada estrato, representada por  $\sigma_h$ .

Un problema de especial interés es la determinación de su control o cuánto que pertenecen al llamado estrato de inclusión forzosa o estrato cierto, es el que se muestra en todo por necesidad. En el llamado estrato de exclusión forzosa o estrato imposible de



Si se sustituye para la varianza de los términos desarrollados anteriormente:

$$\varepsilon^2 = k^2 \left[ \frac{\left( \sum_{h=1}^r N_h \sigma_h \right)^2}{n} - \sum_{h=1}^r N_h \sigma_h^2 \right],$$

y se despeja el lado derecho de la ecuación para  $n$ :

$$\frac{1}{n} = \frac{\frac{\varepsilon^2}{k^2} + \sum_{h=1}^r N_h \sigma_h^2}{\left( \sum_{h=1}^r N_h \sigma_h \right)^2}.$$

Se puede concluir que el tamaño de muestra total y su asignación óptima se calculan como:

$$n = \frac{\left( \sum_{h=1}^r N_h \sigma_h \right)^2}{\frac{\varepsilon^2}{k^2} + \sum_{h=1}^r N_h \sigma_h^2} \quad \text{y} \quad n_h = \frac{n N_h \sigma_h}{\sum_{h=1}^r N_h \sigma_h}$$

En general, para calcular  $n$  se necesita conocer:

- El nivel de confianza  $(1 - \alpha)$ , y el error máximo admisible  $\varepsilon$ .
- Los tamaños de los estratos  $N_1, N_2, N_3, \dots, N_r$ , que se suponen conocidos.
- La varianza de cada estrato, representada por  $\sigma_h^2$ .

"Un problema de especial interés es la determinación de los puntos o valores que definen al llamado estrato de inclusión forzosa o **estrato cierto**, en el que la selección se hace con probabilidad 1; y el llamado estrato de exclusión forzosa o **estrato imposible**, en

el que la selección se hace con probabilidad  $0^{n-2}$ . En el primero, el tamaño de muestra será del 100% de la población del estrato, ya que todos los elementos son incluidos. Y en el de exclusión forzosa será de 0.

### 3.4. Una comparación entre MAE y MAS

Una de las ventajas que se tienen al estratificar es la ganancia en precisión o disminución de la varianza. Una comparación entre los dos métodos es examinar la varianza del estimador  $\hat{Y}$  de MAS y de MAE con asignación proporcional y óptima.

Los estimadores de  $Y$  en MAS y MAE son:

$\hat{Y} = N\bar{y}$	En MAS
$\hat{Y} = N\bar{y}_{ev}$	En MAE
$\hat{Y}_{prop} = N \sum_{h=1}^L w_h \bar{y}_h$	En MAE con asignación proporcional
$\hat{Y}_{op} = N \sum_{h=1}^L W_h \bar{y}_h$	En MAE con asignación óptima

<sup>2</sup> AZORIN, F., 1971, *op. cit.*, p. 111.

Las varianzas, en términos de la CPF, de  $\hat{Y}$  de MAS y MAE son:

<p>En MAS</p> $\sigma^2(\hat{Y}) = N(N - n) \frac{\sigma^2}{n} = \frac{N^2 \sigma^2}{n} (1 - f_h)$
<p>En MAE con asignación proporcional</p> $\sigma^2(\hat{Y}_{prop}) = \frac{N}{n} \sum_{h=1}^r N_h \sigma_h^2 (1 - f_h)$
<p>En MAE con asignación óptima</p> $\sigma^2(\hat{Y}_{op}) = \frac{1}{n} \sum_{h=1}^r N_h \sigma_h \sum_{h=1}^r N_h \sigma_h (1 - f_h)$

En la práctica, la CPF puede ser ignorada siempre y cuando la fracción de muestreo no exceda un 5% y, para muchos propósitos prácticos, aún si es del 10%. El efecto de ignorar la CPF es sobrestimar la varianza del error estándar del estimador<sup>3</sup>.

Para hacer la comparación entre MAS y MAE, Cochran<sup>4</sup> hace uso del análisis de varianza. Supóngase que la CPF puede ser ignorada

Se parte de la siguiente igualdad de la suma de cuadrados:

$$\sum_{h=1}^r \sum_{i=1}^{N_h} (y_{hi} - \bar{Y})^2 = \sum_{h=1}^r \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2 + \sum_{h=1}^r N_h (\bar{Y}_h - \bar{Y})^2$$

<sup>3</sup> COCHRAN, William G: *Técnicas de muestreo*, México, Continental, p. 48.

<sup>4</sup> COCHRAN, W. G., 1961, *op. cit.*, p. 138 y AZORIN, F., 1971, *op. cit.*, p. 130.

De la varianza de  $Y$

$$N\sigma^2 = \sum_{k=1}^K \sum_{i=1}^{N_k} (y_{ki} - \bar{y})^2$$

sustituyendo en la igualdad anterior para tenerla en función de varianzas de los estratos

$$N\sigma^2 = \sum_{k=1}^K N_k \sigma_k^2 + \sum_{k=1}^K N_k (\bar{y}_k - \bar{y})^2$$

A esta igualdad se le conoce como la descomposición de varianzas; el primer término del lado derecho se le conoce como **intra-varianza** (varianza dentro de los estratos); y al segundo término, como **inter-varianza** (varianza entre los estratos).

Para comparar las varianzas de  $\hat{Y}$  y  $\hat{Y}_{prop}$  se multiplica la ecuación de descomposición de varianzas por  $\frac{N}{n}$  para que la unidad de la izquierda sea igual a la varianza en el MAS cuando la OPE es ignorada

$$\frac{N^2}{n} \sigma^2 = \frac{N}{n} \sum_{k=1}^K N_k \sigma_k^2 + \frac{N}{n} \sum_{k=1}^K N_k (\bar{y}_k - \bar{y})^2$$

Al sustituir la varianza de MAS y de OPE en **ecuación (1)** y **ecuación (2)** respectivamente en el lado izquierdo y el primer término de esta ecuación, respectivamente se tiene que:

$$\sigma^2 + \sigma^2_{prop} = \sum_{k=1}^K \frac{N_k}{n} \sigma_k^2 + \frac{N}{n} \sum_{k=1}^K N_k (\bar{y}_k - \bar{y})^2$$

Por lo que el segundo término de esta ecuación es:

$$\frac{N}{n} \sum_{k=1}^K N_k (\bar{y}_k - \bar{y})^2$$

De la varianza de  $\mathbf{Y}$ :

$$N \sigma^2 = \sum_{h=1}^L \sum_{i=1}^{N_h} (y_{hi} - \bar{Y})^2,$$

sustituyendo en la igualdad anterior para tenerla en función de varianzas, se tiene:

$$N \sigma^2 = \sum_{h=1}^L N_h \sigma_h^2 + \sum_{h=1}^L N_h (\bar{Y}_h - \bar{Y})^2.$$

A esta igualdad se le conoce como la descomposición de varianzas: al primer término del lado derecho se le conoce como *intravarianza* (varianza dentro de los estratos); y al segundo término, como *intervarianza* (varianza entre los estratos).

Para comparar las varianzas de  $\hat{Y}$  y  $\hat{Y}_{pp}$ , se multiplica la ecuación de descomposición de varianzas por  $\frac{N}{n}$  para que la entidad de la izquierda se iguale a la varianza en el MAS cuando la CPF es ignorada:

$$\frac{N^2}{n} \sigma^2 = \frac{N}{n} \sum_{h=1}^L N_h \sigma_h^2 + \frac{N}{n} \sum_{h=1}^L N_h (\bar{Y}_h - \bar{Y})^2.$$

Al sustituir la varianza del MAS y del MAE con asignación proporcional en el lado izquierdo y el primer término del lado derecho respectivamente, se tiene que:

$$\sigma^2(\hat{Y}) = \sigma^2(\hat{Y}_{pp}) + \frac{N}{n} \sum_{h=1}^L N_h (\bar{Y}_h - \bar{Y})^2.$$

Puesto que el segundo término del lado derecho es:

$$\frac{N}{n} \sum_{h=1}^L N_h (\bar{Y}_h - \bar{Y})^2 > 0,$$

entonces:

$$\sigma^2(\hat{Y}) \geq \sigma^2(\hat{Y}_{prop}).$$

Así queda la comparación entre la varianza del parámetro del total en MAS y MAE con asignación proporcional.

Lo siguiente es comparar la relación que existe en el MAE entre las varianzas de las dos asignaciones. En este caso, se parte de la diferencia de las respectivas varianzas,

$$\sigma^2(\hat{Y}_{prop}) - \sigma^2(\hat{Y}_{op}) = \frac{N}{n} \sum_{h=1}^r N_h \sigma_h^2 - \frac{\left( \sum_{h=1}^r N_h \sigma_h \right)^2}{n}.$$

Para simplificar esta ecuación se define el siguiente término:

$$\bar{\sigma} = \sum_{h=1}^r W_h \sigma_h ;$$

se sustituye en la diferencia de varianzas, se despejan<sup>5</sup> y se asocian los términos respectivos, para obtener:

$$\sigma^2(\hat{Y}_{prop}) - \sigma^2(\hat{Y}_{op}) = \frac{N}{n} \sum_{h=1}^r N_h (\sigma_h - \bar{\sigma})^2 ,$$

despejando al lado derecho la varianza  $\sigma^2(\hat{Y}_{op})$ :

$$\sigma^2(\hat{Y}_{prop}) = \sigma^2(\hat{Y}_{op}) + \frac{N}{n} \sum_{h=1}^r N_h (\sigma_h - \bar{\sigma})^2 ,$$

---

<sup>5</sup> COCHRAN, W. G., 1981, *op. cit.*, p. 141.

donde:

$$\frac{N}{n} \sum_{h=1}^L N_h (\sigma_h - \bar{\sigma})^2 > 0.$$

Por lo tanto:

$$\sigma^2(\hat{Y}_{ppp}) \geq \sigma^2(\hat{Y}_{dp}).$$

Para confirmar la relación entre MAS y MAE con asignación óptima, se sustituye  $\sigma^2(\hat{Y}_{ppp})$  en función de  $\sigma^2(\hat{Y}_{dp})$  en la igualdad de  $\sigma^2(\hat{Y})$ :

$$\sigma^2(\hat{Y}) = \left[ \sigma^2(\hat{Y}_{dp}) + \frac{N}{n} \sum_{h=1}^L N_h (\sigma_h - \bar{\sigma})^2 \right] + \frac{N}{n} \sum_{h=1}^L (\bar{Y}_h - \bar{Y})^2, \text{ ó}$$

$$\sigma^2(\hat{Y}) = \sigma^2(\hat{Y}_{dp}) + \frac{N}{n} \left[ \sum_{h=1}^L N_h (\sigma_h - \bar{\sigma})^2 + \sum_{h=1}^L (\bar{Y}_h - \bar{Y})^2 \right],$$

y el término:

$$\frac{N}{n} \left[ \sum_{h=1}^L N_h (\sigma_h - \bar{\sigma})^2 + \sum_{h=1}^L (\bar{Y}_h - \bar{Y})^2 \right] > 0.$$

Por lo que:

$$\sigma^2(\hat{Y}) \geq \sigma^2(\hat{Y}_{ppp}).$$

Así, la relación entre las tres varianzas está dada por:

$$\sigma^2(\hat{Y}) \geq \sigma^2(\hat{Y}_{ppp}) \geq \sigma^2(\hat{Y}_{dp}).$$

Esto sucede si se cumple la teoría básica de MAS y de MAE.

Es poco probable que en la práctica no se cumplan las desigualdades anteriores; y aunque es posible que ocurra, para comprobarlo se requiere aplicar los tres métodos al mismo estudio e inferir los resultados. Se ha demostrado teóricamente<sup>6</sup> que el MAE con asignación óptima tiene mayor precisión, después el MAE con asignación proporcional y luego el MAS.

---

<sup>6</sup> COCHRAN, William G.: *Comparison of Methods for Determining Stratum Boundaries*, Bulletin of International Statistical Institute, 32 (2), Tokyo, 1961, pp. 345-358.



## CAPITULO IV

# EL CRITERIO DE DALENIUS O REGLA $\text{cum}\sqrt{f(y)}$ PARA ENCONTRAR FRONTERAS DE ESTRATOS

### 4.1. Introducción

En la formación de estratos, las principales consideraciones a tener son la determinación de:

- La variable de estratificación
- El número de estratos
- La determinación de las fronteras de los estratos
- La asignación de la muestra en cada estrato

En el capítulo anterior se revisaron dos procedimientos de asignación de muestra en cada estrato; en el presente capítulo se mostrará la determinación del número de estratos y las fronteras de estratos sobre una variable de estratificación.

La consideración básica para la formación de estratos es que deben ser homogéneos internamente. Puede utilizarse cualquier conocimiento previo, como la intuición personal o de juicio para producir semejanza dentro de un mismo estrato, pero

entre los métodos matemáticos para seleccionar las mejores fronteras para los estratos<sup>1</sup> está el denominado de **estratificación óptima**, por Tore Dalenius, en varios trabajos durante la década de los cincuenta.

En 1950, Dalenius<sup>2</sup> planteó los supuestos para formar fronteras bajo la situación ideal de trabajar con una población teórica, y encontró la solución exacta a la frontera óptima mediante procesos iterativos cuando es asignación proporcional y asignación óptima respectivamente. En 1951, contempló<sup>3</sup>, entre otros aspectos, la estratificación óptima cuando se realiza sobre una variable de estratificación específica y no sobre la variable de estimación. En 1959, dio énfasis<sup>4</sup> al planteamiento de la solución exacta para determinar las fronteras de los estratos y ofrecería una aproximación a la solución para situaciones prácticas.

## 4.2. La variable de estratificación

Para encontrar las fronteras de los estratos se necesita el conocimiento previo de la población  $\Omega = \{\omega_1, \dots, \omega_j, \dots, \omega_N\}$ , donde  $Y$  es la variable asociada  $Y(\omega_j) = y_j$ , a esa población, llamada **variable de estimación**. En la situación ideal se dispone de la función

---

<sup>1</sup> KISH, Leslie: *Muestreo de encuestas*, México, Trillas, 1979, p. 133.

<sup>2</sup> DALENIUS, Tore: *The problem of optimum stratification*, E.U., Scandinavian Journal of Statistics (33), 1950, pp. 203-213.

<sup>3</sup> DALENIUS, Tore y GURNEY, M.: *The problem of optimum stratification II*, E.U., Scandinavian Journal of Statistics (34), 1951, pp. 133-148.

<sup>4</sup> DALENIUS, T. y HODGES, S. P.: *Minimum variance stratification*, Journal American Statistics Association, V. 54, 1959, pp. 88-101.

de probabilidad de  $Y$ ,  $f(y)$ ; entonces, se forman los estratos estableciendo las fronteras en  $f(y)$ .

Si se carece de la información de  $Y$ , que es lo sucede en la práctica, entonces se usa alguna otra variable  $X$ , por ejemplo: los valores de  $Y$  en algún censo reciente o una variable altamente correlacionada a ésta.

Se asume entonces que la variable de estimación  $Y$  y la variable de estratificación  $X$  están relacionadas por un modelo<sup>5</sup> de regresión de la forma:

$$Y = \alpha + \beta X + \eta ,$$

donde:

- $\eta$  es la desviación aleatoria de  $Y$  a la recta de regresión  $E(Y) = \alpha + \beta X$ ,
- $\beta$  y  $\eta$  no están correlacionadas entre sí,
- $E(\eta) = 0 \quad \forall X$ ,
- $E(\eta^2) = \gamma \quad \forall X$ .

### 4.3. El número de estratos

La cuestión sobre cuál debe ser el número de estratos  $r$  puede ser resuelta si se calcula y compara la intravarianza e intervarianza de  $Y$  por el efecto de adición de otro estrato.

---

<sup>5</sup> DALENIUS, Tore y GURNEY, M., 1951, *op. cit.*, p. 134 y sig.

Nuevamente, si se considera la descomposición de la varianza para estimar  $\bar{Y}$  se obtiene que<sup>6</sup>:

$$N\sigma^2 = \sum_{h=1}^r N_h \sigma_h^2 + \sum_{h=1}^r N_h (\bar{Y}_h - \bar{Y})^2.$$

Cuando la intravarianza sea un mínimo y la intervarianza un máximo, se habrá logrado hacer estratos homogéneos internamente. Puede demostrarse que en ese punto es donde el llamado número de estratos óptimo  $r_{op}$  alcanza una mínima varianza para el estimador. Esto se logra mediante un proceso iterativo de  $r$ .

El proceso iterativo es el cálculo repetido para cambios de  $r$ . El rango de valores de  $r$  cabe en dos alternativas: Cuando la condición de los estratos es tener cuando menos dos unidades en cada uno por restricción de varianza; es decir, cuando  $N_h \geq 2$ , el rango de  $r$  es:

$$1 \leq r \leq \frac{N}{2},$$

donde  $r$  es el máximo entero.

Cuando no existe la condición mínima de  $N_h \geq 2$ , y la varianza por estrato se calcula por otros métodos<sup>7</sup>, el rango de  $r$  es:

$$1 \leq r \leq N.$$

Un aumento en el número de estratos más allá del óptimo implica trabajo extra en la planificación y extracción de la muestra.

---

<sup>6</sup> COCHRAN, William G.: *Técnicas de muestreo*, México, Continental, 1971, p. 141.

<sup>7</sup> *Idem.*, p. 190.

Se reconoce<sup>8</sup> también que, el duplicar el número de estratos más allá de un número razonable, no resulta en una reducción proporcional de la varianza cuando la estratificación para la estimación de  $Y$  se hace con base en una variable de estratificación  $X$ .

Cuando la asignación es óptima, con la regresión lineal de la variable de estimación  $Y = \alpha + \beta X + \eta$ , existe un modelo<sup>9</sup> para medir el efecto en la varianza del estimador al aumentar el número  $r$  de estratos. Este modelo se representa por:

$$\frac{\rho^2}{r^2} + (1 - \rho^2),$$

donde  $\rho$  es el coeficiente de correlación entre  $X$  y  $Y$ .

La parte  $\left(\frac{\rho^2}{r^2}\right)$  disminuye con el cuadrado del número de estratos, pero  $(1 - \rho^2)$  no se afecta por este aumento. Así, la varianza del estimador se aproxima a cierto nivel después de la creación de un número moderado de estratos. Es decir, por más estratos que se aumenten ya no disminuirá considerablemente la varianza del estimador.

Cochran<sup>10</sup> probó el modelo  $\frac{\rho^2}{r^2} + (1 - \rho^2)$  para una función de distribución rectangular con asignación de la forma  $n_h = \frac{n}{r}$  (llamada asignación igual) y el modelo de regresión  $Y = \alpha + \beta X + \eta$ ; los resultados indicaron que, a menos que el coeficiente de correlación exceda de 0.95, debe esperarse poca reducción en la varianza del estimador cuando  $r > 6$  estratos.

---

<sup>8</sup> *Idem.*, pp. 181 y sig.

<sup>9</sup> KISH, Leslie: *Muestreo de encuestas*, México, Trillas, 1979, p. 131.

<sup>10</sup> COCHRAN, William G. 1976, *op. cit.* p. 182.

#### 4.4. Fronteras de los estratos. Regla $\text{cum}\sqrt{f(y)}$

##### 4.4.1. Supuestos

Aunque muchas aplicaciones de muestreo estratificado se definen desde una población finita  $\Omega = \{\omega_1, \dots, \omega_r, \dots, \omega_N\}$ , es conveniente para propósitos de investigación teórica <sup>11</sup> distribuir una población hipotética  $\Pi$  representada por una función de densidad  $f(y)$ , donde  $a \leq y \leq b$  y  $Y$  es la variable asociada a esa población hipotética.

Con el fin de formar  $r$  estratos, se dejan  $y_0 = a$  y  $y_r = b$  como el valor menor y mayor de  $Y$  respectivamente, y a  $y_1, y_2, \dots, y_{r-1}$  como las fronteras intermedias entre los estratos, es decir, donde la variable  $Y$  va a ser cortada.

La media y la varianza de  $f(y)$  son:

$$\mu = \int_a^b y f(y) dy = \sum_{h=1}^r W_h \mu_h, \text{ y}$$

$$\sigma^2 = \int_a^b y^2 f(y) dy - \mu^2.$$

La ponderación o frecuencia relativa  $W_h$  y la media  $\mu_h$  para cada estrato son<sup>12</sup>:

$$W_h = \int_{y_{h-1}}^{y_h} f(y) dy, \text{ y}$$

<sup>11</sup> DALENIUS, Tore. 1950, *op. cit.*, p. 203.

<sup>12</sup> DALENIUS, T. y HODGES, S. P., 1959, *op. cit.*, p. 89.

$$\frac{\partial W_h}{\partial y_h} = f(y_h),$$

$$W_h \mu_h = \int_{y_{h-1}}^{y_h} f(t) dt.$$

Nótese que al pasar a  $W_h$  al lado derecho de la ecuación para dejarla en función de  $\mu_h$  se obtiene que:

$$\mu_h = \frac{\int_{y_{h-1}}^{y_h} f(t) dt}{\int_{y_{h-1}}^{y_h} f(t) dt}.$$

La varianza de  $\mathbf{Y}$  en el estrato  $h$ ,  $\sigma_h^2$ , es:

$$W_h \sigma_h^2 = \int_{y_{h-1}}^{y_h} t^2 f(t) dt - W_h \mu_h^2,$$

despejando a  $W_h$  al lado izquierdo de la ecuación anterior:

$$\sigma_h^2 = \frac{\int_{y_{h-1}}^{y_h} t^2 f(t) dt}{\int_{y_{h-1}}^{y_h} f(t) dt} - \mu_h^2.$$

Para una muestra<sup>13</sup>, la media y su varianza son:

$$\hat{\mu} = \sum_{h=1}^L W_h \hat{\mu}_h,$$

$$\hat{\mu} = \sum_{h=1}^L W_h \bar{y}_h = \bar{y}_{est}.$$

<sup>13</sup> *Idem.*, p. 89.

entonces:

$$\sigma^2(\hat{\mu}) = \sigma^2(\bar{y}_{rv}).$$

Si la muestra se reparte con asignación proporcional<sup>14</sup>:

$$\sigma^2(\hat{\mu}_{prop}) = \frac{1}{n} \sum_{h=1}^r W_h \sigma_h^2,$$

mientras que si se distribuye con asignación óptima<sup>15</sup>:

$$\sigma^2(\hat{\mu}_{op}) = \frac{1}{n} \left( \sum_{h=1}^r W_h \sigma_h \right)^2.$$

El objetivo es determinar las mejores fronteras de  $\mathbf{Y}$ , para las cuales  $\sigma^2(\hat{\mu})$  es mínimo cuando el estimador es la media poblacional  $\hat{\mu}$ . El problema es que  $\sigma^2(\hat{\mu})$  esta en función de las fronteras entre los estratos  $y_1, y_2, \dots, y_{r-1}$ . Obviamente,  $\sigma^2(\hat{\mu}_{prop})$  y  $\sigma^2(\hat{\mu}_{op})$  hacen que el tamaño del estrato dependa de la asignación que se usa al hacer la estratificación.

Para que la estratificación con asignación proporcional sea óptima en el conjunto de fronteras  $y_1, y_2, \dots, y_{r-1}$ , Dalenius<sup>16</sup> hace:

$$\sigma^2(\hat{\mu}_{prop}) = \frac{1}{n} \sum_{h=1}^r W_h \sigma_h^2,$$

<sup>14</sup> DALENIUS, Tore. 1950, *op. cit.*, p. 203.

<sup>15</sup> *Idem*, p. 204.

Bajo la suposición de una investigación teórica no se considera la CPF.

<sup>16</sup> *Idem*, p. 205.



un mínimo. Para simplificar a  $\sigma^2(\hat{\mu}_{ppp})$ , define a  $\Phi_{ppp}$ :

$$\Phi_{ppp} = \Phi_{ppp}(y_1, y_2, \dots, y_{r-1}) = \sum_{h=1}^r W_h \sigma_h^2 .$$

Dalenius<sup>17</sup> encuentra que  $\sigma^2(\hat{\mu}_{ppp})$  y  $\Phi_{ppp}$  son mínimos para el mismo conjunto de puntos.

También demuestra con un caso simple la forma general de encontrar las  $r-1$  fronteras óptimas:

Empezando con el caso simple de  $r=2$  se tienen  $y_0, y_1, y_2$  como las fronteras a encontrar. Los extremos son conocidos ya, por lo que  $y_1$  es el punto frontera entre un estrato y otro.

$$\Phi_{ppp} = W_1 \sigma_1^2 + W_2 \sigma_2^2 ,$$

donde:

$$W_1 = \int_{y_0}^{y_1} f(t) dt , y$$

$$W_2 = \int_{y_1}^{y_2} f(t) dt .$$

se deriva a  $\Phi_{ppp}$  con respecto a  $y_1$  y se iguala a 0:

$$\frac{\partial \Phi_{ppp}}{\partial y_1} = \frac{\partial W_1}{\partial y_1} \sigma_1^2 + \frac{\partial \sigma_1^2}{\partial y_1} W_1 + \frac{\partial W_2}{\partial y_1} \sigma_2^2 + \frac{\partial \sigma_2^2}{\partial y_1} W_2 .$$

---

<sup>17</sup> *Idem*, p. 205.

Donde:

$$\frac{\partial W_1}{\partial y_1} = f(y_1),$$

$$\frac{\partial W_2}{\partial y_1} = -f(y_1),$$

$$\frac{\partial \sigma_1^2}{\partial y_1} = \frac{f(y_1)}{W_1} [(y_1 - \mu_1)^2 - \sigma_1^2] \quad y$$

$$\frac{\partial \sigma_2^2}{\partial y_1} = \frac{f(y_1)}{W_2} [\sigma_2^2 - (y_1 - \mu_2)^2]$$

Por lo tanto:

$$\begin{aligned} \frac{\partial \Phi_{prop}}{\partial y_1} &= f(y_1)\sigma_1^2 + \frac{W_1 f(y_1)}{W_1} [(y_1 - \mu_1)^2 - \sigma_1^2] + \\ &\quad - f(y_1)\sigma_2^2 - \frac{W_2 f(y_1)}{W_2} [(y_1 - \mu_2)^2 - \sigma_2^2] \end{aligned}$$

que es igual a:

$$\frac{\partial \Phi_{prop}}{\partial y_1} = f(y_1) [(y_1 - \mu_1)^2 - (y_1 - \mu_2)^2]$$

Si:

$$\frac{\partial \Phi_{prop}}{\partial y_1} = 0,$$

se obtiene la solución:

$$y_1 = \frac{\mu_1 + \mu_2}{2}.$$

Para el caso general de  $r$  estratos, se tiene el problema de estratificación óptima con asignación proporcional para  $r-1$  puntos de estratificación. Obsérvese que  $W_h$ ,  $\mu_h$  y  $\sigma_h^2$  solamente son funciones de  $y_h$  y  $y_{h+1}$ . Entonces:

$$\Phi_{pmp} = W_1\sigma_1^2 + W_2\sigma_2^2 + \dots + W_h\sigma_h^2 + \dots + W_r\sigma_r^2,$$

y su respectiva derivada:

$$\frac{\partial \Phi_{pmp}}{\partial y_h} = \frac{\partial W_h}{\partial y_h} \sigma_h^2 + \frac{\partial \sigma_h^2}{\partial y_h} W_h + \frac{\partial W_{h+1}}{\partial y_h} \sigma_{h+1}^2 + \frac{\partial \sigma_{h+1}^2}{\partial y_h} W_{h+1},$$

$$\text{para } h = \overline{1, r-1}.$$

Donde:

$$\frac{\partial W_h}{\partial y_h} = f(y_h),$$

$$\frac{\partial W_{h+1}}{\partial y_h} = -f(y_h),$$

$$\frac{\partial \sigma_h^2}{\partial y_h} = \frac{f(y_h)}{W_h} [(y_h - \mu_h)^2 - \sigma_h^2] \quad \text{y}$$

$$\frac{\partial \sigma_{h+1}^2}{\partial y_h} = \frac{f(y_h)}{W_{h+1}} [\sigma_{h+1}^2 - (y_h - \mu_{h+1})^2]$$

Por lo tanto, se tiene que:

$$\frac{\partial \Phi_{pmp}}{\partial y_h} = f(y_h) [(y_h - \mu_h)^2 - (y_h - \mu_{h+1})^2]$$

y cuando:

$$\frac{\partial \Phi_{pp}}{\partial y_h} = 0,$$

se obtiene la solución general:

$$y_h = \frac{\mu_h + \mu_{h+1}}{2}, \quad h = \overline{1, r-1}.$$

La frontera óptima de estratificación puede ser encontrada por medio de un proceso iterativo de cortes de frontera ya que  $\mu_h$  depende de  $y_h$ .

Dalenius expone, con gran parecido a la discusión anterior, la estratificación óptima con asignación óptima para las fronteras  $y_1, y_2, \dots, y_{r-1}$ , es decir, hace:

$$\sigma^2(\hat{\mu}_{op}) = \frac{1}{n} \left( \sum_{h=1}^r W_h \sigma_h \right)^2,$$

un mínimo. Para simplificar también a  $\sigma^2(\hat{\mu}_{op})$  define a  $\Phi_{op}$ , como:

$$\Phi_{op} = \Phi_{op}(y_1, y_2, \dots, y_{r-1}) = \sum_{h=1}^r W_h \sigma_h.$$

Dalenius encuentra<sup>18</sup> que  $\sigma^2(\hat{\mu}_{op})$  y  $\Phi_{op}$  son mínimos para el mismo conjunto de puntos.

Se usa el artificio de  $\Phi_{op}$  porque la derivada es más simple.

El conjunto de puntos de estratificación se encuentra con la solución de:

$$\frac{\partial \Phi_{op}}{\partial y_h} = 0.$$

---

<sup>18</sup> *Idem*, p. 210.

Empezando con el caso simple de  $r = 2$  se tiene:

$$\Phi_{op} = W_1 \sigma_1 + W_2 \sigma_2,$$

y su respectiva derivada:

$$\frac{\partial \Phi_{op}}{\partial y_1} = \frac{\partial W_1}{\partial y_1} \sigma_1 + \frac{\partial \sigma_1}{\partial y_1} W_1 + \frac{\partial W_2}{\partial y_1} \sigma_2 + \frac{\partial \sigma_2}{\partial y_1} W_2.$$

Donde:

$$\frac{\partial W_1}{\partial y_1} = f(y_1),$$

$$\frac{\partial W_2}{\partial y_1} = -f(y_1),$$

$$\frac{\partial \sigma_1}{\partial y_1} = \frac{f(y_1)}{2W_1\sigma_1} [(y_1 - \mu_1)^2 - \sigma_1^2] \quad y$$

$$\frac{\partial \sigma_2}{\partial y_1} = -\frac{f(y_1)}{2W_2\sigma_2} [(y_1 - \mu_2)^2 - \sigma_2^2]$$

Por lo tanto:

$$\frac{\partial \Phi_{op}}{\partial y_1} = f(y_1)\sigma_1 + \frac{W_1 f(y_1)}{2W_1\sigma_1} [(y_1 - \mu_1)^2 - \sigma_1^2] +$$

$$-f(y_1)\sigma_2 + \frac{W_2 f(y_1)}{2W_2\sigma_2} [(y_1 - \mu_2)^2 - \sigma_2^2]$$

$$\frac{\partial \Phi_{op}}{\partial y_1} = \frac{f(y_1)}{2} \left[ \frac{\sigma_1^2 + (y_1 - \mu_1)^2}{\sigma_1} - \frac{\sigma_2^2 + (y_1 - \mu_2)^2}{\sigma_2} \right].$$

Si:

$$\frac{\partial \Phi_{op}}{\partial y_1} = 0,$$

la solución está en:

$$\frac{(y_1 - \mu_1)^2 + \sigma_1^2}{\sigma_1} = \frac{(y_1 - \mu_2)^2 + \sigma_2^2}{\sigma_2}.$$

Para poblaciones  $f(y)$  simétricas alrededor de la media  $\mu$ , la solución óptima para el caso de  $r=2$  es cuando  $y_1 = \mu$ . Un ejemplo trivial es el proporcionado por la población rectangular  $f(y) = 1$ . Entonces:

$$\Phi_{op} = \frac{1}{12} [y_1^2 + (1 - y_1)^2]$$

resulta en  $y_1 = \frac{1}{2}$  como el punto de solución óptima.

En analogía a los resultados anteriores, la solución al caso general de estratificación óptima con asignación óptima de  $r-1$  estratos se obtiene con:

$$\Phi_{op} = W_1 \sigma_1 + W_2 \sigma_2 + \dots + W_h \sigma_h + \dots + W_r \sigma_r,$$

Igualmente  $W_h$ ,  $\mu_h$  y  $\sigma_h$  solamente dependen de  $y_h$  y  $y_{h+1}$ .

y la respectiva derivada de  $\Phi_{op}$  es:

$$\frac{\partial \Phi_{op}}{\partial y_h} = \frac{\partial W_h}{\partial y_h} \sigma_h + \frac{\partial \sigma_h}{\partial y_h} W_h + \frac{\partial W_{h+1}}{\partial y_h} \sigma_{h+1} + \frac{\partial \sigma_{h+1}}{\partial y_h} W_{h+1},$$

para  $h = 1, r-1$ .

Donde:

$$\frac{\partial W_h}{\partial y_h} = f(y_h),$$

$$\frac{\partial W_{h+1}}{\partial y_h} = -f(y_h),$$

$$\frac{\partial \sigma_h}{\partial y_h} = \frac{f(y_h)}{2W_h \sigma_h} [(y_h - \mu_h)^2 - \sigma_h^2] \quad y$$

$$\frac{\partial \sigma_{h+1}}{\partial y_h} = -\frac{f(y_h)}{2W_{h+1} \sigma_{h+1}} [(y_h - \mu_{h+1})^2 - \sigma_{h+1}^2]$$

Por lo tanto:

$$\frac{\partial \Phi_{op}}{\partial y_h} = f(y_h) \sigma_h + \frac{W_h f(y_h)}{2W_h \sigma_h} [(y_h - \mu_h)^2 - \sigma_h^2] +$$

$$-f(y_h) \sigma_{h+1} + \frac{W_{h+1} f(y_h)}{2W_{h+1} \sigma_{h+1}} [(y_h - \mu_{h+1})^2 - \sigma_{h+1}^2]$$

$$\frac{\partial \Phi_{op}}{\partial y_h} = \frac{f(y_h)}{2} \left[ \frac{\sigma_h^2 + (y_h - \mu_h)^2}{\sigma_h} - \frac{\sigma_{h+1}^2 + (y_h - \mu_{h+1})^2}{\sigma_{h+1}} \right].$$

Y cuando:

$$\frac{\partial \Phi_{op}}{\partial y_h} = 0,$$

la solución general está en:

$$\frac{(y_h - \mu_h)^2 + \sigma_h^2}{\sigma_h} = \frac{(y_h - \mu_{h+1})^2 + \sigma_{h+1}^2}{\sigma_{h+1}}, \quad h = \overline{1, r-1}.$$

y la frontera óptima  $y_h$  se encuentra aplicando un proceso iterativo de esta relación.

Si la solución exacta sólo depende de los parámetros de cada estrato (media y varianza) y no depende de los ponderadores poblacionales, entonces es indistinto su uso para los estimadores del total y la media.

Por otra parte, es obvio que en muchas aplicaciones, la variable de estratificación no es idéntica a la variable de estimación. Cuando se realiza sobre una variable de estratificación  $X$ , que está linealmente relacionada a la variable de estimación  $Y$  de la forma siguiente:

$$Y = \alpha + \beta X + \eta ,$$

Dalenius<sup>19</sup> da la solución exacta a la estratificación óptima con asignación proporcional, la cual se calcula con procesos iterativos de:

$$y_h = \frac{\alpha + \beta \mu_{x_h} + \alpha + \beta \mu_{x_{h+1}}}{2} .$$

Para  $y_h = \alpha + \beta x_h$ , la ecuación anterior se reduce a:

$$x_h = \frac{\mu_{x_h} + \mu_{x_{h+1}}}{2} .$$

La solución exacta a la estratificación óptima sobre una variable de estratificación con asignación óptima, la cual se calcula con procesos iterativos, está dada por<sup>20</sup>:

$$\frac{(\beta^2 \sigma_{x_h}^2 + 2\gamma) + \beta^2 (x_h - \mu_{x_h})^2}{\sqrt{(\beta^2 \sigma_{x_h}^2 + \gamma)}} = \frac{(\beta^2 \sigma_{x_{h+1}}^2 + 2\gamma) + \beta^2 (x_h - \mu_{x_{h+1}})^2}{\sqrt{(\beta^2 \sigma_{x_{h+1}}^2 + \gamma)}} .$$

<sup>19</sup> DALENIUS, T. y GURNEY, M. 1951, *op. cit.*, p. 135.

<sup>20</sup> *Idem*, p.136.



#### 4.4.2. Una aproximación a la solución exacta

La solución teórica para la asignación óptima condicionada a la frontera óptima  $y_h$ , entre los estratos  $h$  y  $h+1$ , para satisfacer la relación<sup>21</sup>:

$$\frac{(y_h - \mu_h)^2 + \sigma_h^2}{\sigma_h} = \frac{(y_h - \mu_{h+1})^2 + \sigma_{h+1}^2}{\sigma_{h+1}}$$

Esta solución óptima no es práctica porque los parámetros en la ecuación dependen de  $y_h$ , por lo que se tiene que recurrir a procesos iterativos.

Otros investigadores han propuesto varios criterios para formar estratos que se aproximan a la solución óptima propuesta por Dalenius. Algunos de estos criterios serán citados cronológicamente a continuación:

El criterio de Hansen, Hurwitz y Madow<sup>22</sup> establece constante al estimador del total para los diferentes estratos, es decir:

$$W_h \mu_h = \text{constante} .$$

Un segundo criterio, dado por Aoyama<sup>23</sup>, es la regla de estratificación equidistante, la cual considera que:

$$(y_h - y_{h-1}) = \text{constante} .$$

<sup>21</sup> DALENIUS, Tore. 1950, *op. cit.*, p. 212.

<sup>22</sup> HANSEN, M. L., HURWITZ, W. N., y MADOW, W.G.: *Sample surveys methods and theory*, Vol. 1, New York: John Wiley and Sons, Inc., 1953, p. 219.

<sup>23</sup> AOYAMA, H., *A study of the stratified random sampling*, *Annals of Mathematical Statistics*, Vol. 6 (1), 1954, pp. 1-36.

Aoyama derivó la regla equidistante aplicando el teorema del valor medio a la solución exacta dada por Dalenius y asumiendo que la variación de la densidad es pequeña en cada estrato.

El criterio de Dalenius y Gurney<sup>24</sup> propone que los valores iguales de  $W_h\sigma_h$  para los estratos conducirán a fronteras que están cerca de las óptimas:

$$W_h\sigma_h = W_{h+1}\sigma_{h+1}, \text{ ó}$$

$$W_h\sigma_h = \text{constate.}$$

Estos tres criterios no constituyen un procedimiento simple para establecer directamente un conjunto de fronteras óptimas porque sus parámetros también dependen de las fronteras<sup>25</sup>. Es decir, se tiene que recurrir a procesos iterativos, como en el caso de la solución exacta de Dalenius. Cuando los estratos se dan en clases, el criterio de Dalenius y Gurney puede ayudar a combinar algunas de ellas y quizá a dividir otras para obtener buenas fronteras.

La igualdad de valores  $W_h\sigma_h$  también significa igualdad de tamaños de muestra  $n_h$  en asignación óptima. Por tanto, la asignación óptima también tiende a resultar en muestras de asignación igual  $n_h = n_{h+1}$ .

Otro es el criterio de Sethi<sup>26</sup>, que ajusta la distribución de probabilidades poblacional a una distribución teórica donde las fronteras óptimas están predeterminadas.

Se ha demostrado, con el criterio de Ekman<sup>27</sup>, que si se crean estratos con

<sup>24</sup> DALENIUS, T. y HODGES, S. P., 1959, *op. cit.*, p.90.

<sup>25</sup> KISH, L., 1979, *op. cit.*, p. 134.

<sup>26</sup> SETHI, V. K.: *A note on optimum stratification of populations for estimating the population means*, The Australian Journal of Statistics, 5, 1963, pp. 20-33.

<sup>27</sup> HESS, I., SETHI, V. K., Balakrishnan, T. R. : *Stratification: A practical investigation*, Journal American Association, V. 61, 1966, pp. 74-90.

valores iguales de:

$$W_h(y_h - y_{h-1}) = \text{constante}$$

se obtienen buenas aproximaciones a las fronteras óptimas. Esto también puede servir de regla para combinar y dividir clases, pero no para obtener la solución directa de las fronteras óptimas.

Se menciona al último el criterio de Dalenius porque tiene más importancia en esta investigación. Dalenius supone <sup>28</sup> que la distribución de  $\mathbf{Y}$  dentro de los estratos  $f(y_h)$  se comporta como una distribución rectangular cuando el número de estratos es grande; esto es que, se generan estratos tan angostos hasta que cada  $f(y_h)$  parezca un rectángulo de altura  $f(y_h)$  y de base  $(y_h - y_{h-1})$ , entonces:

$$W_h = \int_{y_{h-1}}^{y_h} f(t) dt \doteq (y_h - y_{h-1})f(y_h),$$

y la desviación estándar de  $f(y_h)$  en el estrato  $h$ ,  $\sigma_h$ , es:

$$\sigma_h = \frac{y_h - y_{h-1}}{\sqrt{12}}.$$

Por el teorema del valor medio, existe un valor  $f(y_h)$  de  $f(y)$  en el  $h$ -ésimo estrato tal que:

$$\begin{aligned} \sqrt{12} \sum_{h=1}^r W_h \sigma_h &\doteq \sum_{h=1}^r f(y_h) (y_h - y_{h-1})^2 \\ &= \sum_{h=1}^r [f(y_h) (y_h - y_{h-1})]. \end{aligned}$$

---

<sup>28</sup> DALENIUS, T. y HODGES, S. P., *op. cit.*, p. 81.

Si se introduce la transformación:

$$Z(y) = \int_x^y \sqrt{f(t)} dt ,$$

se tiene que:

$$Z(y_h) - Z(y_{h-1}) = \int_{y_{h-1}}^{y_h} \sqrt{f(t)} dt \doteq \sqrt{f(y_h)} (y_h - y_{h-1}) .$$

Donde  $f(y_h)$  es un valor constante de  $f(y)$  en el estrato  $h$ . Al sustituir estas aproximaciones se encuentra que:

$$\sqrt{12} \sum_{h=1}^k W_h \sigma_h \doteq \sum_{h=1}^k [Z(y_h) - Z(y_{h-1})]^2 .$$

Ya que  $[Z(y_r) - Z(y_0)]$  es fijo (pues  $y_r$  es el máximo valor; y  $y_0$ , el mínimo valor), se verifica que la suma de la derecha se minimiza al hacer  $[Z(y_h) - Z(y_{h-1})]$  constante.

En resumen, dada  $f(y)$ , la regla es formar la distribución acumulada de  $\sqrt{f(y)}$  y escoger las fronteras  $y_h$  de tal manera que éstas formen estratos iguales en la escala  $cum\sqrt{f(y)}$ , donde *cum* significa la respectiva distribución acumulada.

La solución exacta supone la estratificación de una población hipotética pero, como se mencionó, Dalenius<sup>29</sup> mostró la relación de la solución exacta cuando la estratificación no es llevada sobre la variable de estimación  $\mathbf{X}$ ; "aunque se necesita más investigación, el resultado de Dalenius sugiere que la regla  $cum\sqrt{f(y)}$  aplicada a  $\mathbf{x}$  debe

---

<sup>29</sup> DALENIUS, T. y GURNEY, M., 1951, *op. cit.*, pp. 134 a 136.

dar una estratificación eficiente para la variable de estimación  $Y$ , que tiene una regresión lineal en  $X$  con alta correlación<sup>30</sup>.

En la práctica la variable de estratificación  $x$  está asociada a una población finita, donde  $f(x)$  es la función de distribución de frecuencias. La construcción de  $f(x)$  se define con la determinación de intervalos de clase y el conteo de eventos que pertenecen a ellos, este conteo es la frecuencia de la variable  $x$  en cada intervalo.

Los límites del intervalo juegan un papel importante en la búsqueda de las fronteras. Ellos están relacionados con el rango del intervalo<sup>31</sup> ( $Ra_i$ ):

$$Ra_i = Ls - Li,$$

donde  $Ls$  es el límite superior y  $Li$  es el límite inferior del intervalo.

Este rango también tiene una relación con el rango de los datos de la población ( $Ra$ ) y el número de intervalos necesario para cubrirla.

Donde:

$$Ra = x_{(n)} - x_{(1)}.$$

Regularmente el número de intervalos necesario para cubrir a la población está entre 5 y 20 intervalos<sup>32</sup>. En el caso extremo, el rango del intervalo puede ser de tal manera que se tenga la frecuencia de un evento en cada uno.

---

<sup>30</sup> COCHRAN, William G., 1981, *op. cit.*, p. 179.

<sup>31</sup> El rango de un intervalo de frecuencia es la diferencia entre los límites superior e inferior.

<sup>32</sup> Se suele tomar entre 5 y 20 intervalos de frecuencia, según los datos. Estos intervalos son del mismo tamaño, pero también es posible usar intervalos de distintos tamaños. Véase SPIEGEL, Murray R.: *Estadística*, McGraw-Hill, México, 1994, p. 39.

Puede intuirse de manera lógica, crear intervalos de igual rango cuando se da a priori el número de intervalos, con la siguiente relación:

$$Ra_i = \frac{Ra}{I},$$

donde  $I$  es el número total de intervalos.

Y cuando se da a priori el dato del rango del intervalo ( $Ra_i$ ), entonces:

$$I = \frac{Ra}{Ra_i}.$$

Formada la distribución de frecuencias de  $x$ , la regla de Dalenius establece calcular la raíz de la frecuencia  $\sqrt{f(x)}$  y su acumulada  $cum\sqrt{f(x)}$ . Estos cálculos se esquematizan en el cuadro 4.1.

Para formar estratos iguales a escala *cum* se hace la partición equitativa de  $cum\sqrt{f(x)}$ .

Esta partición se obtiene verificando el máximo valor acumulado:

$$\max(cum\sqrt{f(x)}) = \sum_{k=1}^I \sqrt{f_k(x)},$$

donde:  $f_k(x) = \sum_{b_{k-1}}^{b_k} c \in (L_{k-1}, L_k)$ , para toda  $k = \overline{1, I}$ , es la frecuencia o conteo del número de eventos ( $c$ ) que pertenece a un intervalo determinado.

Cuadro 4.1

Fórmulas de cálculo de  $f(x)$ ,  $\sqrt{f(x)}$  y  $cum\sqrt{f(x)}$  para la regla de Dalenius aplicada a X.

k No.	Intervalos <sup>33</sup> de clase		Frecuencia de X (Núm. de eventos) $f(x)$	Raíz cuadrada de la frecuencia $\sqrt{f(x)}$	Acumulada de la raíz de frecuencia $cum\sqrt{f(x)}$
	Inferior $Li$	Superior $Ls$			
1	$Li$	$Ls_1$	$f_1(x) = \sum_{Li}^{Ls_1} e \in (Li, Ls_1)$	$\sqrt{f_1(x)}$	$\sqrt{f_1(x)}$
2	$Ls_1$	$Ls_2$	$f_2(x) = \sum_{Ls_1}^{Ls_2} e \in (Ls_1, Ls_2)$	$\sqrt{f_2(x)}$	$\sum_{k=1}^2 \sqrt{f_k(x)}$
3	$Ls_2$	$Ls_3$	$f_3(x) = \sum_{Ls_2}^{Ls_3} e \in (Ls_2, Ls_3)$	$\sqrt{f_3(x)}$	$\sum_{k=1}^3 \sqrt{f_k(x)}$
...	.....	.....	.....	.....	.....
g	$Ls_{g-1}$	$Ls_g$	$f_g(x) = \sum_{Ls_{g-1}}^{Ls_g} e \in (Ls_{g-1}, Ls_g)$	$\sqrt{f_g(x)}$	$\sum_{k=1}^g \sqrt{f_k(x)}$
...	.....	.....	.....	.....	.....
l	$Ls_{l-1}$	$Ls_l$	$f_l(x) = \sum_{Ls_{l-1}}^{Ls_l} e \in (Ls_{l-1}, Ls_l)$	$\sqrt{f_l(x)}$	$\sum_{k=1}^l \sqrt{f_k(x)}$

Y se forman estratos de igual rango a escala  $cum$ , dividiendo al  $max(cum\sqrt{f(x)})$  entre el número óptimo de estratos  $r^{op}$ , de la manera siguiente:

<sup>33</sup> En los intervalos, el lado inferior es abierto, y el lado superior es cerrado.

$$Ra_{cum} = \frac{\max(\text{cum}\sqrt{f(x)})}{r^{op}}$$

donde  $Ra_{cum}$  es la longitud o rango del estrato a escala *cum*.

Las fronteras de los estratos a escala *cum* quedan determinadas para  $r$  estratos por el siguiente cálculo:

$$f_h = h \times Ra_{cum}, \text{ y}$$

$$f_{i_h} = f_{h-1} = (h-1) \times Ra_{cum}, \text{ para } h = \overline{1, r},$$

donde  $f_h$  es la frontera superior y  $f_{i_h} = f_{h-1}$  es la frontera inferior del estrato.

Ahora bien, las fronteras de los estratos a escala *cum*  $(f_{h-1}, f_h)$  corresponden a cortes de las fronteras de los estratos a escala  $f(x)$  o escala original  $(x_{h-1}, x_h)$ .

$(f_{h-1}, f_h)$  corresponden a:

$$\begin{cases} x_0 = x_{(0)} \\ x_h \text{ es tal que } \sum_{k=1}^h \sqrt{f_k(x)} \leq f_h \text{ para } g = \overline{1, h} \text{ y } h = \overline{1, r-1} \\ x_r = x_{(N)} \end{cases}$$

Así concluye la correspondencia de las fronteras  $x_0, \dots, x_h, \dots, x_r$  para los estratos de la variable de estratificación  $\mathbf{x}$  con la aproximación a la solución exacta.



ESTADO DE LA UNIÓN  
VALOR DE LA PRODUCCIÓN

## CAPITULO V

### APLICACIÓN A UN PROBLEMA PRÁCTICO. EL CASO DE UNA "RAMA DE ACTIVIDAD ECONÓMICA EN MÉXICO"

En este capítulo se aplica el criterio de la formación de estratos óptimos y el proceso de selección de muestra y obtención de los estimadores descritos en los primeros capítulos.

Se aplicará sobre información económica de formación bruta de capital fijo<sup>1</sup> del Banco de México. Ésta se refiere a 63 de las 72 ramas de actividad económica del país, según la Clasificación Mexicana de Actividades Económicas (CMAE)<sup>2</sup>.

El capítulo se compone de cuatro apartados a fin de cumplir con su objetivo y con el de la tesis. Así, primeramente se comentan los motivos que dieron lugar a la aplicación de la técnica descrita, una breve descripción de la población en estudio y la información disponible. Posteriormente, se expone el resultado del muestreo aleatorio simple de la población. En seguida, se presentan los resultados de aplicar el criterio de Dalenius para la estratificación óptima y el proceso de selección y de estimación mediante muestreo estratificado; y por último, se compararán los dos diseños utilizados.

Cabe señalar que la aplicación de la técnica se ejemplifica ampliamente para una rama de la actividad productiva nacional, por ser suficiente para mostrar las ventajas de

---

<sup>1</sup> La formación bruta de capital fijo de las empresas, según Las Naciones Unidas, comprende el valor de los bienes adquiridos por las industrias (públicas y privadas), los productores de servicios y las instituciones privadas sin fines de lucro, destinados a incrementar sus activos fijos; menos sus desechos y ventas netas de bienes de capital usados.  
<sup>2</sup> CMAE, Clasificación Mexicana de Actividades Económicas, 1970.

la técnica; se trata de la Industria de Refrescos y Aguas Gaseosas denominada "Rama 22". Sin embargo, se realizan los cálculos para todas las ramas y se citan en los cuadros seis y siete del apéndice B, éstos son el tamaño de población y de muestra por estrato para cada una de las 63 de las 72 ramas del CMAE.

### 5.1. Antecedentes de la aplicación y de la información

El objetivo de aplicación de la técnica es revisar empíricamente los principios de optimalidad del criterio de Dalenius en la formación de estratos. Es muy común encontrar literatura relacionada con muestreo estratificado en cuanto a las ventajas obtenidas en la confianza y error estadísticos de un muestreo basado en una selección por estratos. También se sabe que estas ventajas son más amplias a medida que los estratos son más homogéneos en su interior y más heterogéneos en su exterior.

Para hacer la aplicación, se utiliza la encuesta del Banco de México conocida como "Encuesta sobre Formación de Capital Fijo"<sup>3</sup>. Ésta surge por la necesidad de disponer con series históricas sobre acervos brutos<sup>4</sup>, depreciación<sup>5</sup> y formación bruta de capital fijo de diferentes unidades económicas.

---

<sup>3</sup> Esta Encuesta tiene como antecedentes genéricos a tres proyectos principales que también han sido desarrollados por el mismo Banco. Estos son: "Cuentas Nacionales y Acervos de Capital, consolidados y por tipo de actividad económica, 1950-1967"; "Acervos y Formación de Capital, 1960-1975"; e "Inversión Fija del Sector Empresarial, 1975-1980".

<sup>4</sup> Los acervos brutos en un lapso determinado son el valor de todos los bienes dados de alta y que están en operación.

<sup>5</sup> La depreciación en un periodo dado es la estimación de la pérdida de valor de los bienes de capital debido al desgaste o la obsolescencia prevista. El concepto también es conocido como consumo de capital fijo.

La información disponible comprende el período de 1960 a la fecha en miles de pesos a precios corrientes y constantes de 1980.

El objetivo de la encuesta es medir el activo fijo de las unidades económicas, por lo que la población está teóricamente integrada por todos los agentes del proceso productivo que desempeñan una actividad económica determinada, es decir, las empresas.

Esta población existe; pero no es posible conocerla en forma total para efectos de análisis. Sin embargo, es factible conocerla parcialmente. Para ello se dispone de algunas listas o directorios especializados: el principal es el directorio de empresas consideradas por la Secretaría de Hacienda y Crédito Público como causantes mayores en 1989.

Sólo el directorio de causantes mayores de la SHCP asocia a la población con variables cuantitativas; esto permite clasificar a las empresas de acuerdo a la actividad económica que realizan.

La población bajo estudio está clasificada en subpoblaciones que a su vez se consideran un dominio de estudio. No todos los dominios de estudio son sujetos a estimaciones por muestreo; algunos por su importancia en el contexto económico o bien por el reducido número de unidades que contienen quedan sujetas a censo (ver cuadro 3, del apéndice B).

Los números y nombres de las subpoblaciones de empresas corresponden a las diversas ramas de actividad productiva según la Clasificación Mexicana de Actividades Económicas (CMAE) 1970. La clasificación se compone de nueve Grandes Divisiones y nueve Divisiones que integran a la Gran División III, y en su interior se organizan el total de 72 ramas de actividad económica (ver cuadro 1, apéndice B).

La principal variable de estimación en la encuesta es el total de formación bruta de capital fijo por rama.

No existe información de los valores reales de formación de capital fijo del universo de empresas que componen a las ramas objeto de estudio, dato necesario en el muestreo para clasificar según la variable de estimación.

Sin embargo, el directorio de causantes mayores de la SHCP de 1989 asocia una variable cuantitativa a la población de empresas; ésta es el ingreso de las empresas.

La relación entre el ingreso y la formación de capital se mide con el cálculo del coeficiente de correlación entre  $X$  y  $Y$ .

Sea:

- $X$  La variable Ingreso, y
- $Y$  La variable Formación Bruta de Capital Fijo.

Para estimar el coeficiente de correlación entre  $X$  y  $Y$  se consideran los valores e  $Y$  estimados de la encuesta de 1989 a nivel de rama (ver cuadro 4, apéndice B), lo cual es válido pues es un dato histórico, y para  $X$  se considera el valor del ingreso agregado de las empresas en cada rama. Es decir, las observaciones son los valores de formación bruta de capital para  $Y$  y el ingreso para  $X$  agregados por rama.

El coeficiente de correlación se estimó (cuadro 5.1) con el paquete de cómputo Micro TSP<sup>6</sup>; el resultado es de 0.36 para el conjunto de las 63 ramas que forman el marco muestral de la encuesta.

---

<sup>6</sup> TSP es un conjunto de herramientas de predicción y regresión desarrolladas para computadoras personales

Cuadro 5.1

Cálculo del coeficiente de correlación

Referencia de Gran División	Número de ramas que la forman	Coefficiente de Correlación
1	4	IND
2	6	0.709
3	48	0.626
4	1	NEC
5	1	NEC
6	2	-1
7	2	-1
8	2	1
9	3	IND
Total	72	0.360

Fuente: Marco muestral de la Encuesta sobre Formación de Capital Fijo de Banco de México.

Nota: IND: Información no disponible para toda la Gran División 1, y dos ramas para la Gran División 9.

NEC: No existe coeficiente de correlación. Sólo contienen una rama la Gran División 4 y 5.

En general, si el coeficiente de correlación es de  $\rho > \pm 0.5$ , entonces se dice que la variable  $Y$  está explicada satisfactoriamente por la variable  $X$ .

La correlación también es calculada por gran división suponiendo que cada una se comporta de diferente manera. En los resultados del cuadro 5.1, las Grandes Divisiones seis, siete y ocho sólo contienen dos ramas cada una, por lo que el resultado de su coeficiente de correlación puede ser no significativo. El coeficiente de correlación para la Gran División II es de 0.709 y para la Gran División III es de 0.626, los cuales dan mejor correlación entre las variables  $Y$  y  $X$  que para el grupo completo de las 72 ramas.

Aunque se tiene evidencia parcial de que la correlación es buena entre  $Y$  y  $X$  para todas las ramas, es la única variable cuantitativa con la que cuenta el directorio de 1989

**Cuadro 5.1**

**Cálculo del coeficiente de correlación**

Referencia de Gran División	Número de ramas que la forman	Coefficiente de Correlación
1	4	IND
2	6	0.799
3	48	0.626
4	1	NEC
5	1	NEC
6	2	-1
7	2	-1
8	2	1
9	3	IND
<b>Total</b>	<b>72</b>	<b>0.360</b>

Fuente : Marco muestral de la Encuesta sobre Formación de Capital Fijo de Banco de México.

Nota: IND: Información no disponible para toda la Gran División 1, y dos ramas para la Gran División 9.

NEC: No existe coeficiente de correlación. Solo contienen una rama la Gran División 4 y 5.

En general, si el coeficiente de correlación es de  $\rho > \pm 0.5$ , entonces se dice que la variable **Y** está explicada satisfactoriamente por la variable **X**.

La correlación también es calculada por gran división suponiendo que cada una se comporta de diferente manera. En los resultados del cuadro 5.1, las Grandes Divisiones seis, siete y ocho sólo contienen dos ramas cada una, por lo que el resultado de su coeficiente de correlación puede ser no significativo. El coeficiente de correlación para la Gran División II es de 0.799 y para la Gran División III es de 0.626, los cuales dan mejor correlación entre las variables **Y** y **X** que para el grupo completo de las 72 ramas.

Aunque se tiene evidencia parcial de que la correlación es buena entre **Y** y **X** para todas las ramas, es la única variable cuantitativa con la que cuenta el directorio de 1989.

## 5.2. Estimación con muestreo aleatorio simple

Para aplicar el MAS a la Rama 22 (ver Cuadro 3, apéndice B), se retoman los aspectos que se citaron en el capítulo II:

- El proceso de selección

Se determina el tamaño de la muestra por seleccionar considerando que se desea estimar el total de  $Y$  para la rama 22 dado el estimador  $\hat{Y}$  en muestreo aleatorio simple, mediante la variable correlacionada  $X$ :

$$n = \frac{N^2 k^2 \sigma^2}{\epsilon^2 + N k^2 \sigma^2}$$

Donde:

El tamaño de la población  $N$ , son las 126 empresas de la Rama 22.

El nivel de confianza  $(1 - \alpha)$  es del 90%, su respectiva desviación  $k$  correspondiente a una distribución normal es igual a 1.64<sup>7</sup>.

Para la encuesta, el error admisible  $\epsilon$  es del 5% del total poblacional de la variable "Ingreso" ( $X$ ). Es decir, se supone que los resultados de "Formación Bruta de Capital Fijo" ( $Y$ ) también tienen margen de error del 5% respecto al valor del estadístico.

$$\epsilon = |Y - \hat{Y}| = 0.05 \hat{Y}$$

---

<sup>7</sup> Este dato se puede confirmar en las Tablas de la Distribución Normal que indiquen el área al nivel de confianza determinado. Literatura de Estadística.

El error admisible en función de  $\bar{x}$  es :

$$\epsilon = 0.05 \hat{X} = 0.05 \sum_{i=1}^{126} x_i = 0.05 \times 3,798 = 190 \text{ mill. de pesos,}$$

es decir, se dispone a correr el riesgo de que el resultado para la rama 22 se aleje 190 millones de pesos del valor real al 90% de confianza.

Por otra parte, la varianza poblacional de  $\mathbf{x}$  se calcula con:

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N}$$

Donde la media poblacional de  $\mathbf{x}$  es:

$$\bar{X} = \frac{\sum_{i=1}^N x_i}{N} = \frac{\sum_{i=1}^{126} x_i}{126} = 30 \text{ mill. de pesos.}$$

Entonces:

$$\sigma^2 = \frac{\sum_{i=1}^{126} (x_i - 30)^2}{126} = 1,161 \text{ unidades cuadradas.}$$

Y la desviación estándar es:

$$\sigma = \sqrt{\sigma^2} = \sqrt{1,161} = 34 \text{ unidades.}$$

Sustituyendo los valores anteriores en la ecuación de  $n$ , el cálculo de tamaño de muestra es de:

$$\begin{aligned} n &= \frac{(126)^2 (1.64)^2 (1.161)}{(190)^2 + (126)(1.64)^2(1.161)} \\ &= 115.4 \text{ empresas a seleccionar.} \end{aligned}$$



La selección de la muestra de 115.4 se redondea a 116 empresas. Esta selección es aleatoria con la ayuda de tablas de números aleatorios<sup>8</sup> o con ayuda de programas de cómputo.

**Cuadro 5.2**  
**116 números aleatorios generados entre 1 y 126,**  
**elemento seleccionado *i*.**

1	13	26	39	54	66	81	95	107	119
2	14	27	40	55	67	83	96	108	120
3	15	28	41	56	68	84	97	109	121
4	16	30	42	57	69	85	98	110	122
5	17	31	43	58	70	86	99	111	123
6	18	32	44	59	72	87	100	112	124
7	19	33	46	80	73	88	101	113	125
8	20	34	47	61	74	89	102	114	126
9	21	35	49	62	75	91	103	115	
10	22	36	50	63	77	92	104	116	
11	23	37	51	64	78	93	105	117	
12	25	38	53	65	80	94	106	118	

Fuente: Número aleatorios generados en Microsoft Excel V. 5.0<sup>9</sup>.

Los 116 números aleatorios presentados en el cuadro 5.2 equivalen a los indicadores de las empresa a seleccionar.

En el cuadro 5.3, se observa la asociación del ingreso relacionado al indicador de la empresa según el número aleatorio seleccionado para la población de la rama 22 (ver cuadro 3 del apéndice B), de acuerdo al tamaño de muestra requerido.

De esta manera queda seleccionada la muestra para explicar a la población.

<sup>8</sup> Puede consultarse la tabla de números aleatorios en SPIEGEL, Murray R.: *Estadística*, McGraw-Hill, México, 1994, p.545.

<sup>9</sup> La instrucción ejecutada en el paquete Excel para generar números aleatorios es:  
 =ALEATORIO.ENTRE(a,b)

Donde a y b son los números mínimo y máximo de la lista a seleccionar. En el caso específico de 116 empresas a seleccionar de 126, la instrucción fue "=aleatorio.entre(1,126)" generado 116 veces.

Cuadro 5.3

Ingreso asociado al elemento seleccionado *i* para la muestra de 116 empresas.

(Millones de pesos)

$X_i =$ Ingreso asociado a la empresa <i>i</i>			
$X_1 = 6.7$	$X_{30} = 12.1$	$X_{59} = 22.5$	$X_{88} = 36.9$
$X_2 = 6.7$	$X_{31} = 12.3$	$X_{60} = 22.9$	$X_{89} = 37.2$
$X_3 = 6.8$	$X_{32} = 12.3$	$X_{61} = 23.6$	$X_{90} = 37.5$
$X_4 = 6.8$	$X_{33} = 12.4$	$X_{62} = 24.1$	$X_{91} = 37.8$
$X_5 = 6.9$	$X_{34} = 13.5$	$X_{63} = 24.2$	$X_{92} = 38.9$
$X_6 = 6.9$	$X_{35} = 13.7$	$X_{64} = 24.4$	$X_{93} = 39.0$
$X_7 = 7.0$	$X_{36} = 13.7$	$X_{65} = 24.6$	$X_{94} = 40.4$
$X_8 = 7.1$	$X_{37} = 13.8$	$X_{66} = 24.6$	$X_{95} = 40.8$
$X_9 = 7.2$	$X_{38} = 14.3$	$X_{67} = 24.7$	$X_{96} = 41.9$
$X_{10} = 7.2$	$X_{39} = 14.4$	$X_{68} = 25.0$	$X_{97} = 43.5$
$X_{11} = 7.6$	$X_{40} = 14.8$	$X_{69} = 25.2$	$X_{98} = 43.8$
$X_{12} = 7.6$	$X_{41} = 14.9$	$X_{70} = 25.9$	$X_{99} = 45.5$
$X_{13} = 7.9$	$X_{42} = 15.7$	$X_{71} = 26.1$	$X_{100} = 45.7$
$X_{14} = 8.4$	$X_{43} = 16.2$	$X_{72} = 26.8$	$X_{101} = 48.2$
$X_{15} = 8.6$	$X_{44} = 16.6$	$X_{73} = 27.1$	$X_{102} = 51.2$
$X_{16} = 8.8$	$X_{45} = 16.8$	$X_{74} = 27.3$	$X_{103} = 52.0$
$X_{17} = 9.0$	$X_{46} = 16.9$	$X_{75} = 27.6$	$X_{104} = 59.4$
$X_{18} = 9.2$	$X_{47} = 18.5$	$X_{76} = 27.6$	$X_{105} = 60.0$
$X_{19} = 9.2$	$X_{48} = 18.9$	$X_{77} = 27.7$	$X_{106} = 66.9$
$X_{20} = 9.3$	$X_{49} = 19.3$	$X_{78} = 28.3$	$X_{107} = 67.0$
$X_{21} = 9.4$	$X_{50} = 20.0$	$X_{79} = 28.7$	$X_{108} = 70.8$
$X_{22} = 9.5$	$X_{51} = 20.1$	$X_{80} = 29.4$	$X_{109} = 71.6$
$X_{23} = 9.9$	$X_{52} = 20.6$	$X_{81} = 30.3$	$X_{110} = 71.7$
$X_{24} = 10.1$	$X_{53} = 20.7$	$X_{82} = 30.6$	$X_{111} = 79.0$
$X_{25} = 10.1$	$X_{54} = 20.8$	$X_{83} = 31.4$	$X_{112} = 85.5$
$X_{26} = 10.6$	$X_{55} = 21.1$	$X_{84} = 32.6$	$X_{113} = 115.7$
$X_{27} = 11.0$	$X_{56} = 21.2$	$X_{85} = 36.5$	$X_{114} = 174.0$
$X_{28} = 11.1$	$X_{57} = 22.2$	$X_{86} = 36.6$	$X_{115} = 188.9$
$X_{29} = 11.4$	$X_{58} = 22.5$	$X_{87} = 36.7$	$X_{116} = 260.6$

Fuente: Marco muestral de la Encuesta de Formación de Capital de Banco de México. Por razones de confidencialidad, los nombres de las empresas no se mencionan, ni se mantiene el indicador de la población. (Ver cuadro 3, apéndice B).

• El proceso de estimación

El proceso de estimación consiste en calcular los estadísticos o estimadores de los parámetros poblacionales que interesa estimar.

El parámetro a estimar es el total poblacional de  $Y$  que está denotado por  $\hat{Y}$ , mediante la variable correlacionada  $X$ , cuyo estimador es  $\hat{X}$ :

$$\hat{X} = N\bar{x},$$

donde:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{\sum_{i=1}^{116} x_i}{116} = 31 \text{ mill. de pesos,}$$

por lo tanto:

$$\hat{X} = (126)(31) = 3,906 \text{ mill. de pesos.}$$

Se calcula el error estándar de  $\hat{X}$  para evaluar la precisión del estimador, dado por:

$$\sigma(\hat{X}) = \frac{N\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N}},$$

donde  $\sigma$  es la desviación estándar de la muestra, y al sustituir valores:

$$\sigma(\hat{X}) = \frac{126 \sqrt{1249}}{\sqrt{116}} \sqrt{\frac{126 - 116}{126}} = 116.5,$$

con una corrección por finitud de:

$$\text{CPF} = \frac{N - n}{N} = \frac{126 - 116}{126} = 0.0794.$$

El intervalo al 90% de confianza para  $\hat{X}$  es:

$$\left( \hat{X} - 1.64 \sigma(\hat{X}) < X < \hat{X} + 1.64 \sigma(\hat{X}) \right).$$

Se sustituyen valores:

$$(3,906 - 1.64 \times 116.5 < X < 3,906 + 1.64 \times 116.5).$$

Entonces, el valor real de  $X$  esta dentro del intervalo:

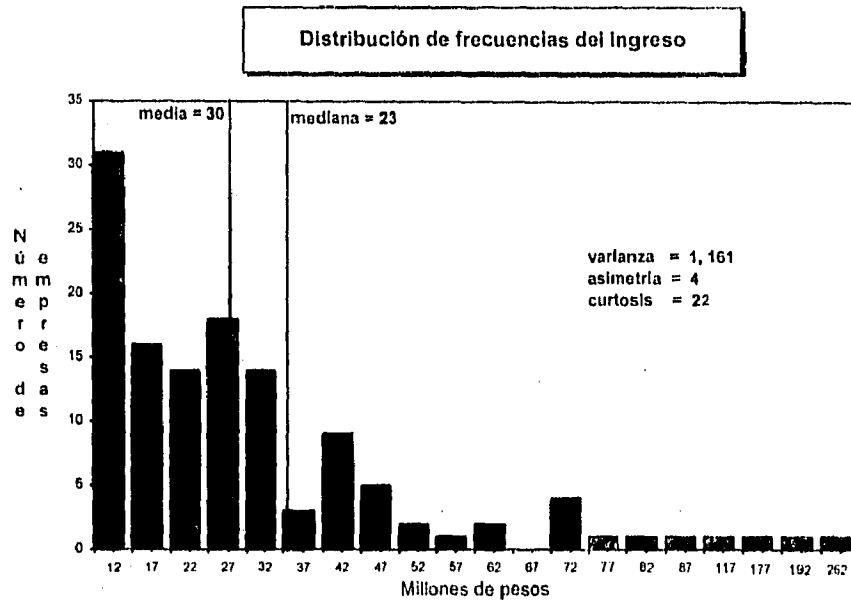
$$(3,715 < X < 4,097),$$

es decir, el valor real de  $X$  puede estar desde 3,715 hasta 4,097 millones de pesos como estimación del total de Ingreso para la rama 22.

### 5.3. Estimación con muestreo aleatorio estratificado

Se analiza intuitivamente la variable asociada  $x$  a la población de la Rama 22, mediante la función de distribución de frecuencias que se puede observar en la siguiente gráfica:

Gráfica 5.1



Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Rama 22, datos de 1989. (Ver cuadro 3, apéndice B).

El Ingreso asociado a la población de la Rama 22 se comporta como una población con asimetría positiva, cuyo coeficiente de asimetría es igual a cuatro, comparada con una población simétrica con coeficiente de asimetría cero.

En el caso del coeficiente de curtosis igual a 22, describe a una distribución leptocúrtica, lo que sugiere datos muy concentrados en pocos intervalos, y por la asimetría que presenta, estos intervalos son los del menor valor de  $X$ .

En otras palabras, el Ingreso de las empresas se puede observar como el más alto relativamente a las demás, y para la mayoría de las empresas su ingreso es el más alto, el mediano a mínimo.

El comportamiento de esta variable muestra que para un número de empresas que son relativamente pequeñas, el ingreso es el más alto, como resultado de la estructura de la industria, que es una industria de alta concentración. En otras palabras, el ingreso de las empresas más grandes es el más bajo, y el ingreso de las empresas más pequeñas es el más alto. Este comportamiento es el resultado de la estructura de la industria, que es una industria de alta concentración. En otras palabras, el ingreso de las empresas más grandes es el más bajo, y el ingreso de las empresas más pequeñas es el más alto. Este comportamiento es el resultado de la estructura de la industria, que es una industria de alta concentración.

El comportamiento de esta variable muestra que para un número de empresas que son relativamente pequeñas, el ingreso es el más alto, como resultado de la estructura de la industria, que es una industria de alta concentración.

CONCLUSIONES

El análisis de la estructura de la industria muestra que la industria es altamente concentrada, con un número limitado de empresas que controlan la mayor parte del ingreso. Este comportamiento es el resultado de la estructura de la industria, que es una industria de alta concentración.

El análisis de la estructura de la industria muestra que la industria es altamente concentrada, con un número limitado de empresas que controlan la mayor parte del ingreso.

En otras palabras, el Ingreso de las empresas es para algunas muy alto, o alto relativamente a las demás, y para la mayoría de las empresas su Ingreso es de valor mediano a mínimo.

El comportamiento de esta población en estudio, y en general de poblaciones con tendencias asimétricas (ver cuadro 5 del apéndice B), indica que es mejor usar el MAE en vez del MAS<sup>10</sup> como procedimiento de muestreo. Las empresas con menores Ingresos, en la cola izquierda de la gráfica, tienen más probabilidad de selección, pues tienen mayor oportunidad por ser superior en número; y las empresas con mayores Ingresos, en la cola derecha de la gráfica, corren el riesgo de no ser seleccionadas, pues son un número mucho menor que las primeras, lo que genera como consecuencia la subestimación del estimador. O en otro caso, que por alguna razón de aleatorización se eligen a más empresas de altos Ingresos, lo que puede ocasionar sobrestimación del estimador.

Para aplicar el MAE a la población de la rama 22, se retoman los aspectos que se citaron en el capítulo III, los cuales se van a desarrollar a continuación:

#### • El proceso de selección

En el proceso de selección, primero se forman los estratos y después se calcula el tamaño de muestra a seleccionar. En la formación de estratos, las principales consideraciones a tener se describieron en el capítulo IV, las cuales son: la variable de

---

<sup>10</sup> Entendiéndose que la selección es aleatoria en ambos casos.

estratificación (que ya se puntualizó en el apartado 5.1.), el número de estratos y la determinación de sus fronteras. Posteriormente la asignación de la muestra en cada estrato.

La determinación del número de estratos se deja como última etapa en la formación de estratos, pues se resolverá con el cálculo y comparación de la varianza cuando se adicionan más estratos.

El camino a seguir para determinar sus fronteras es la aproximación a la estratificación óptima con asignación óptima, propuesta por Dalenius.

Para desarrollar el proceso del criterio de Dalenius, se diseñó una macro<sup>11</sup> adaptada para el paquete de Lotus-123 (ver apéndice C), que estratifica series de datos no mayores a 900 elementos y calcula el tamaño de muestra con asignación óptima.

El criterio de Dalenius dice que dada  $f(y)$ , la regla es formar la distribución acumulada de  $\sqrt{f(y)}$  y escoger las fronteras  $y_h$  de tal manera que éstas formen intervalos iguales en la escala  $\text{cum}\sqrt{f(y)}$ . Lo mismo se dice si se aplica sobre  $X$ .

Desarrollando lo anterior, primero se genera la distribución de frecuencias de  $X$  que es  $f(x)$ .

En la construcción de  $f(x)$ , existen 126 empresas de las cuales cuatro tienen valores atípicos a los de la población en general (ver diagrama de dispersión en la gráfica 5.2).

Si se excluyen las cuatro empresas con valores de la variable  $X$  más altos (véase cuadro 3 del apéndice B), la varianza de  $X$  disminuye de 1,161 a 294 unidades

---

<sup>11</sup> Macro es un programa con una serie de instrucciones que ejecuta cálculos y comandos de Lotus-123.

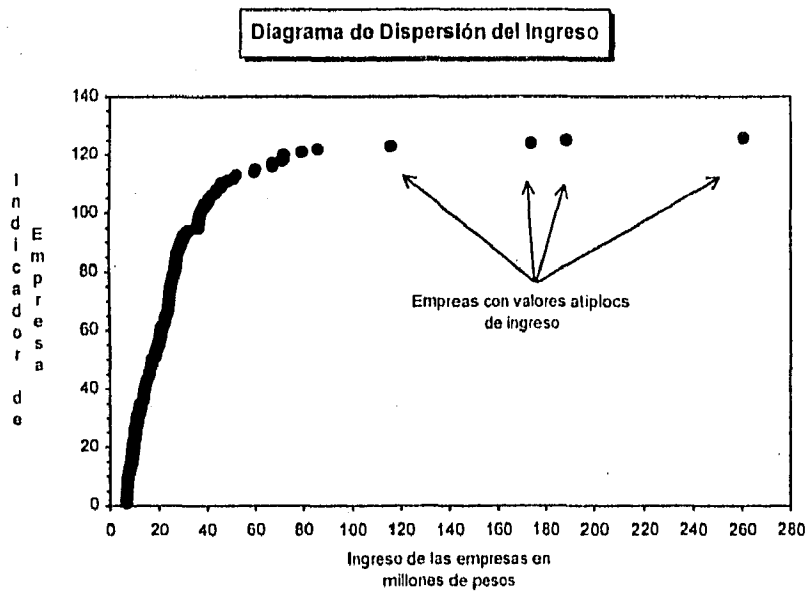


cuadradas. El cálculo del tamaño de muestra depende directamente de la varianza, disminución en la varianza equivale a disminución en tamaño de muestra.

La importante diferencia en la varianza sólo es provocada por el 3.2% de la población (cuatro empresas de 126), así que se construye con ellas un estrato cierto en

$f(x)$ .

Gráfica 5.2.



Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Rama 22, datos de 1989. (Ver cuadro 3, apéndice B).

Las 122 empresas restantes se agrupan en intervalos de frecuencia de rango

$(Ra_i)$  igual a cinco millones de pesos. Donde:

$$Ra = x_{(N-4)} - x_{(1)} = x_{(122)} - x_{(1)}$$

$$= 85.5 - 6.7 = 78.8, \text{ entonces}$$

$$I = \frac{Ra}{Ra_1} = \frac{78.8}{5} = 15.76.$$

El número redondeado de intervalos es de 16, para  $(Ra_1)$  igual a cinco millones de pesos.

De esta manera la distribución consta de 16 intervalos de igual y uno de diferente tamaño (véase cuadro 5.4).

En la construcción de estratos con la regla de Dalenius, se calcula la raíz de la distribución de frecuencias  $\sqrt{f(x)}$ , y posteriormente su respectiva acumulación  $cum\sqrt{f(x)}$  (véase cuadro 5.4).

La construcción de estratos a escala *cum*, se obtiene verificando el máximo valor acumulado  $\max(cum\sqrt{f(x)})$  y se divide entre el número de estratos<sup>12</sup>  $r$ , de la siguiente manera:

$$Ra_{cum}^r = \frac{\max(cum\sqrt{f(x)})}{r}$$

Donde el máximo valor de la acumulación (excluyendo al estrato cierto<sup>13</sup>) es:

$$\max(cum\sqrt{f(x)}) = 37.1.$$

---

<sup>12</sup> El número de estratos se determinará posteriormente con un proceso iterativo de cálculo de varianzas

<sup>13</sup> El estrato cierto ya tiene sus límites de fronteras.

Cuadro 5.4

Cálculo de  $cum\sqrt{f(x)}$  para ingreso

k	Intervalos <sup>14</sup> de clase. límites		Frecuencia de X (Núm. de empresas) $f(x)$	Raíz cuadrada de la frecuencia $\sqrt{f(x)}$	Acumulada de la raíz de frecuencia $cum\sqrt{f(x)}$
	Inferior $LS_{k-1}$ (millones de pesos de 1989)	Superior $LS_k$			
1	6.7	11.7	31	6	5.6
2	11.7	16.7	16	4	9.6
3	16.7	21.7	14	4	13.3
4	21.7	26.7	18	4	17.6
5	26.7	31.7	14	4	21.3
6	31.7	36.7	3	2	23.0
7	36.7	41.7	9	3	26.0
8	41.7	46.7	5	2	28.3
9	46.7	51.7	2	1	29.7
10	51.7	56.7	1	1	30.7
11	56.7	61.7	2	1	32.1
12	61.7	66.7	0	0	32.1
13	66.7	71.7	4	2	34.1
14	71.7	76.7	1	1	35.1
15	76.7	81.7	1	1	36.1
16	81.7	86.7	1	1	37.1

Estrato cierto:

17	86.7	261.7	4		
----	------	-------	---	--	--

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Rama 22, datos de 1989. (Ver cuadro 3, apéndice B).

El valor óptimo de  $r$  se conoce con el cálculo iterativo de la varianza, paralelamente se encuentran las fronteras de los estratos para diferentes valores de  $r$ .

Para ejemplificar el comportamiento de los cortes de frontera de  $x$  en cada estrato,  $r$  recorre los valores desde dos (que es el mínimo número de estratos en MAE) hasta el valor de nueve.

<sup>14</sup> En los intervalos, el lado inferior es abierto, y el lado superior es cerrado.

Entonces el rango del estrato a escala *cum* queda:

$$Ra_{cum}^r = \frac{\max(cum\sqrt{f(x)})}{r} = \frac{37.1}{r}, \text{ donde } r = \overline{2,9}.$$

El cálculo desarrollado de  $Ra_{cum}^r$  para cada caso de  $r = \overline{2,9}$ , puede observarse en el cuadro 5.5.

**Cuadro 5.5**

**Longitud del intervalo a escala *cum*  
para  $r = \overline{2,9}$**

$r$	$Ra_{cum}^r = \frac{37.1}{r}$
2	18.55
3	12.36
4	9.28
5	7.42
6	6.18
7	5.30
8	4.64
9	4.12

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Rama 22, datos de 1989.

Las fronteras del estrato cierto quedan implícitas con los límites superior e inferior del intervalo que contiene a las cuatro últimas empresas.

Las fronteras a escala *cum*, para  $r = \overline{2,9}$ , varían para cada caso porque  $Ra_{cum}^r$  varía según  $r$ . Empezando con el caso de  $r=2$ , estas fronteras son:

Para el estrato  $h=1$ :

$$f_{h-1} = f_0 = (1 - 1) \times Ra_{cum}^2 = 0, \text{ y}$$

$$f_h = f_1 = 1 \times Ra_{cum}^2 = 18.55.$$

Para el estrato  $h=2$ :

$$f_{h-1} = f_1 = (2 - 1) \times Ra_{cum}^2 = 18.55, \text{ y}$$

$$f_h = f_2 = 2 \times Ra_{cum}^2 = 37.1.$$

Ahora bien, las fronteras a escala *cum*, redondeadas a un decimal (0,18.6) y (18.6, 37.1), corresponden a las fronteras  $(x_0, x_1)$  y  $(x_1, x_2)$  a escala  $f(x)$ .

La correspondencia de fronteras, para  $r = 2$ , es de la manera siguiente :

$$x_h = \begin{cases} h = 0 & x_0 = x_{(0)} = 6.7 \\ h = 1 & x_1 \text{ es tal que su respectiva } \sum_{k=1}^1 \sqrt{f_k(x)} \leq f_1 = 18.55 \\ & \Rightarrow x_1 = 26.7 \text{ y su respectiva } \sum_{k=1}^4 \sqrt{f_k(x)} = 17.6 \leq 18.55 \\ h = 2 & x_2 = x_{(N-4)} = 85.5 \quad \text{o} \\ & x_2 = 86.7 \quad \text{de } f_2 = 37.1 \quad \text{que corresponde al intervalo } 16 \end{cases}$$

Los resultados anteriores se resumen en el siguiente cuadro:

Cuadro 5.6.1

Correspondencia de fronteras de escala *cum* con escala original para el caso de dos estratos.

Escala <i>cum</i>		Escala original	
Fronteras de estratos	Número de oventos	Fronteras de estratos	Empresas
$f_{h-1}   f_h$		$x_{h-1}   x_h$	$N_h$
0.0 - 18.6	4	6.7 - 26.7	79
18.6 - 37.1	12	26.7 - 86.7	43
37.1 - 41.1	4	86.7 - 261.7	4

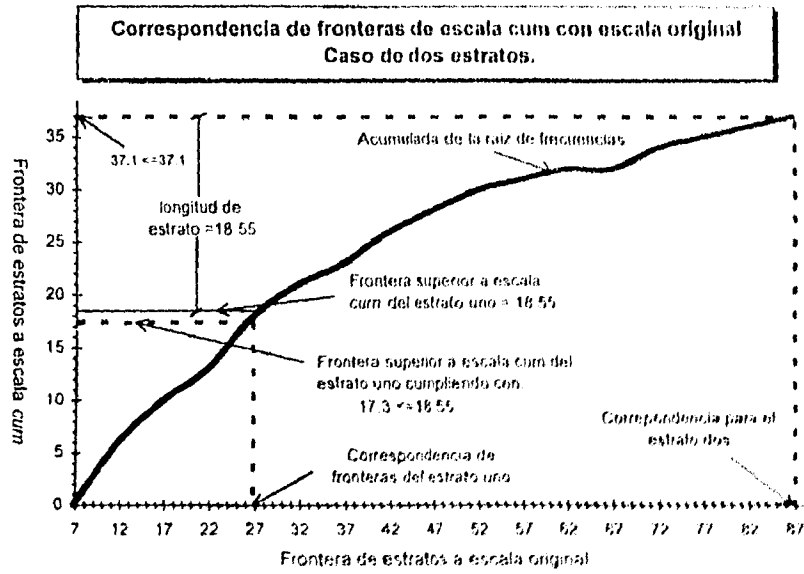
Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Rama 22, datos de 1989. (Ver cuadro 3, apéndice B).

El cuadro anterior puede ser apreciado en la gráfica 5.3, donde las líneas pausadas indican la correspondencia de las fronteras de cada estrato a escala *cum*,

$\sum_{k=1}^K \sqrt{f_k(s)} \leq f_h$ , con las fronteras a escala original  $x_h$ ; y la línea continua representa la frontera  $f_h$  a escala *cum* que resultó del cálculo con  $Ra_{cum}^2$ .

Los casos de  $r$  de tres a nueve estratos se observan en los cuadros del 5.6.2 al 5.6.8, donde cada renglón representa un estrato y el último renglón es el estrato cierto.

Gráfica 5.3



Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Barras de México, Rama 22, datos de 1989 (Ver cuadro 3, apéndice B)

Cuadro 5.6.2

Correspondencia de fronteras de escala cum con escala original para el caso de tres estratos.

Escala cum		Escala original		
Fronteras de estratos	Número de eventos	Fronteras de estratos		Empresas
$f_{1j} - f_{1j-1}$		$x_{1j} - x_{1j-1}$	$x_{1j}$	$n_j$
7.3 - 12.4	2	5.7 - 15.7		47
12.4 - 24.7	4	15.7 - 36.7		49
24.7 - 37.1	11	36.7 - 88.7		76
37.1 - 41.1	4	88.7 - 281.7		4

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Barras de México, Rama 22, datos de 1989 (Ver cuadro 3, apéndice B)

**Ejemplo 1**

Correspondencia de los datos de la tabla 1 con la original para el caso de cuatro atributos.

Tabla 1		Tabla original	
Elementos de atributos	Elementos de atributos	Elementos de atributos	Elementos de atributos
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
11	11	11	11
12	12	12	12
13	13	13	13
14	14	14	14
15	15	15	15
16	16	16	16
17	17	17	17
18	18	18	18
19	19	19	19
20	20	20	20
21	21	21	21
22	22	22	22
23	23	23	23
24	24	24	24
25	25	25	25
26	26	26	26
27	27	27	27
28	28	28	28
29	29	29	29
30	30	30	30
31	31	31	31
32	32	32	32
33	33	33	33
34	34	34	34
35	35	35	35
36	36	36	36
37	37	37	37
38	38	38	38
39	39	39	39
40	40	40	40
41	41	41	41
42	42	42	42
43	43	43	43
44	44	44	44
45	45	45	45
46	46	46	46
47	47	47	47
48	48	48	48
49	49	49	49
50	50	50	50

**Ejemplo 2**

Correspondencia de los datos de la tabla 2 con la original para el caso de cuatro atributos.

Tabla 2		Tabla original	
Elementos de atributos	Elementos de atributos	Elementos de atributos	Elementos de atributos
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
11	11	11	11
12	12	12	12
13	13	13	13
14	14	14	14
15	15	15	15
16	16	16	16
17	17	17	17
18	18	18	18
19	19	19	19
20	20	20	20
21	21	21	21
22	22	22	22
23	23	23	23
24	24	24	24
25	25	25	25
26	26	26	26
27	27	27	27
28	28	28	28
29	29	29	29
30	30	30	30
31	31	31	31
32	32	32	32
33	33	33	33
34	34	34	34
35	35	35	35
36	36	36	36
37	37	37	37
38	38	38	38
39	39	39	39
40	40	40	40
41	41	41	41
42	42	42	42
43	43	43	43
44	44	44	44
45	45	45	45
46	46	46	46
47	47	47	47
48	48	48	48
49	49	49	49
50	50	50	50



Cuadro 5.6.3

Correspondencia de fronteras de escala *cum* con escala original para el caso de cuatro estratos.

Escala <i>cum</i>		Escala original		
Fronteras de estratos	Número de eventos	Fronteras de estratos		Empresas
$f_{h-1} \mid f_h$		$x_{h-1} \mid x_h$	$N_h$	
0.0 - 9.3	1	6.7 - 11.7		31
9.3 - 18.6	3	11.7 - 26.7		48
18.6 - 27.8	3	26.7 - 41.7		26
27.8 - 37.1	9	41.7 - 86.7		17
37.1 - 41.1	4	86.7 - 261.7		4

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Rama 22, datos de 1989. (Ver cuadro 3, apéndice B).

Cuadro 5.6.4

Correspondencia de fronteras de escala *cum* con escala original para el caso de cinco estratos.

Escala <i>cum</i>		Escala original		
Fronteras de estratos	Número de eventos	Fronteras de estratos		Empresas
$f_{h-1} \mid f_h$		$x_{h-1} \mid x_h$	$N_h$	
0.0 - 7.4	1	6.7 - 11.7		31
7.4 - 14.9	2	11.7 - 21.7		30
14.9 - 22.3	2	21.7 - 31.7		32
22.3 - 29.7	3	31.7 - 46.7		17
29.7 - 37.1	8	46.7 - 86.7		12
37.1 - 41.1	4	86.7 - 261.7		4

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Rama 22, datos de 1989. (Ver cuadro 3, apéndice B).

**Cuadro 5.6.5**

**Correspondencia de fronteras de escala *cum* con escala original para el caso de seis estratos.**

Escala <i>cum</i>		Escala original	
Fronteras de estratos	Número de eventos	Fronteras de estratos	Empresas
$f_{h-1} \mid f_h$		$x_{h-1} \mid x_h$	
0.0 - 6.2	1	6.7 - 11.7	31
6.2 - 12.4	1	11.7 - 16.7	16
12.4 - 18.6	2	16.7 - 26.7	32
18.6 - 24.7	2	26.7 - 36.7	17
24.7 - 30.9	4	36.7 - 56.7	17
30.9 - 37.1	6	56.7 - 86.7	9
37.1 - 41.1	4	86.7 - 261.7	4

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Rama 22, datos de 1989. (Ver cuadro 3, apéndice B).

**Cuadro 5.6.6**

**Correspondencia de fronteras de escala *cum* con escala original para el caso de siete estratos.**

Escala <i>cum</i>		Escala original	
Fronteras de estratos	Número de eventos	Fronteras de estratos	Empresas
$f_{h-1} \mid f_h$		$x_{h-1} \mid x_h$	
0.0 - 6.2	1	6.7 - 11.7	31
6.2 - 12.4	1	11.7 - 16.7	16
12.4 - 18.6	2	16.7 - 26.7	32
18.6 - 24.7	2	26.7 - 36.7	17
24.7 - 30.9	4	36.7 - 56.7	17
30.9 - 37.1	6	56.7 - 86.7	9
37.1 - 41.1	4	86.7 - 261.7	4

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Rama 22, datos de 1989. (Ver cuadro 3, apéndice B).

Cuadro 5.6.7

Correspondencia de fronteras de escala *cum* con escala original para el caso de ocho estratos.

Escala <i>cum</i>		Escala original	
Fronteras de estratos	Número de eventos	Fronteras de estratos	Empresas
$f_{h-1}   f_h$		$x_{h-1}   x_h$	
0.0 - 4.7	0	0.0 - 6.6	0
4.7 - 9.3	1	6.6 - 21.7	61
9.3 - 13.9	2	21.7 - 21.7	0
12.4 - 18.6	1	21.7 - 26.7	18
18.6 - 23.2	2	26.7 - 36.7	17
23.2 - 27.8	1	36.7 - 46.7	14
27.8 - 32.5	5	46.7 - 76.7	10
32.5 - 37.1	4	76.7 - 86.7	2
37.1 - 41.1	4	86.7 - 261.7	4

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Rama 22, datos de 1989. (Ver cuadro 3, apéndice B).

Cuadro 5.6.8

Correspondencia de fronteras de escala *cum* con escala original para el caso de nueve estratos.

Escala <i>cum</i>		Escala original	
Fronteras de estratos	Número de eventos	Fronteras de estratos	Empresas
$f_{h-1}   f_h$		$x_{h-1}   x_h$	
0.0 - 4.1	0	0.0 - 6.6	0
4.1 - 8.2	1	6.6 - 21.7	61
8.2 - 12.4	1	21.7 - 26.7	18
12.4 - 16.5	1	26.7 - 31.7	14
16.5 - 20.6	1	31.7 - 36.7	3
20.6 - 24.7	2	36.7 - 41.7	9
24.7 - 28.6	2	41.7 - 56.7	8
28.6 - 33.0	4	56.7 - 81.7	8
33.0 - 37.1	4	81.7 - 86.7	1
37.1 - 41.1	4	86.7 - 261.7	4

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Rama 22, datos de 1989. (Ver cuadro 3, apéndice B).

Puede observarse que para los casos de  $r$  de siete a nueve, se formaron estratos sin número de eventos, esto significa que la población para algunos de ellos es de cero porque ninguna empresa pertenece a los rangos de dichos estrato.

Por lo tanto, generar más de siete estratos ya no es significativo para esta rama porque se violan los supuestos de  $N_h > 0$  para toda  $h$ . En adelante sólo se considerarán las iteraciones de dos a seis estratos.

Determinadas las fronteras de los estratos, se calcula el tamaño de muestra y su asignación para cada estrato.

Se desea estimar el total poblacional para la rama 22 de la variable  $Y$  con el estimador  $\hat{Y}$ , mediante la variable de estratificación  $X$ , así que se considera el cálculo del tamaño de muestra total  $n$  con:

$$n = \frac{\left( \sum_{h=1}^r N_h \sigma_h \right)^2}{\frac{e^2}{k^2} + \sum_{h=1}^r N_h \sigma_h^2}$$

Y la asignación óptima de muestra por estrato  $n_h$  se calcula con:

$$n_h = \frac{n N_h \sigma_h}{\sum_{h=1}^r N_h \sigma_h}$$

En el cuadro 5.7, se muestran los cálculos de  $(N_h \sigma_h)$  y  $(N_h \sigma_h^2)$  sobre  $X$  para cada  $r$  de dos a seis.

Con los resultados del cuadro 5.7, se desarrolla el cálculo de tamaño de muestra sobre los mismos supuestos de confianza y error que en MAS (ver cuadro 5.8).

**Cuadro 5.7**

**Cálculo de  $(N_h \sigma_h)$  y  $(N_h \sigma_h^2)$  del ingreso (millones de pesos)**  
**Para  $r = 2,6$ .**

Indicador de estrato	$(N_h \sigma_h)$ para cada caso de estratos					$(N_h \sigma_h^2)$ para cada caso de estratos				
	$r=2$	$r=3$	$r=4$	$r=5$	$r=6$	$r=2$	$r=3$	$r=4$	$r=5$	$r=6$
$h=1$	493	136	45	45	45	3,163	413	70	70	70
$h=2$	681	226	217	89	23	10,773	1,044	979	264	33
$h=3$		381	126	78	92		5,575	606	189	204
$h=4$			227	87	50			3,039	400	150
$h=5$				93	81				857	390
$h=6$					71					557
<b>Suma</b>	<b>691</b>	<b>743</b>	<b>615</b>	<b>392</b>	<b>362</b>	<b>13,936</b>	<b>7,032</b>	<b>4,694</b>	<b>1,780</b>	<b>1,464</b>

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Rama 22, datos de 1989. (Ver cuadro 3, apéndice B).

**Cuadro 5.8**

**Cálculo de tamaño de muestra total y asignación por estrato**  
**Para  $r = 2,6$ .**

Indicador de estrato	Caso de número de estratos				
	$r=2$	$r=3$	$r=4$	$r=5$	$r=6$
$h=1$	27	7	2	2	2
$h=2$	36	11	10	3	1
$h=3$		18	6	3	3
$h=4$			10	2	2
$h=5$				5	3
$h=6$					3
estrato cierto	4	4	4	4	4
<b>Total de muestra</b>	<b>67</b>	<b>40</b>	<b>32</b>	<b>19</b>	<b>18</b>

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Rama 22, datos de 1989. (Ver cuadro 3, apéndice B).

Como puede observarse en el cuadro 5.8, el tamaño de muestra disminuye mientras se incrementa el número de estratos. También al aumentar  $r$ ,  $(N_h \sigma_h)$  y  $(N_h \sigma_h^2)$  disminuyen y es más homogéneo el comportamiento dentro de los estratos.

Una vez formados los estratos se determina su número óptimo. Para determinarlo se calcula la varianza descompuesta en intravarianza e intervarianza.

El estrato óptimo es aquel donde la intravarianza es mínima y la intervarianza es máxima, y el recorrido de  $r$  es de dos a seis estratos.

Se retoman las fórmulas de cálculo de la intravarianza e intervarianza, respectivamente sus ecuaciones son:

$$\frac{1}{N} \sum_{h=1}^r N_h \sigma_h^2 \quad \text{y} \quad \frac{1}{N} \sum_{h=1}^r N_h (\bar{Y}_h - \bar{Y})^2 .$$

El cálculo de varianzas, mostrado en el cuadro 5.9, muestra como número óptimo el caso de seis estratos, porque la intravarianza es mínima y la intervarianza es máxima en seis.

**Cuadro 5.9**

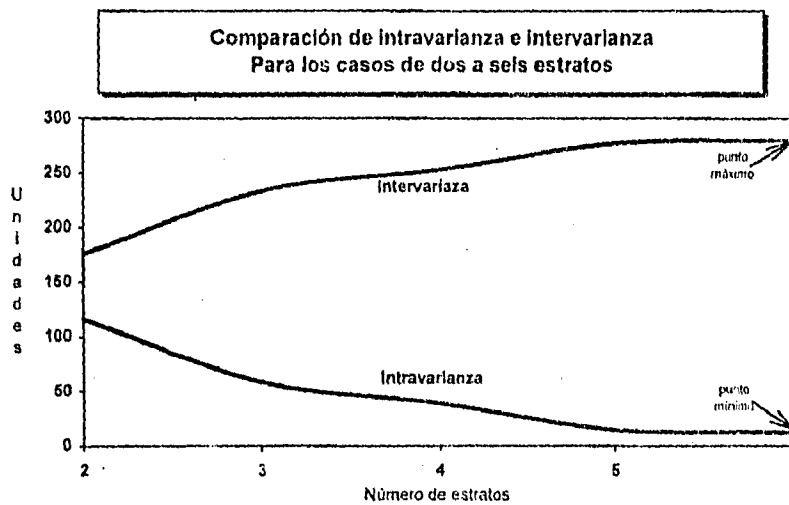
**Cálculo de la Intravarianza e intervarianza para  $r = 2,6$**

Caso de número de estratos	Intravarianza	Intervarianza	F-Fisher
$r = 2$	116	176	1.52
$r = 3$	59	233	4.50
$r = 4$	39	253	6.49
$r = 5$	15	277	18.47
$r = 6$	12	280	23.33

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Rama 22, datos de 1989. (Ver cuadro 3, apéndice B).

Obsérvese, en la gráfica 5.4, que la intervarianza permanece con una línea ascendente y la intravarianza permanece como una línea descendente hasta llegar a seis estratos.

Gráfica 5.4



Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Rama 22, datos de 1989. (Ver cuadro 3, apéndice B).

Es posible que más allá de seis estratos el efecto de intravarianza e intervarianza de la variable  $X$  siga su rumbo óptimo. Pero para este caso, aumentar más estratos ya no es significativo por falta de unidades en algunos de ellos.

Por otra parte, un aumento a más de seis podría no ser significativo si el coeficiente de correlación entre las variables  $X$  y  $Y$  no rebasa a 0.95.

La selección de la muestra en los seis estratos es aleatoria y la selección de empresas en el estrato cierto se hace con probabilidad uno ya que incluye a todos sus elementos.

Dado el número óptimo de estratos igual a seis, el tamaño total de muestra a seleccionar es de 14 (ver cuadro 5.8), además las cuatro del estrato cierto dan como resultado 18 empresas en total.

Los números aleatorios para los seis estratos según sus fronteras se describen en el siguiente cuadro:

**Cuadro 5.10**

**14 número aleatorios generados entre los límites de cada estrato. asociado al elemento  $i$  de la población de empresas.**

Estrato 1	Estrato 2	Estrato 3	Estrato 4	Estrato 5	Estrato 6
10	35	53	84	103	114
23		67	86	107	119
		70		112	122

Fuente: Microsoft Excel versión 5.

La información de ingreso correspondiente a las empresas de la rama 22 (ver cuadro 3 del apéndice B), asociadas a los números aleatorios del cuadro 5.10, queda esquematizada en el cuadro 5.11.



Cuadro 5.11

Ingreso asociado al elemento  $i$  para la muestra estratificada de 18 empresas.  
(Millones de pesos)

Estrato 1	Estrato 2	Estrato 3	Estrato 4	Estrato 5	Estrato 6	Estrato. cierto
$n_1 = 2$	$n_2 = 1$	$n_3 = 3$	$n_4 = 2$	$n_5 = 3$	$n_6 = 3$	$n_7 = 4$
X1 = 7.2	X3 = 12.4	X4 = 18.9	X7 = 27.6	X9 = 39.0	X12 = 59.4	X15 = 115.7
X2 = 9.9		X5 = 24.1	X8 = 27.7	X10 = 43.5	X13 = 71.6	X16 = 174.0
		X6 = 24.6		X11 = 51.2	X14 = 85.5	X17 = 188.9
						X18 = 260.6

Fuente: Marco muestral de la Encuesta sobre Formación de Capital, de Banco de México. Rama 22, datos de 1989. Por razones de confidencialidad, los nombres de las empresas no se mencionan, ni se mantiene el indicador de la población. (Ver cuadro 3, apéndice B).

• El proceso de estimación

Seleccionada la muestra, queda por concluir el cálculo de la estimación del total poblacional de  $Y$ , cuyo estimador se denota por  $\hat{Y}$ , con la variable de estratificación de ingreso  $X$ .

$$\hat{X} = N\bar{x}_{cm}$$

Donde  $N$  es la población de las 126 empresas (incluyendo al estrato cierto).

Y:

$$\bar{x}_{cm} = \sum_{h=1}^7 W_h \bar{x}_h \quad \text{de}$$

$$\bar{x}_h = \frac{\sum_{i=1}^{n_h} x_{hi}}{n_h} \quad \text{y} \quad W_h = \frac{N_h}{N}$$

El cálculo de  $\bar{x}_h$  y de  $\bar{x}_{est}$  puede observarse en el cuadro 5.12.

Cuadro 5.12

Cálculo de la media del estrato y media ponderada

Indicador de estrato	$N_h$	$n_h$	$W_h$	$\bar{x}_h$	$\bar{x}_{est}$
$h=1$	31	2	0.246	9	2.2
$h=2$	16	1	0.127	14	1.8
$h=3$	32	3	0.254	22	5.6
$h=4$	17	2	0.135	30	4.0
$h=5$	17	3	0.135	42	5.7
$h=6$	9	3	0.071	70	5.0
Estrato cierto	4	4	0.032	185	5.9
Suma	126	18	1.000		30.2

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Rama 22, datos de 1989. (Ver cuadro 3, apéndice B).

Y el cálculo del estadístico con asignación óptima es:

$$\hat{X}_{op} = N \bar{x}_{est} = 126 \times 30.2 = 3,805 \text{ mill. de pesos.}$$

Para valuar la precisión del estimador se calcula el error estándar del estadístico, con la ecuación:

$$\sigma(\hat{Y}_{op}) = \frac{\left( \sum_{h=1}^r N_h \sigma_h \right)^2}{n} - \sum_{h=1}^r N_h \sigma_h^2$$

cuyo cálculo se obtiene desarrollando los resultados del cuadro 5.13.

**Cuadro 5.13**

Cálculos para el error estándar del estadístico  $\hat{X}_{op}$ .

Indicador de estrato	$N_h$	$f_h$	$\sigma^2_h$	$N_h \sigma^2_h$	$N_h \sigma_h$
$h=1$	31	0.065	2	74	48
$h=2$	16	0.063	2	35	24
$h=3$	32	0.094	9	273	93
$h=4$	17	0.118	9	159	52
$h=5$	17	0.177	24	415	84
$h=6$	9	0.333	70	627	75
Estrato cierto	4	1.000	3,548	14193	238
Total	126			15,776	614

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Rama 22, datos de 1989. (Ver cuadro 3, apéndice B).

El cálculo del error estándar es:

$$\sigma(\hat{X}_{op}) = \sqrt{\frac{(614)^2}{18} - 15,776} = 72.$$

El intervalo de confianza de  $X$  es:

$$\begin{aligned} (\hat{X}_{op} - k\sigma(\hat{X}_{op})) < X < \hat{X}_{op} + k\sigma(\hat{X}_{op}), \\ (3,798 - 1.64 \times 72 < X < 3,798 + 1.64 \times 72), \\ (3,680 < X < 3,916). \end{aligned}$$

El valor real de  $X$  puede ser desde 3,680 hasta 3,916 millones de pesos como estimación del valor total de  $X$ .

Tanto en MAS como en MAE, el valor que se desea medir es el estimador  $\hat{Y}$  y no  $\hat{X}$ , pero para efectos de comparación de métodos es suficiente la estimación de la variable correlacionada.

### 5.3. Comparación de resultados

Un principio de muestreo es seleccionar aquel método cuyo error estándar del estimador sea el menor. La precisión está en función inversa al error estándar, el mínimo error estándar equivale a la máxima precisión. De los métodos expuestos anteriormente, el error estándar del estimador del total de ingreso de la rama 22 es:

Para MAS:

$$\sigma^2(\hat{X})_{mas} = 116.5.$$

Y para MAE con asignación óptima:

$$\sigma^2(\hat{X})_{mae} = 72.$$

El error estándar del estimador en MAE es menor al error estándar del estimador en MAS; entonces, el método MAE con asignación óptima es más preciso que el MAS.

Además, la eficiencia de los estimadores es :

$$\sigma^2(\hat{X})_{mae} < \sigma^2(\hat{X})_{mas},$$

$$72 < 116.5.$$

O su eficiencia relativa:

$$Ef = \frac{\sigma^2(\hat{X})_{mae}}{\sigma^2(\hat{X})_{mas}} = \frac{72}{116.5} = 0.62 < 1.$$

La eficiencia relativa de  $\sigma^2(\hat{X})_{mae}$  con respecto a  $\sigma^2(\hat{X})_{mas}$  es menor que la unidad; entonces, el estimador de  $X$  es más eficiente con el procedimiento de MAE con asignación óptima que con el de MAS.

Los intervalos de confianza de los estimadores de  $X$  para MAS y MAE, respectivamente, son:

$$(3,715 < X < 4,097), \text{ y}$$

$$(3,680 < X < 3,916).$$

Se conoce el valor real de la variable ingreso, pues es la variable correlacionada a la variable de estimación (formación bruta de capital fijo):

$$X = 3,798 \text{ mill. de pesos.}$$

Entonces, se puede confirmar el error muestral para cada estimador:

$$\varepsilon = \hat{X} - X.$$

En MAS:

$$\varepsilon = 3,906 - 3,798 = 108 \text{ mill. de pesos.}$$

En MAE:

$$\varepsilon = 3,805 - 3,798 = 7 \text{ mill. de pesos.}$$

El error muestral es significativamente mayor en MAS que en MAE, así la estimación con el proceso de MAE con asignación óptima produce menor error de muestreo que con MAS.

En resumen, el estimador del total para la rama 22 es más preciso, mas eficiente y de menor error muestral con el método de MAE con asignación óptima que con el método de MAS.

Las comparaciones están de acuerdo a la muestra seleccionada en cada uno de los procedimientos para dicha rama.

## CONCLUSIONES

Describir el criterio de Dalenius en la construcción de estratos para estimar el total poblacional y compararlo con el muestreo aleatorio simple ha sido el objetivo de la investigación.

El criterio de Dalenius o regla  $cum\sqrt{f}$  es la aproximación a la solución exacta de estratificación óptima. El procedimiento detallado se encuentra desarrollado en el capítulo cuatro.

Dalenius supone como estimador a la media poblacional de la variable de estimación; sin embargo, se concluyó que para los fines de aplicación, el estimador es el total poblacional y la variable de estratificación es la correlacionada a la variable de estimación.

El supuesto sobre la solución exacta cuando se estima al total poblacional es la misma ecuación que cuando se estima a la media poblacional. Se dedujo, que si la solución exacta sólo depende de los parámetros de cada estrato (media y varianza) y no depende de los ponderadores poblacionales, entonces es indistinto su uso para los estimadores del total y la media.

Para el supuesto sobre la variable correlacionada a la variable de estimación, Dalenius sugiere que: "la regla  $cum\sqrt{f(y)}$  empleada a una variable de estratificación debe dar una estratificación eficiente para la variable de estimación con la que tiene una alta correlación".

## CONCLUSIONES

115

Los datos de ingreso asociados a las empresas de la industria de refrescos y aguas gaseosas (rama 22) ejemplificaron el proceso del censo de Dabrutec y su optimidad en el muestreo aleatorio estratificado en comparación con el muestreo aleatorio simple.

En la aplicación de dicho proceso, se creó un programa de cómputo (finado en Lotus-123) útil para estratificar variables de hasta 999 elementos. Para el muestreo aleatorio simple, el programa calcula el tamaño de muestra para el muestreo aleatorio estratificado, el programa estratifica según el criterio de Latham con divisiones de estratos de dos a nueve, calcula el tamaño de muestra para cada uno de las divisiones, y calcula la tabla de intravarianza e intervarianza para seleccionar el óptimo número de estratos.

"El principio a seguir en la selección de un método de muestreo es el de obtener el resultado deseado con la precisión requerida a un costo mínimo, o sea, con la máxima precisión o mínimo error al costo dado, o sea, el costo más bajo de los recursos disponibles."

En el cálculo del tamaño muestral por errores humanos, en todos los momentos, se fueron considerando la variación sistemática, la precisión del muestreo y el costo muestral.

En la comparación de los dos métodos, se concluyó que para la rama 22 se obtiene un muestreo más eficiente y de menor costo muestral por el método del muestreo aleatorio estratificado con asignación óptima que por el método del muestreo aleatorio simple.



Los datos de ingreso asociados a las empresas de la industria de refrescos y aguas gaseosas (rama 22) ejemplificaron el proceso del criterio de Dalenius y su optimidad en el muestreo aleatorio estratificado en comparación con el muestreo aleatorio simple.

En la aplicación de dicho procesos, se creó un programa de cómputo (macro en Lotus-123) útil para estratificar variables de hasta 900 elementos. Para el muestreo aleatorio simple, el programa calcula el tamaño de muestra; para el muestreo aleatorio estratificado, el programa: estratifica según el criterio de Dalenius con iteraciones de estratos de dos a nueve, calcula el tamaño de muestra para cada una de las iteraciones y calcula la tabla de intravarianza e intervarianza para seleccionar el número óptimo de estratos.

"El principio a seguir en la selección de un método de muestreo es el de obtener el resultado deseado con la precisión requerida o error aceptado a costo mínimo, o con la máxima precisión o mínimo error al costo dado, haciendo el uso más eficaz de los recursos disponibles"<sup>1</sup>.

En el cálculo del tamaño muestral con ambas técnicas, los costos de muestreo no fueron considerados. Se compara solamente la precisión del estimador y el error muestral.

En la comparación de los dos métodos, el estimador del total para la rama 22 es más preciso, más eficiente y de menor error muestral con el método del muestreo aleatorio estratificado con asignación óptima que con el método del muestreo aleatorio simple.

---

<sup>1</sup> SUKHATME, Pandurang V.: *Sampling theory of survey with applications*, The Indian Society of Agricultural, New Delhi, India, 1953, p. 3.

Por lo tanto, el mejor método de muestreo para la rama 22 es el muestreo aleatorio estratificado.

El muestreo aleatorio estratificado se aplicó con el criterio de Dalenius a las 63 ramas del marco muestral de la encuesta de acervos y formación de capital, y se obtuvieron resultados análogos a los de esta investigación.

El estudio se hizo sobre una variable económica, así también puede extenderse a otro tipo de variables: demográficas, sociales, etc., siempre y cuando estén en una escala de intervalo.

## BIBLIOGRAFIA

1. AOYAMA, H., *A study of the stratified random sampling*, Annals of Mathematical Statistics, V. 6 (1), 1954, pp. 1-36.
2. AZORIN, Francisco: *Curso de muestreo y aplicaciones*, México, Aguilar, 1971.
3. COCHRAN, William G.: *Técnicas de muestreo*, México, Continental, 1981.
4. COCHRAN, William G.: *Comparison of Methods for Determining Stratum Boundaries*, Bulletin of International Statistical Institute, 32 (2), Tokyo, 1961, pp. 345-358.
5. DALENIUS, Tore: *The problem of optimum stratification*, E.U., Scandinavian Journal of Statistics (33), 1950, pp. 203-213.
6. DALENIUS, Tore y GURNEY, M.: *The problem of optimum stratification II*, E.U., Scandinavian Journal of Statistics (34), 1951, pp. 133-148.
7. DALENIUS, T. y HODGES, S. P.: *Minimum variance stratification*, Journal American Statistics Association, Vol. 54, 1959, pp. 88-101.
8. HANSEN, M. L., HURWITZ, W. N., y MADOW, W.G.: *Sample surveys methods and theory*, Vol. 1, New York: John Wiley and Sons, Inc., 1953.
9. HESS, I., SETHI, V. K., BALAKRISHNAN, T. R.: *Stratification: A practical investigation*, Journal American Association, Vol. 61, 1966, pp. 74-90.
10. KISH, Leslie: *Muestreo de encuestas*, México, Trillas, 1979.
11. SUKHATME, Pandurang V.: *Sampling theory of survey with applications*, The Indian Society of Agricultural, New Delhi, India, 1953.
12. SPIEGEL, Murray R.: *Estadística*, México, McGraw-Hill, 1994.
13. MENDENHALL, SCHEAFFER, WACKERLY: *Estadística Matemática con Aplicaciones*. México, Grupo Editorial Iberoamericana, 1986.

## APÉNDICES

### Apéndice A: Algunos parámetros de una función de distribución

Es posible definir una serie de parámetros que nos permitan describir las características de una distribución de probabilidades.

Para una distribución de probabilidades poblacional, dichos parámetros se denominan **parámetros poblacionales**, dado que consideran a toda una población.

Y para una distribución de probabilidades muestral, se denominan **parámetros muestrales** porque sólo consideran a la muestra seleccionada y son usados para inferir a los parámetros poblacionales. Cuando esto último sucede, los parámetros muestrales son denominados estimadores.

Los parámetros son características que describen una distribución de datos, existen parámetros de tendencia central, parámetros de dispersión y otros parámetros como el coeficiente de asimetría y el coeficiente de curtosis.

Los parámetros de tendencia central	Los parámetros de dispersión	Otros Parámetros
<ul style="list-style-type: none"><li>• La moda</li><li>• La mediana</li><li>• El valor esperado o la media</li></ul>	<ul style="list-style-type: none"><li>• El rango</li><li>• La varianza</li><li>• La desviación estándar</li><li>• El coeficiente de variación</li></ul>	<ul style="list-style-type: none"><li>• El coeficiente de correlación</li><li>• El coeficiente de asimetría</li><li>• El coeficiente de curtosis</li></ul>

A continuación, una breve descripción de cada uno.

**Moda ( $Mo$ )**

"La moda de un conjunto de números es el valor que ocurre con mayor frecuencia; es decir, el valor más frecuente. La moda puede no existir, e incluso no ser única en caso de existir".<sup>1</sup>

**Mediana ( $Me$ )**

La mediana de una distribución de frecuencias "es el valor central o la media de los valores centrales".<sup>2</sup>

También se define como el valor de la variable que corresponde al 50% de la función de distribución acumulada.

**Valor esperado  $E(Y)$** 

"En general se define el valor esperado de una variable aleatoria  $Y$  como sigue: Sea la variable aleatoria discreta con valores posibles  $(y_1, y_2, \dots, y_N, \dots)$  y

---

<sup>1</sup> SPIEGEL, Murray R.: *Estadística*, México, McGraw-Hill, 1994, p. 63.

<sup>2</sup> *Idem*, p. 63.

$p(y_1), p(y_2), \dots, p(y_N), \dots$  las probabilidades de seleccionar  $(y_1, y_2, \dots, y_N, \dots)$ , respectivamente. Entonces, el valor esperado de  $\mathbf{Y}$ ,  $E(Y)$ , está definido por<sup>3</sup>:

$$E(Y) = \sum_{i=1}^N y_i p(y_i).$$

Donde la suma contempla los valores posibles para  $\mathbf{Y}$ . En el caso de implicar a los  $N$  valores, el valor esperado se llama media poblacional, también llamado promedio ponderado.

Cuando la muestra toma  $n$  valores observados, se llama valor esperado (o promedio ponderado) de la muestra y se denota por:

$$\bar{y} = \sum_{i=1}^n y_i p(y_i).$$

#### Propiedades del valor esperado:

Supóngase que  $a$  es una constante y  $\mathbf{Y}$  es una variable aleatoria, entonces:

i) El valor esperado de una constante,  $\mathbf{Y} = a$ :

$$E(\mathbf{Y}) = E(a) = a,$$

ii) El valor esperado de  $(a\mathbf{Y})$ :

$$E(a\mathbf{Y}) = aE(\mathbf{Y}),$$

---

<sup>3</sup> MENDENHALL, SCHEAFFER, WACKERLY: *Estadística Matemática con Aplicaciones* México, Grupo Editorial Iberoamericana, 1986, p. 97.

iii) El valor esperado de  $(\mathbf{Y} + a)$ :

$$E(\mathbf{Y} + a) = E(\mathbf{Y}) + a.$$

### Rango ( $Ra$ )

"Es la diferencia entre el máximo valor observado y el mínimo valor observado"<sup>4</sup> de la variable  $\mathbf{Y}$ . Sea  $y_1, y_2, \dots, y_N$  los valores de  $\mathbf{Y}$  asociados a los elementos de la población. Cuando  $y_{(1)} < y_{(2)} < \dots < y_{(N)}$ , entonces  $y_{(a)}$  es una estadística de orden, y

$$Ra = y_{(a)} - y_{(1)}.$$

### Varianza $\sigma^2$

"La varianza de una variable aleatoria  $\mathbf{Y}$  está definida como el valor esperado de

$(Y - E(Y))^2$ "<sup>5</sup>. Es decir:

$$\sigma^2 = E[\mathbf{Y} - E(\mathbf{Y})]^2.$$

---

<sup>4</sup> SPIEGEL, M., 1994, *op. cit.*, p. 91.

<sup>5</sup> MENDENHALL, SCHEAFFER, WACKERLY, 1986, *op. cit.*, p. 99.

A partir de la definición de valor esperado:

$$\sigma^2 = \sum_{i=1}^N [y_i - E(\mathbf{Y})]^2 p(y_i),$$

$$\sigma^2 = \sum_{i=1}^N (y_i - \bar{Y})^2 p(y_i) = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{Y})^2,$$

$\sigma^2$  es llamada la varianza poblacional. Al numerador se le llama usualmente la suma de las desviaciones cuadradas.

En el caso de una muestra que toma  $n$  datos, la varianza muestral es

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2.$$

Para muestras muy grandes es indiferente usar  $(n - 1)$  que  $n$  como divisor.

#### Propiedades de la varianza:

Sea  $a$  una constante cualquiera y  $\mathbf{Y}$  una variable aleatoria:

i) La varianza de  $\mathbf{Y} = a$ , es

$$\sigma^2 = \sigma^2(a) = 0,$$

ii) La varianza de  $(a \mathbf{Y})$ , es

$$\sigma^2(a \mathbf{Y}) = a^2 \sigma^2,$$



esta ecuación muestra que cuando la escala de  $Y$  cambia por  $a$ , la varianza se incrementa por  $a^2$  veces.

iii) La varianza de  $(Y + a)$ , es

$$\sigma^2(Y + a) = \sigma^2,$$

esta ecuación muestra que  $\sigma^2$  es independiente del origen.

#### Desviación estándar $\sigma$ , $s$

"Es la raíz cuadrada positiva de la varianza"<sup>6</sup>.

La desviación estándar poblacional es:

$$\sigma = \sqrt{\sigma^2}.$$

La desviación estándar muestral es:

$$s = \sqrt{s^2}.$$

#### Coefficiente de variación ( $C$ )

Es una magnitud adimensional y permite calcular un porcentaje de dispersión con respecto a la media. Está en términos relativos.

---

<sup>6</sup> *Ibidem*.

El coeficiente de variación de una distribución es definido como la desviación estándar dividido por la media.

$$C^2 = \frac{\sum_{i=1}^N \left( \frac{y_i - \bar{Y}}{\bar{Y}} \right)^2}{N},$$

$$C^2 = \frac{\sum_{i=1}^N \frac{(y_i - \bar{Y})^2}{N}}{\bar{Y}^2} = \frac{\sigma^2}{\bar{Y}^2}.$$

$C^2$  es llamada varianza relativa de la distribución. Y

$$C = \sqrt{C^2} = \frac{\sigma}{\bar{Y}},$$

$C$  es llamado coeficiente de variación.

### Coeficiente de correlación ( $\rho$ )

"El coeficiente de correlación de rangos de Spearman ( $\rho$ ) se calcula al utilizar los rangos como las mediciones aparejadas de dos variables,  $x$  y  $y$ "<sup>7</sup>. Así:

$$\rho = \frac{n \sum_{i=1}^n x_i y_i - \left( \sum_{i=1}^n x_i \right) \left( \sum_{i=1}^n y_i \right)}{\sqrt{\left[ \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 \right] \left[ \sum_{i=1}^n y_i^2 - \left( \sum_{i=1}^n y_i \right)^2 \right]}}$$

<sup>7</sup> Idem, p. 649.

**Coefficiente de curtosis (K)**

"La curtosis mide cuán puntiaguda es una distribución, en general por referencia a la función distribución normal"<sup>8</sup>. Para la distribución normal, la curtosis tiene un valor de 3.

En casos especiales, si:

- $K < 3$ , la distribución es platicúrtica,
- $K = 3$ , la distribución es mesocúrtica, y
- $K > 3$ , la distribución es leptocúrtica

**Coefficiente de asimetría (As)**

"Se conoce como asimetría o sesgo, a la medida en la cual se aparta de la simetría una función de distribución"<sup>9</sup>. El coeficiente de asimetría en la distribución normal tiene el valor de cero.

$$As = \frac{\bar{Y} - Mo}{\sigma}$$

Si la curva de frecuencias de una distribución tiene a la derecha una cola más larga que a la izquierda, se dice sesgada a la derecha, o de sesgo positivo,  $As > 0$ . En caso contrario, sesgada a la izquierda o de sesgo negativo,  $As < 0$ . En el caso de  $As = 0$ , entonces la distribución es simétrica.

---

<sup>8</sup> SPIEGEL, M., 1994. *op. cit.*, p. 119.

<sup>9</sup> *Idem*, p 118.

**Apéndice B: Tablas de referencia**

Cuadro I

**CLASIFICACIÓN MEXICANA DE ACTIVIDADES ECONÓMICAS 1970.**

<b>GRAN DIVISION 1</b>	<b>AGROPECUARIO, SILVICULTURA Y PESCA</b>
RAMA 1	AGRICULTURA
RAMA 2	GANADERÍA
RAMA 3	SILVICULTURA
RAMA 4	CAZA Y PESCA
<b>GRAN DIVISION 2</b>	<b>MINERÍA</b>
RAMA 5	EXTRACCIÓN Y BENEFICIO DE CARBÓN Y GRAFITO Y FAB DE SUS DERIVADOS
RAMA 6	EXTRACCIÓN DE PETRÓLEO
RAMA 7	EXTRACCIÓN Y BENEFICIO DE MINERAL DE HIERRO.
RAMA 8	EXTRACCIÓN, BENEFICIO Y FUNDICIÓN Y REFINACIÓN DE MINERAL METÁLICOS NO FERROSOS.
RAMA 9	EXPLOTACIÓN DE CANTERAS
RAMA 10	EXTRACCIÓN Y BENEFICIO DE OTROS MINERALES NO METÁLICOS

Continuación de Cuadro I

## CLASIFICACIÓN MEXICANA DE ACTIVIDADES ECONÓMICAS 1970.

GRAN DIVISIÓN 3	INDUSTRIA MANUFACTURERA
DIVISIÓN I	PRODUCTOS ALIMENTICIOS, BEBIDAS Y TABACO
RAMA 11	CARNES Y LÁCTEOS
RAMA 12	PREPARACIÓN DE FRUTAS Y LEGUMBRES
RAMA 13	MOLIENDA DE TRIGO
RAMA 14	MOLIENDA DE NIXTAMAL
RAMA 15	BENEFICIO Y MOLIENDA DE CAFÉ
RAMA 16	AZÚCAR
RAMA 17	ACEITES Y GRASAS COMESTIBLES
RAMA 18	ALIMENTOS PARA ANIMALES
RAMA 19	OTROS PRODUCTOS ALIMENTICIOS
RAMA 20	BEBIDAS ALCOHÓLICAS
RAMA 21	CERVEZA Y MALTA
RAMA 22	REFRESCOS Y AGUAS GASEOSAS
RAMA 23	TABACO
DIVISIÓN II	TEXTILES, PRENDAS DE VESTIR E INDUSTRIA DEL CUERO
RAMA 24	HILADOS Y TEJIDOS DE FIBRAS BLANDAS
RAMA 25	HILADOS Y TEJIDOS DE FIBRAS DURAS
RAMA 26	OTRAS INDUSTRIAS TEXTILES

Continuación de Cuadro I

## CLASIFICACIÓN MEXICANA DE ACTIVIDADES ECONÓMICAS 1970.

RAMA 27	PRENDAS DE VESTIR
RAMA 28	CUERO Y CALZADO
DIVISIÓN III	INDUSTRIA DE LA MADERA Y PRODUCTOS DE MADERA
RAMA 29	ASERRADEROS, TRIPLAY Y TABLEROS
RAMA 30	OTROS PRODUCTOS DE MADERA Y CORCHO
DIVISIÓN IV	PRODUCTOS DE PAPEL
RAMA 31	PAPEL, PRODUCTOS DE PAPEL E IMPRENTA
RAMA 32	IMPRENTAS Y EDITORIALES
DIVISIÓN V	SUSTANCIAS QUÍMICAS, DERIVADOS DEL PETRÓLEO
RAMA 33	PETRÓLEO Y DERIVADOS
RAMA 34	PETROQUÍMICA BÁSICA
RAMA 35	QUÍMICA BÁSICA
RAMA 36	ABONOS Y FERTILIZANTES
RAMA 37	RESINAS SINTÉTICAS Y ARTIFICIALES
RAMA 38	PRODUCTOS FARMACÉUTICOS
RAMA 39	JABONES, DETERGENTES Y COSMÉTICOS
RAMA 40	OTROS PRODUCTOS QUÍMICOS
RAMA 41	PRODUCTOS DE HULE
RAMA 42	ARTÍCULOS DE PLÁSTICO

Continuación de Cuadro I

## CLASIFICACIÓN MEXICANA DE ACTIVIDADES ECONÓMICAS 1970.

DIVISIÓN VI	PRODUCTOS DE MINERALES NO METÁLICOS EXCEPTUANDO DERIVADOS DEL PETRÓLEO Y EL CARBÓN
RAMA 43	VIDRIO Y PRODUCTOS DE VIDRIO
RAMA 44	CEMENTO
RAMA 45	PRODUCTOS A BASE DE MINERALES NO METÁLICOS
DIVISIÓN VII	INDUSTRIAS METÁLICAS BÁSICAS
RAMA 46	INDUSTRIAS BÁSICAS DEL HIERRO Y ACERO
RAMA 47	INDUSTRIAS BÁSICAS DE METALES NO FERROSOS
DIVISIÓN VIII	PRODUCTOS METÁLICOS, MAQUINARIA Y EQUIPO
RAMA 48	MUEBLES METÁLICOS
RAMA 49	PRODUCTOS METÁLICOS ESTRUCTURALES
RAMA 50	OTROS PRODUCTOS METÁLICOS, EXCEPTO MAQUINARIA
RAMA 51	MAQUINARIA Y EQUIPO NO ELÉCTRICO
RAMA 52	MAQUINARIA Y APARATOS ELÉCTRICOS
RAMA 53	APARATOS ELECTRODOMÉSTICOS
RAMA 54	EQUIPOS Y APARATOS ELECTRÓNICOS
RAMA 56	AUTOMÓVILES
RAMA 57	CARROCERÍAS Y PARTES PARA AUTOMÓVIL
RAMA 58	EQUIPO Y MATERIAL DE TRANSPORTE

Continuación de Cuadro I

## CLASIFICACIÓN MEXICANA DE ACTIVIDADES ECONÓMICAS 1970.

DIVISIÓN IX	OTRAS INDUSTRIAS MANUFACTURERAS
RAMA 59	OTRAS INDUSTRIAS MANUFACTURERAS
GRAN DIVISIÓN 4	CONSTRUCCIÓN
RAMA 60	CONSTRUCCIÓN
GRAN DIVISIÓN 5	ELECTRICIDAD, GAS Y AGUA
RAMA 61	ELECTRICIDAD
GRAN DIVISIÓN 6	COMERCIO, RESTAURANTES Y HOTELES
RAMA 62	COMERCIO
RAMA 63	RESTAURANTES Y HOTELES
GRAN DIVISIÓN 7	TRANSPORTE Y COMUNICACIONES
RAMA 64	TRANSPORTE
RAMA 65	COMUNICACIONES
GRAN DIVISIÓN 8	SERVICIOS FINANCIEROS, SEGUROS Y BIENES INMUEBLES
RAMA 66	SERVICIOS FINANCIEROS
RAMA 67	ALQUILER DE INMUEBLES



Continuación de Cuadro I

## CLASIFICACIÓN MEXICANA DE ACTIVIDADES ECONÓMICAS 1970.

<b>GRAN DIVISIÓN 9</b>	<b>SERVICIOS PROFESIONALES</b>
RAMA 68	SERVICIOS PROFESIONALES
RAMA 69	SERVICIOS DE EDUCACIÓN
RAMA 70	SERVICIOS MÉDICOS
	<b>OTROS SERVICIOS</b>
RAMA 71	SERVICIOS DE ESPARCIMIENTO
RAMA 72	SERVICIOS DE REPARACIÓN

**Cuadro II**  
**MÉTODOS DE ESTIMACIÓN POR RAMA: MUESTREO O CENSO,**  
**Y TAMAÑO DE POBLACIÓN PARA CADA UNA.**

R a m a		Muestreo	Censo	Tamaño de población
5	Extracción y beneficio de carbón y grafito y fabricación de sus derivados		/	10
7	Extracción y beneficio de mineral de hierro		/	8
8	Extracción, beneficio, fundición y refinación de minerales metálicos no ferrosos		/	38
9	Explotación de canteras y extracción de arena, grava y arcilla		/	9
10	Extracción y beneficio de otros minerales no metálicos		/	14
11	Carnes y lácteos	/		123
12	Preparación de frutas y legumbres	/		46
13	Molenda de trigo	/		52
14	Molenda de maíz	/		15
15	Beneficio y molenda de café		/	20
16	Azúcar	/		77
17	Aceites y grasas comestibles	/		28
18	Alimentos para animales	/		25
19	Otros productos alimenticios	/		62
20	Bebidas alcohólicas	/		25
21	Cerveza y maiz	/		18
22	Refrescos y aguas gasificadas	/		23
23	Tabaco	/		4
24	Alambres y teleros de fibras blancas	/		12
25	Alambres y teleros de fibras puras	/		6
26	Otros productos textiles	/		20
27	Prendas de vestir	/		20
28	Cuero y calzados	/		20
29	Aserradero, muela y tabernos	/		16
30	Otros productos de madera y corcho	/		62
31	Papel y cartón	/		11
32	Imprentas y editoriales	/		10
33	Química básica	/		1
34	Química sintética y artificiales	/		12
35	Producción farmacéutica	/		12

Fuente: Muestreo de la Encuesta sobre Población de Censos de 1980, Censo de 1980.

Continuación Cuadro II

**MÉTODOS DE ESTIMACIÓN POR RAMA: MUESTREO O CENSO,  
Y TAMAÑO DE POBLACIÓN PARA CADA UNA.**

	R a m a	Muestreo	Censo	Tamaño de población
39	Jabones, detergentes y cosméticos	X		39
40	Otros productos químicos	X		146
41	Productos de hule	X		21
42	Artículos de plástico	X		107
43	Vidrio y productos de vidrio	X		26
44	Cemento		X	19
45	Productos a base de minerales no metálicos	X		79
46	Industria básica de hierro y acero	X		104
47	Industrias básicas de metales no ferrosos	X		30
48	Muebles metálicos	X		21
49	Productos metálicos estructurales	X		8
50	Otros productos metálicos, excepto maquinaria	X		133
51	Maquinaria y equipo no eléctrico	X		142
52	Maquinaria y aparatos eléctricos	X		47
53	Aparatos electro - domésticos	X		15
54	Aparatos y equipos electrónicos	X		56
55	Aparatos y equipos eléctricos	X		121
56	Automóviles		X	12
57	Carrocerías, motores, partes y accesorios para automóviles	X		85
58	Equipo y material de transporte	X		16
59	Otras industrias manufactureras	X		66
60	Construcción		X	138
61	Electricidad, gas y agua		X	2
63	Restaurantes y hoteles	X		87
64	Transporte	X		68
65	Comunicaciones		X	4
66	Servicios financieros, seguros y bienes inmuebles	X		248
67	Alquiler de inmuebles	X		118
70	Servicios médicos		X	12
71	Servicios de esparcimiento	X		39
72	Otros servicios		X	22

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Datos de 1989.

Cuadro III

INGRESO ASOCIADO A LA EMPRESA *i* PARA LA POBLACIÓN DE LA RAMA 22.<sup>10</sup>

$X_i$ = Ingreso asociado a la empresa <i>i</i>			
$X_1 = 6.7$	$X_{33} = 12.3$	$X_{65} = 22.9$	$X_{97} = 36.7$
$X_2 = 6.7$	$X_{34} = 12.3$	$X_{66} = 23.6$	$X_{98} = 36.9$
$X_3 = 6.8$	$X_{35} = 12.4$	$X_{67} = 24.1$	$X_{99} = 37.2$
$X_4 = 6.8$	$X_{36} = 13.5$	$X_{68} = 24.2$	$X_{100} = 37.5$
$X_5 = 6.9$	$X_{37} = 13.7$	$X_{69} = 24.4$	$X_{101} = 37.8$
$X_6 = 6.9$	$X_{38} = 13.7$	$X_{70} = 24.6$	$X_{102} = 38.9$
$X_7 = 7.0$	$X_{39} = 13.8$	$X_{71} = 24.6$	$X_{103} = 39.0$
$X_8 = 7.1$	$X_{40} = 14.3$	$X_{72} = 24.6$	$X_{104} = 40.4$
$X_9 = 7.2$	$X_{41} = 14.4$	$X_{73} = 24.7$	$X_{105} = 40.8$
$X_{10} = 7.2$	$X_{42} = 14.8$	$X_{74} = 25.0$	$X_{106} = 41.9$
$X_{11} = 7.6$	$X_{43} = 14.9$	$X_{75} = 25.2$	$X_{107} = 43.5$
$X_{12} = 7.6$	$X_{44} = 15.7$	$X_{76} = 25.6$	$X_{108} = 43.8$
$X_{13} = 7.9$	$X_{45} = 16.2$	$X_{77} = 25.9$	$X_{109} = 45.5$
$X_{14} = 8.4$	$X_{46} = 16.2$	$X_{78} = 26.1$	$X_{110} = 45.7$
$X_{15} = 8.6$	$X_{47} = 16.6$	$X_{79} = 26.3$	$X_{111} = 48.2$
$X_{16} = 8.8$	$X_{48} = 16.8$	$X_{80} = 26.8$	$X_{112} = 51.2$
$X_{17} = 9.0$	$X_{49} = 16.8$	$X_{81} = 27.1$	$X_{113} = 52.0$
$X_{18} = 9.2$	$X_{50} = 16.9$	$X_{82} = 27.1$	$X_{114} = 59.4$
$X_{19} = 9.2$	$X_{51} = 18.5$	$X_{83} = 27.3$	$X_{115} = 60.0$
$X_{20} = 9.3$	$X_{52} = 18.5$	$X_{84} = 27.6$	$X_{116} = 66.9$
$X_{21} = 9.4$	$X_{53} = 18.9$	$X_{85} = 27.6$	$X_{117} = 67.0$
$X_{22} = 9.5$	$X_{54} = 19.3$	$X_{86} = 27.7$	$X_{118} = 70.8$
$X_{23} = 9.9$	$X_{55} = 20.0$	$X_{87} = 28.3$	$X_{119} = 71.6$
$X_{24} = 10.0$	$X_{56} = 20.1$	$X_{88} = 28.7$	$X_{120} = 71.7$
$X_{25} = 10.1$	$X_{57} = 20.6$	$X_{89} = 29.4$	$X_{121} = 79.0$
$X_{26} = 10.1$	$X_{58} = 20.7$	$X_{90} = 29.6$	$X_{122} = 85.5$
$X_{27} = 10.6$	$X_{59} = 20.8$	$X_{91} = 30.3$	$X_{123} = 115.7$
$X_{28} = 11.0$	$X_{60} = 21.1$	$X_{92} = 30.6$	$X_{124} = 174.0$
$X_{29} = 11.0$	$X_{61} = 21.2$	$X_{93} = 31.4$	$X_{125} = 188.9$
$X_{30} = 11.1$	$X_{62} = 22.2$	$X_{94} = 32.6$	$X_{126} = 260.6$
$X_{31} = 11.4$	$X_{63} = 22.5$	$X_{95} = 36.5$	
$X_{32} = 12.1$	$X_{64} = 22.5$	$X_{96} = 36.6$	

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Rama 22, datos de 1989.

<sup>10</sup> Por razones de confidencialidad no se exponen los nombres de cada empresa.

## Cuadro IV

**FORMACIÓN BRUTA DE CAPITAL POR RAMA  
Y OTROS RESULTADOS DE LA ENCUESTA DE 1989.**  
(Miles de pesos de 1980)

RAMA	ACERVOS BRUTOS	ACERVOS NETOS	DEPRECIACIÓN ACUMULADA	FORMACIÓN BRUTA	FORMACIÓN NETA	DEPRECIACIÓN ANUAL
5	17,768	7,420	10,347	1,317	230	1,087
7	26,438	14,489	11,949	3,618	1,916	1,702
8	92,762	55,009	37,753	5,724	-299	6,023
9	8,428	5,168	3,260	1,409	791	618
10	18,320	9,023	9,297	815	-226	1,041
11	15,883	7,327	8,556	1,216	-42	1,259
12	4,081	1,758	2,324	452	108	344
13	13,433	5,909	7,524	975	8	967
14	2,661	1,252	1,409	223	80	144
15	9,813	4,469	5,344	819	209	610
16	25,541	12,383	13,158	5,924	4,491	1,432
17	9,353	3,883	5,470	186	-339	525
18	1,542	684	858	144	29	115
19	23,244	10,313	12,930	739	-966	1,705
20	5,756	2,182	3,574	203	-114	317
21	25,471	11,322	14,149	1,119	-97	1,216
22	21,017	8,108	12,909	1,268	-730	1,998
23	4,820	2,164	2,656	370	87	284
24	43,554	17,239	26,315	1,875	-1,214	3,089
25	7,821	3,463	4,358	70	-236	306
26	3,249	1,505	1,744	157	-33	190
27	11,815	4,717	7,097	997	-476	1,473
28	6,184	2,973	3,211	455	12	443
29	6,954	2,095	4,859	167	-395	562
30	3,579	1,098	2,481	149	-134	283
31	42,901	20,599	22,302	2,779	379	2,400
32	14,804	6,595	8,209	506	-669	1,175
35	61,171	24,857	36,315	1,439	-1,767	3,206
36	9,411	4,473	4,939	321	-240	561
37	44,404	17,947	26,457	1,565	-863	2,428

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Datos de 1989. Resultados preliminares a febrero de 1996.

Continuación de cuadro IV

**FORMACIÓN BRUTA DE CAPITAL POR RAMA  
Y OTROS RESULTADOS DE LA ENCUESTA DE 1989.**  
(Miles de pesos de 1980)

RAMA	ACERVOS BRUTOS	ACERVOS NETOS	DEPRECIACIÓN ACUMULADA	FORMACIÓN BRUTA	FORMACIÓN NETA	DEPRECIACIÓN ANUAL
38	46,681	26,580	20,101	4,223	1,267	2,955
39	14,335	7,208	7,128	1,151	252	899
40	14,011	7,143	6,868	939	-165	1,105
41	17,204	8,210	8,994	838	-1	839
42	16,335	6,779	9,556	562	-636	1,198
43	18,537	8,034	10,503	720	-633	1,353
44	55,978	24,087	31,891	738	-2,285	3,022
45	16,495	7,502	8,994	718	-462	1,180
46	172,212	76,718	95,494	6,929	-3,111	10,040
47	9,890	5,083	4,807	361	-247	608
48	3,847	1,422	2,426	108	-179	287
49	5,831	2,391	3,439	178	-236	414
50	25,476	10,980	14,496	1,404	-293	1,697
51	36,102	16,285	19,817	1,524	-947	2,471
52	10,312	4,055	6,257	387	-525	911
53	5,032	1,703	3,329	346	-68	414
54	19,294	9,639	9,654	970	-458	1,428
55	23,579	10,551	13,029	1,075	-518	1,593
56	82,499	40,009	42,490	3,907	-2,478	6,386
57	25,197	10,761	14,436	952	-737	1,689
58	8,704	4,088	4,616	183	-344	527
59	12,780	6,203	6,577	498	-168	666
60	25,984	9,995	15,989	1,542	-818	2,360
61	646,191	435,466	210,725	24,099	5,815	18,284
62	474,772	212,441	262,332	31,976	-2,817	34,793
63	51,420	30,074	21,346	1,288	-731	2,019
64	364,270	216,650	147,620	7,756	-7,564	15,321
65	136,218	74,308	61,910	16,850	8,317	8,533
66	115,516	68,678	46,839	15,842	6,462	9,380
67	102,763	38,526	64,237	5,355	-1,056	6,411
70	7,811	3,577	4,234	278	-97	375
71	18,351	10,007	8,344	1,754	469	1,285
72	15,977	8,146	7,632	1,391	-434	1,825

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Datos de 1989.  
Resultados preliminares a febrero de 1996.

**Cuadro V**  
**ALGUNOS PARÁMETROS POBLACIONALES DE**  
**LA VARIABLE INGRESO.**

Rama	Total	Media	Varianza	Coef. De Curtosis	Coef. De Asimetría
5	707,803	71	9,547	5.72	2.30
7	499,456	62	3,239	-0.60	0.83
8	4,541,329	120	47,858	7.22	2.70
9	250,428	28	1,063	0.68	1.53
10	1,099,639	79	17,651	9.64	2.99
11	4,088,434	41	6,166	30.16	5.08
12	1,475,188	32	820	4.18	1.96
13	3,351,370	36	5,512	32.64	5.23
14	629,019	42	3,539	11.57	3.28
15	768,577	38	7,473	19.21	4.35
16	1,593,327	32	424	2.30	1.32
17	601,371	24	239	0.98	1.24
18	1,450,087	50	3,851	3.48	1.95
19	3,264,632	49	9,888	22.07	4.39
20	872,756	44	7,195	6.58	2.77
21	3,339,989	239	107,108	0.36	1.47
22	3,798,291	30	1,161	21.69	4.14
23	806,991	202	30,781	-3.21	0.03
24	3,531,444	23	747	26.26	4.54
25	161,929	27	1,436	5.40	2.30
26	817,624	23	717	11.82	3.15
27	1,604,424	19	441	11.27	3.25
28	1,263,152	18	499	28.03	4.81
29	268,427	19	361	6.68	2.56
30	1,095,594	23	2,779	36.50	5.82
31	8,215,808	71	25,173	45.53	5.97
32	1,749,817	21	531	9.13	2.89
35	5,655,274	70	12,155	17.50	3.59
37	5,270,132	91	19,016	6.00	2.35
38	3,906,011	36	2,413	25.39	4.39

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Datos de 1989. Coeficientes calculados con el paquete de cómputo Micro SPSS para Windows.

Continuación de cuadro V

**ALGUNOS PARÁMETROS POBLACIONALES DE  
LA VARIABLE INGRESO.**

Rama	Total	Media	Varianza	Coef. De Curtosis	Coef. De Asimetría
39	3,770,188	97	34,823	5.85	2.65
40	4,641,194	32	3,163	73.00	7.65
41	1,801,643	86	20,724	5.80	2.42
42	3,008,776	28	1,997	34.85	5.18
43	2,694,044	104	8,289	-0.75	0.73
44	2,735,725	144	13,049	0.39	0.82
45	2,297,674	29	947	6.26	2.37
46	9,387,094	90	94,560	33.72	5.70
47	4,311,540	144	132,405	17.91	4.04
48	523,129	25	988	3.78	2.18
49	209,403	26	308	-1.38	0.43
50	4,687,500	35	3,674	39.49	5.59
51	5,469,448	39	14,539	98.80	9.36
52	1,861,315	40	2,205	6.79	2.54
53	615,079	41	873	-0.97	.74
54	1,939,844	35	3,886	20.30	4.43
55	5,018,236	41	5,337	23.91	4.54
56	15,056,406	1,255	2,118,910	-1.47	0.66
57	3,900,918	46	3,646	10.18	2.68
58	422,372	26	860	7.62	2.64
59	1,642,171	25	2,197	47.13	6.50
60	5,939,102	43	3,125	40.91	5.50
63	2,188,272	25	1,284	17.70	3.91
64	7,199,126	106	98,091	20.11	4.45
65	4,176,061	2,088	7,437,727	N.D.	N.D.
66	62,163,855	251	501,205	50.21	6.43
67	2,992,108	25	2,062	49.01	6.25
70	174,455	14	44	0.83	0.95
71	1,807,927	49	14,102	29.18	5.22
72	1,164,994	73	16,356	6.20	2.61

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Datos de 1989. Coeficientes calculados con el paquete de cómputo Micro SPSS para Windows.

Nota: N.D. son datos no disponibles para la rama.



**Cuadro VI**  
**TAMAÑO DE POBLACIÓN POR ESTRATO**  
**PARA CADA RAMA DE ACTIVIDAD ECONÓMICA**

Rama	Estrato										N
	1	2	3	4	5	6	7	8	9	10	
5	10	0	0	0	0	0	0	0	0	0	10
6	0	0	0	0	0	0	0	0	0	0	0
7	8	0	0	0	0	0	0	0	0	0	8
8	9	4	4	5	8	6	0	0	0	0	38
9	9	0	0	0	0	0	0	0	0	0	9
10	14	0	0	0	0	0	0	0	0	0	14
11	6	10	13	12	36	44	0	0	0	0	123
12	5	13	6	8	14	0	0	0	0	0	46
13	6	8	14	24	38	0	0	0	0	0	92
14	15	0	0	0	0	0	0	0	0	0	15
15	20	0	0	0	0	0	0	0	0	0	20
16	1	8	12	15	13	0	0	0	0	0	50
17	3	6	6	10	9	0	0	0	0	0	34
18	2	4	5	5	9	0	0	0	0	0	29
19	5	11	8	16	23	0	0	0	0	0	66
20	2	5	12	0	0	0	0	0	0	0	20
21	14	0	0	0	0	0	0	0	0	0	14
22	4	9	17	17	32	16	29	0	0	0	126
23	4	0	0	0	0	0	0	0	0	0	4
24	4	23	38	28	60	0	0	0	0	0	156
25	6	0	0	0	0	0	0	0	0	0	6
26	3	9	21	0	0	0	0	0	0	0	35
27	3	13	28	35	0	0	0	0	0	0	83
28	2	10	19	37	0	0	0	0	0	0	70
29	14	0	0	0	0	0	0	0	0	0	14
30	2	15	28	0	0	0	0	0	0	0	48
31	8	8	7	8	15	25	22	22	0	0	115
32	5	17	19	38	0	0	0	0	0	0	83
33	0	0	0	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0
35	6	7	8	9	12	11	20	0	0	0	74
36	0	0	0	0	0	0	0	0	0	0	0
37	2	9	9	14	24	0	0	0	0	0	58
38	3	5	10	10	13	16	22	29	0	0	108
39	4	5	4	6	9	11	0	0	0	0	39
40	3	13	23	23	35	48	0	0	0	0	146

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Datos de 1989.



Continuación de cuadro VI

**TAMAÑO DE POBLACIÓN POR ESTRATO  
PARA CADA RAMA DE ACTIVIDAD ECONÓMICA**

Rama	Estrato										N
	1	2	3	4	5	6	7	8	9	10	
41	21	0	0	0	0	0	0	0	0	0	21
42	2	12	14	35	44	0	0	0	0	0	107
43	1	6	5	7	7	0	0	0	0	0	26
44	19	0	0	0	0	0	0	0	0	0	19
45	3	9	11	16	11	29	0	0	0	0	79
46	4	9	10	16	29	33	0	0	0	0	104
47	4	7	7	10	0	0	0	0	0	0	30
48	21	0	0	0	0	0	0	0	0	0	21
49	8	0	0	0	0	0	0	0	0	0	8
50	4	10	12	27	28	52	0	0	0	0	133
51	6	14	25	27	64	0	0	0	0	0	142
52	3	5	5	11	12	9	0	0	0	0	47
53	15	0	0	0	0	0	0	0	0	0	15
54	2	9	12	12	21	0	0	0	0	0	56
55	3	10	8	17	12	28	38	0	0	0	121
56	12	0	0	0	0	0	0	0	0	0	12
57	2	8	9	7	13	22	23	0	0	0	85
58	2	3	4	6	0	0	0	0	0	0	16
59	3	7	16	10	28	0	0	0	0	0	66
60	3	11	12	20	28	38	25	0	0	0	138
61	2	0	0	0	0	0	0	0	0	0	2
62	0	0	0	0	0	0	0	0	0	0	0
63	2	10	12	26	38	0	0	0	0	0	87
64	4	6	7	11	18	22	0	0	0	0	68
65	4	0	0	0	0	0	0	0	0	0	4
66	15	14	12	14	18	16	35	41	82	0	248
67	2	12	14	37	53	0	0	0	0	0	118
68	0	0	0	0	0	0	0	0	0	0	0
69	0	0	0	0	0	0	0	0	0	0	0
70	12	0	0	0	0	0	0	0	0	0	12
71	3	5	6	8	13	0	0	0	0	0	39
72	2	6	3	5	0	0	0	0	0	0	22
<b>Total</b>	<b>322</b>	<b>414</b>	<b>529</b>	<b>661</b>	<b>787</b>	<b>426</b>	<b>214</b>	<b>92</b>	<b>82</b>	<b>0</b>	<b>3599</b>

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Datos de 1989.

Cuadro VII

TAMAÑO DE MUESTRA POR ESTRATO PARA  
CADA RAMA DE ACTIVIDAD ECONÓMICA

Rama	Estrato										n	
	1	2	3	4	5	6	7	8	9	10		
5	10	0	0	0	0	0	0	0	0	0	0	10
6	0	0	0	0	0	0	0	0	0	0	0	0
7	5	0	0	0	0	0	0	0	0	0	0	5
8	9	3	1	1	2	1	0	0	0	0	0	17
9	9	0	0	0	0	0	0	0	0	0	0	9
10	13	0	0	0	0	0	0	0	0	0	0	13
11	6	3	3	2	3	2	0	0	0	0	0	19
12	5	10	1	1	1	0	0	0	0	0	0	18
13	6	7	6	5	4	0	0	0	0	0	0	28
14	15	0	0	0	0	0	0	0	0	0	0	15
15	19	0	0	0	0	0	0	0	0	0	0	19
16	1	3	3	3	2	0	0	0	0	0	0	12
17	3	6	2	1	1	0	0	0	0	0	0	13
18	2	4	5	1	1	0	0	0	0	0	0	13
19	5	10	2	2	1	0	0	0	0	0	0	20
20	2	6	4	0	0	0	0	0	0	0	0	12
21	14	0	0	0	0	0	0	0	0	0	0	14
22	4	2	2	1	2	1	1	0	0	0	0	13
23	4	0	0	0	0	0	0	0	0	0	0	4
24	4	13	5	1	2	0	0	0	0	0	0	25
25	6	0	0	0	0	0	0	0	0	0	0	6
26	3	9	6	0	0	0	0	0	0	0	0	18
27	3	9	5	3	0	0	0	0	0	0	0	20
28	2	8	4	5	0	0	0	0	0	0	0	19
29	14	0	0	0	0	0	0	0	0	0	0	14
30	2	15	3	0	0	0	0	0	0	0	0	20
31	8	2	3	1	1	2	1	1	0	0	0	19
32	5	10	3	3	0	0	0	0	0	0	0	21
33	0	0	0	0	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0
35	6	2	3	2	1	1	1	0	0	0	0	16
36	0	0	0	0	0	0	0	0	0	0	0	0
37	2	9	3	2	1	0	0	0	0	0	0	17
38	3	1	2	1	1	1	1	1	0	0	0	11
39	4	5	2	2	2	1	0	0	0	0	0	16

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Datos de 1989.

Continuación de Cuadro VII

**TAMAÑO DE MUESTRA POR ESTRATO PARA  
CADA RAMA DE ACTIVIDAD ECONÓMICA**

Rama	Estrato										n
	1	2	3	4	5	6	7	8	9	10	
40	3	5	4	2	2	2	0	0	0	0	18
41	21	0	0	0	0	0	0	0	0	0	21
42	2	11	3	4	2	0	0	0	0	0	22
43	1	4	3	4	1	0	0	0	0	0	13
44	19	0	0	0	0	0	0	0	0	0	19
45	3	6	2	3	1	2	0	0	0	0	17
46	4	9	1	2	1	1	0	0	0	0	18
47	4	7	1	1	0	0	0	0	0	0	13
48	20	0	0	0	0	0	0	0	0	0	20
49	7	0	0	0	0	0	0	0	0	0	7
50	4	5	2	4	2	2	0	0	0	0	19
51	6	8	4	2	3	0	0	0	0	0	23
52	3	5	1	3	2	1	0	0	0	0	15
53	13	0	0	0	0	0	0	0	0	0	13
54	2	7	3	2	1	0	0	0	0	0	15
55	3	7	1	3	1	1	1	0	0	0	17
56	11	0	0	0	0	0	0	0	0	0	11
57	2	4	4	1	2	1	1	0	0	0	15
58	2	3	2	2	0	0	0	0	0	0	9
59	3	4	7	1	3	0	0	0	0	0	18
60	3	5	1	2	1	2	1	0	0	0	15
61	2	0	0	0	0	0	0	0	0	0	2
62	0	0	0	0	0	0	0	0	0	0	0
63	2	10	3	2	2	0	0	0	0	0	19
64	4	6	2	2	1	1	0	0	0	0	16
65	4	0	0	0	0	0	0	0	0	0	4
66	15	2	3	2	1	1	1	1	1	0	27
67	2	11	2	3	2	0	0	0	0	0	20
68	0	0	0	0	0	0	0	0	0	0	0
69	0	0	0	0	0	0	0	0	0	0	0
70	12	0	0	0	0	0	0	0	0	0	12
71	3	3	3	2	1	0	0	0	0	0	12
72	2	4	1	1	0	0	0	0	0	0	8
<b>Total</b>	<b>364</b>	<b>273</b>	<b>54</b>	<b>23</b>	<b>8</b>	<b>3</b>	<b>1</b>	<b>0</b>	<b>932</b>	<b>1449</b>	<b>371</b>

Fuente: Marco muestral de la Encuesta sobre Formación de Capital de Banco de México. Datos de 1989.

## Apéndice C: Macro en Lotus-123 para estratificar

Las macros en Lotus-123 son programas que se ejecutan con el comando ALT-"letra de macro" dentro de la paquetería de Lotus-123.

Para esta investigación se nombra a cada una de las macros ubicándola en el lugar que se indica en el extremo izquierdo. Los datos deben estar en la posición C19 de la hoja de cálculo. Se debe primero ejecutar las macros referentes a mae y por último a mas.

MACRO PARA MUESTREO ALEATORIO SIMPLE, CALCULA EL TAMAÑO DE MUESTRA.

WM MACRO PARA MUESTREO ALEATORIO SIMPLE

A1600<sup>11</sup>

GOTO)K64-TAMAYO DE MUESTRA EN MAS-(DOWN)"n"-(RIGHT)  
(D9\*2\*D4\*2\*D6)/(D8\*2+D9\*D4\*2\*D6)-

MACRO PARA MUESTREO ALEATORIO ESTRATIFICADO, FRONTERAS DE ESTRATOS Y ASIGNACIÓN DE TAMAÑO DE MUESTRA.

VA MACRO INICIAL PARA CREAR DISTRIBUCION DE FRECUENCIAS Y  $cum\sqrt{f(y)}$

A1004

{HOME})(DOWN 2)/RFF2-A3. F12-"CONF:-(DOWN)"K=-  
{DOWN}"MEDIA(Y) -(DOWN)"VARIANZA:-(DOWN)"DESV.  
ST.-(DOWN)"ERROR:-(DOWN)"N-  
-(GOTO)C19~/RNCINGRESO-(END)(DOWN)-  
(GOTO)D3~"80%-(DOWN)1.281(down)@AVG(INGRESO)-(DOWN)@VAR(INGRESO)-  
(down)@STD(INGRESO)-(DOWN)@SUM(INGRESO)\*.05-(DOWN)@COUNT(INGRESO)-

<sup>11</sup> Este número se refiere a la posición en la hoja de cálculo donde debe estar la macro.

```
(GOTO)C19-/DSRD (LEFT 2)(END)(DOWN)-PINGRESO-A-G
(GOTO)D19-/DFD19 .D500-@MIN(INGRESO)-5000-@MAX(INGRESO)+5000-
AVGFF9-/RNCINT(1)- (END)(DOWN)-
(RIGHT)WCS2--/C-(DOWN) (LEFT)(END)(DOWN)(RIGHT)-
(RIGHT)DF19 .F500-@MIN(INGRESO)+5000-@MAX(INGRESO)+5000-
(ENDD)(DOWN)(DOWN)/RE(LEFT 2)-(GOTO)F19-/RNCINT(1)- (END)(DOWN)(UP)-
(GOTO)F19-/DDINGRESO-INT(+)-
(RIGHT 2)@SORT(G19)-
/C-(DOWN) (LEFT)(END)(DOWN)(RIGHT)-(RIGHT)+H10-(DOWN)(UP)+(LEFT)-
/C-(DOWN) (LEFT)(END)(DOWN)(RIGHT)-(GOTO)C14-/C (DOWN 3)-
D14 .Z17-(DOWN)(RIGHT)RED15 b716- INTERVALOS DE-
(DOWN) CLASE-
(GOTO)G15-/FREC -(RIGHT)HAIZ-(RIGHT)'ACUM-
(GOTO)J10-/WCS3-(UP 4)'''-
(PUT D500 I505 1 2 ELEGIR LAS EMPRESAS QUE SE CONSIDEREN A CENSAR)-
(PUT D500 I506 1 4 MARQUE "X" PARA INDICAR HASTA DONDE SE ESTRATIFICA)-
(GOTO)D493-(WAIT @NOW+@TIME(0,0,8))
(HOME)
```

A1041

```
DIST/DDINGRESO-AB19 AB27-(RETURN)
DIST/DDINGRESO-AD19 AD27-(RETURN)
DIST/DDINGRESO-AF19 AF27-(RETURN)
DIST/DDINGRESO-AH19 AH27-(RETURN)
DIST/DDINGRESO-AJ19 AJ30-(RETURN)
DIST/DDINGRESO-AL19 AL30-(RETURN)
DIST/DDINGRESO-AN19 AN30-(RETURN)
DIST/DDINGRESO-AP19 AP30-(RETURN)
```

1B MACRO PARA CREAR INTERVALOS EN LA ESCALA CUM

A1052

```
(GOTO)J19-(END)(DOWN)(LEFT)/RNCUM- (END)(UP)/RNCUM2-(END)(DOWN)-
(GOTO)L16-/DFL16 Z16-(2) - 5-11-
(RIGHT)Ncum 2-(RIGHT 2) Ncum 3-(RIGHT 2) Ncum 4-(RIGHT 2) Ncum 5-
(RIGHT 2) Ncum 6-(RIGHT 2) Ncum 7-(RIGHT 2) Ncum 8-(RIGHT 2) Ncum 9-
(GOTO)K19-@MAX(CUM)-(GOTO)C19-(END)(DOWN)(UP (@max(cum2)-K19)/RNCING- (END)(UP)-
(GOTO)L19-+SK$19/L$16-(DOWN)(UP)+L$19-/C- (DOWN (L$16-2))-(UP)/DDCUM-L19 .L29-
(GOTO)N19-+SK$19/N$16-(DOWN)(UP)+N$19-/C- (DOWN (N$16-2))-(UP)/DDCUM-N19 .N29-
(GOTO)P19-+SK$19/P$16-(DOWN)(UP)+P$19-/C- (DOWN (P$16-2))-(UP)/DDCUM-P19 .P29-
(GOTO)R19-+SK$19/R$16-(DOWN)(UP)+R$19-/C- (DOWN (R$16-2))-(UP)/DDCUM-R19 .R29-
(GOTO)T19-+SK$19/T$16-(DOWN)(UP)+T$19-/C- (DOWN (T$16-2))-(UP)/DDCUM-T19 .T29-
(GOTO)V19-+SK$19/V$16-(DOWN)(UP)+V$19-/C- (DOWN (V$16-2))-(UP)/DDCUM-V19 .v29-
(GOTO)X19-+SK$19/X$16-(DOWN)(UP)+X$19-/C- (DOWN (X$16-2))-(UP)/DDCUM-X19 .X29-
(GOTO)Z19-+SK$19/Z$16-(DOWN)(UP)+Z$19-/C- (DOWN (Z$16-2))-(UP)/DDCUM-Z19 .Z29-
(ENDD)(DOWN)(RIGHT)(DOWN)/RE (END)(DOWN)(LEFT 15)-
(GOTO)D5-@AVG(ING)-(DOWN)@VAR(ING)-(DOWN)@STD(ING)-
(DOWN)@SUM(ING).D5-(DOWN)@COUNT(ING)-
2/RNCDIST-B1041-
(IF M19=0)(GOTO)AB19-+(LEFT 22)(UP)-(BRANCH B1074)
(GOTO)AB19-+(LEFT 22)(DOWN (M19-1))-
(IF M20=0)(DIST)
(DOWN)@MAX(ING)-
(DIST)
3/RNCDIST-B1042-
(IF O19=0)(GOTO)AD19-+(LEFT 24)(UP)-(BRANCH B1079)
(GOTO)AD19-+(LEFT 24)(DOWN (O19-1))-
(DOWN)+(LEFT 24)(DOWN (O19-2+O20))-
(DOWN)@MAX(ING)-
(DIST)
4/RNCDIST-B1043-
(IF O19=0)(GOTO)AF19-+(LEFT 26)(UP)-(BRANCH B1085)
(GOTO)AF19-+(LEFT 26)(DOWN (O19-1))-
(DOWN)+(LEFT 26)(DOWN (O19-2+O20))-
(DOWN)+(LEFT 26)(DOWN (O19-3+O20+O21))-
(DOWN)@MAX(ING)-
(DIST)
5/RNCDIST-B1044-
(IF S19=0)(GOTO)AI19-+(LEFT 28)(UP)-(BRANCH B1092)
```

```

(GOTO)A119-+(LEFT 20)DOWN (S19-1)-
(DOWN)+(LEFT 28)DOWN (S19-2+S20)-
(DOWN)+(LEFT 28)DOWN (S19-3+S20+S21)-
(DOWN)+(LEFT 28)DOWN (S19-4+S20+S21+S22)-
(DOWN)@MAX(ING)-
(DIST)
6 /RNCDIST-B1045-
(IF U19=0)(GOTO)A119-+(LEFT 30)(UP)-(BRANCH B1100)
(GOTO)A119-+(LEFT 30)DOWN (U19-1)-
(DOWN)+(LEFT 30)DOWN (U19-2+U20)-
(DOWN)+(LEFT 30)DOWN (U19-3+U20+U21)-
(DOWN)+(LEFT 30)DOWN (U19-4+U20+U21+U22)-
(DOWN)+(LEFT 30)DOWN (U19-5+U20+U21+U22+U23)-
(DOWN)@MAX(ING)-
(DIST)
7 /RNCDIST-B1046-
(IF W19=0)(GOTO)A119-+(LEFT 32)(UP)-(BRANCH B1109)
(GOTO)A119-+(LEFT 32)DOWN (W19-1)-
(DOWN)+(LEFT 32)DOWN (W19-2+W20)-
(DOWN)+(LEFT 32)DOWN (W19-3+W20+W21)-
(DOWN)+(LEFT 32)DOWN (W19-4+W20+W21+W22)-
(DOWN)+(LEFT 32)DOWN (W19-5+W20+W21+W22+W23)-
(DOWN)+(LEFT 32)DOWN (W19-5+W20+W21+W22+W23+W24)-
(DOWN)@MAX(ING)-
(DIST)
8 /RNCDIST-B1047-
(IF Y19=0)(GOTO)A119-+(LEFT 34)(UP)-(BRANCH B1110)
(GOTO)A119-+(LEFT 34)DOWN (Y19-1)-
(DOWN)+(LEFT 34)DOWN (Y19-2+Y20)-
(DOWN)+(LEFT 34)DOWN (Y19-3+Y20+Y21)-
(DOWN)+(LEFT 34)DOWN (Y19-4+Y20+Y21+Y22)-
(DOWN)+(LEFT 34)DOWN (Y19-5+Y20+Y21+Y22+Y23)-
(DOWN)+(LEFT 34)DOWN (Y19-5+Y20+Y21+Y22+Y23+Y24)-
(DOWN)+(LEFT 34)DOWN (Y19-5+Y20+Y21+Y22+Y23+Y24+Y25)-
(DOWN)@MAX(ING)-
(DIST)
9 /RNCDIST-B1048-
(IF AA19=0)(GOTO)A119-+(LEFT 36)(UP)-(BRANCH B1130)
(GOTO)A119-+(LEFT 36)DOWN (AA19-1)-
(DOWN)+(LEFT 36)DOWN (AA19-2+AA20)-
(DOWN)+(LEFT 36)DOWN (AA19-3+AA20+AA21)-
(DOWN)+(LEFT 36)DOWN (AA19-4+AA20+AA21+AA22)-
(DOWN)+(LEFT 36)DOWN (AA19-5+AA20+AA21+AA22+AA23)-
(DOWN)+(LEFT 36)DOWN (AA19-5+AA20+AA21+AA22+AA23+AA24)-
(DOWN)+(LEFT 36)DOWN (AA19-5+AA20+AA21+AA22+AA23+AA24+AA25)-
(DOWN)+(LEFT 36)DOWN (AA19-5+AA20+AA21+AA22+AA23+AA24+AA25+AA26)-
(DOWN)@MAX(ING)-
(DIST)

```

A1220

```

S1 (GOTO)C19-/RNCST1-(RETURN)
A1 @AVG(S1)-(RETURN)
V1 (RIGHT)@VAR(S1)-(LEFT)(RETURN)
RV /RV (RIGHT)-(RETURN)
S2 (GOTO)C19-/RNCST2-(RETURN)
A2 @AVG(S2)-(RETURN)
V2 (RIGHT)@VAR(S2)-(LEFT)(RETURN)
S3 (GOTO)C19-/RNCST3-(RETURN)
A3 @AVG(S3)-(RETURN)
V3 (RIGHT)@VAR(S3)-(LEFT)(RETURN)
S4 (GOTO)C19-/RNCST4-(RETURN)
A4 @AVG(S4)-(RETURN)
V4 (RIGHT)@VAR(S4)-(LEFT)(RETURN)
S5 (GOTO)C19-/RNCST5-(RETURN)
A5 @AVG(S5)-(RETURN)
V5 (RIGHT)@VAR(S5)-(LEFT)(RETURN)
S6 (GOTO)C19-/RNCST6-(RETURN)
A6 @AVG(S6)-(RETURN)
V6 (RIGHT)@VAR(S6)-(LEFT)(RETURN)
S7 (GOTO)C19-/RNCST7-(RETURN)
A7 @AVG(S7)-(RETURN)
V7 (RIGHT)@VAR(S7)-(LEFT)(RETURN)
S8 (GOTO)C19-/RNCST8-(RETURN)

```



A8 @AVG(ST8)-(RETURN)  
V8 (RIGHT)@VAR(ST8)-(LEFT)(RETURN)

IC MACRO PARA CREAR RANGOS DE LOS ESTRATOS, MEDIAS Y VARIANZAS

A124B

```
(GOTO)J35~"Ymh 2~WC9~(RIGHT)"Sh 2~WC9~(RIGHT)"Ymh 3~WC9~(RIGHT)"Sh 3~WC9~(RIGHT)"Ymh 4~WC9~(RIGHT)"Sh 4~WC9~(RIGHT)"Ymh 5~WC9~(RIGHT)"Sh 5~WC9~(RIGHT)"Ymh 6~WC9~(RIGHT)"Sh 6~WC9~(RIGHT)"Ymh 7~WC9~(RIGHT)"Sh 7~WC9~(RIGHT)"Ymh 8~WC9~(RIGHT)"Sh 8~WC9~(RIGHT)"Ymh 9~WC9~(RIGHT)"Sh 9~WC9~(RIGHT)"
{GOTO}N33~MEDIA Y VARIANZA POR ESTRATO~/RNC51~B1220~/RNC61~B1221~/RNC71~B1222~/RNC81~B1223~/RNC91~B1224~/RNC101~B1225~/RNC111~B1226~/RNC121~B1227~/RNC131~B1228~/RNC141~B1229~/RNC151~B1230~/RNC161~B1231~/RNC171~B1232~/RNC181~B1233~/RNC191~B1234~/RNC201~B1235~/RNC211~B1236~/RNC221~B1237~/RNC231~B1238~/RNC241~B1239~/RNC251~B1240~/RNC261~B1241~/RNC271~B1242~/RNC281~B1243~/RNC291~B1244~
ST1 (IF AC19=0){BRANCH B1262}
{S1}{DOWN (AC19-1)}~(GOTO)J37~@AVG(ST1)-(RIGHT)@VAR(ST1)-(LEFT)(RV)
{S1}{ESC}{DOWN (AC19)}{DOWN (AG20-1)}~(GOTO)J38~(A1)(V1)(RV)
ST2 (IF AE19=0){BRANCH B1265}
{S2}{DOWN (AE19-1)}~(GOTO)N37~@AVG(ST2)-(RIGHT)@VAR(ST2)-(LEFT)(RV)
{S2}{ESC}{DOWN (AE19)}{DOWN (AE20-1)}~(GOTO)N38~(A2)(V2)(RV)
{S2}{ESC}{DOWN (AE20)}{DOWN (AE21-1)}~(GOTO)N39~(A2)(V2)(RV)
ST3 (IF AG19=0){BRANCH B1269}
{S3}{DOWN (AG19-1)}~(GOTO)P37~@AVG(ST3)-(RIGHT)@VAR(ST3)-(LEFT)(RV)
{S3}{ESC}{DOWN (AG19)}{DOWN (AG20-1)}~(GOTO)P38~(A3)(V3)(RV)
{S3}{ESC}{DOWN (AG20)}{DOWN (AG21-1)}~(GOTO)P39~(A3)(V3)(RV)
{S3}{ESC}{DOWN (AG21)}{DOWN (AG22-1)}~(GOTO)P40~(A3)(V3)(RV)
ST4 (IF AI19=0){BRANCH B1274}
{S4}{DOWN (AI19-1)}~(GOTO)R37~@AVG(ST4)-(RIGHT)@VAR(ST4)-(LEFT)(RV){IF AI20=0}{BRANCH B1275}
{S4}{BS}{DOWN (AI19)}{DOWN (AI20-1)}~(GOTO)R38~(A4)(V4)(RV){IF AI21=0}{BRANCH B1276}
{S4}{BS}{DOWN (AI20+AI19)}{DOWN (AI21-1)}~(GOTO)R39~(A4)(V4)(RV){IF AI22=0}{BRANCH B1277}
{S4}{BS}{DOWN (AI19+AI20+AI21)}{DOWN (AI22-1)}~(GOTO)R40~(A4)(V4)(RV){IF AI23=0}{BRANCH B1278}
{S4}{BS}{DOWN (AI19+AI20+AI21+AI22)}{DOWN (AI23-1)}~(GOTO)R41~(A4)(V4)(RV)
ST5 (IF AK19=0){BRANCH B1280}
{S5}{DOWN (AK19-1)}~(GOTO)T37~@AVG(ST5)-(RIGHT)@VAR(ST5)-(LEFT)(RV){IF AK20=0}{BRANCH B1281}
{S5}{BACKSPACE}{DOWN (AK19)}{DOWN (AK20-1)}~(GOTO)T38~(A5)(V5)(RV){IF AK21=0}{BRANCH B1282}
{S5}{BACKSPACE}{DOWN (AK19+AK20)}{DOWN (AK21-1)}~(GOTO)T39~(A5)(V5)(RV){IF AK22=0}{BRANCH B1283}
{S5}{BACKSPACE}{DOWN (AK19+AK20+AK21)}{DOWN (AK22-1)}~(GOTO)T40~(A5)(V5)(RV){IF AK23=0}{BRANCH B1284}
{S5}{BACKSPACE}{DOWN (AK19+AK20+AK21+AK22)}{DOWN (AK23-1)}~(GOTO)T41~(A5)(V5)(RV){IF AK24=0}{BRANCH B1285}
{S5}{BACKSPACE}{DOWN (AK19+AK20+AK21+AK22+AK23)}{DOWN (AK24-1)}~(GOTO)T42~(A5)(V5)(RV)
{S5}{IF AM19=0}{BRANCH B1287}
{S6}{DOWN (AM19-1)}~(GOTO)V37~@AVG(ST6)-(RIGHT)@VAR(ST6)-(LEFT)(RV){IF AM20=0}{BRANCH B1288}
{S6}{BS}{DOWN (AM19)}{DOWN (AM20-1)}~(GOTO)V38~(A6)(V6)(RV){IF AM21=0}{BRANCH B1289}
{S6}{BS}{DOWN (AM19+AM20)}{DOWN (AM21-1)}~(GOTO)V39~(A6)(V6)(RV){IF AM22=0}{BRANCH B1290}
{S6}{BS}{DOWN (AM19+AM20+AM21)}{DOWN (AM22-1)}~(GOTO)V40~(A6)(V6)(RV){IF AM23=0}{BRANCH B1291}
{S6}{BS}{DOWN (AM19+AM20+AM21+AM22)}{DOWN (AM23-1)}~(GOTO)V41~(A6)(V6)(RV){IF AM24=0}{BRANCH B1292}
{S6}{BS}{DOWN (AM19+AM20+AM21+AM22+AM23)}{DOWN (AM24-1)}~(GOTO)V42~(A6)(V6)(RV){IF AM25=0}{BRANCH B1293}
{S6}{BS}{DOWN (AM19+AM20+AM21+AM22+AM23+AM24)}{DOWN (AM25-1)}~(GOTO)V43~(A6)(V6)(RV)
{S7}{IF AO19=0}{BRANCH B1295}
{S7}{DOWN (AO19-1)}~(GOTO)X37~@AVG(ST6)-(RIGHT)@VAR(ST7)-(LEFT)(RV){IF AO20=0}{BRANCH B1296}
{S7}{BS}{DOWN (AO19)}{DOWN (AO20-1)}~(GOTO)X38~(A7)(V7)(RV){IF AO21=0}{BRANCH B1297}
{S7}{BS}{DOWN (AO19+AO20)}{DOWN (AO21-1)}~(GOTO)X39~(A7)(V7)(RV){IF AO22=0}{BRANCH B1298}
{S7}{BS}{DOWN (AO19+AO20+AO21)}{DOWN (AO22-1)}~(GOTO)X40~(A7)(V7)(RV){IF AO23=0}{BRANCH B1299}
{S7}{BS}{DOWN (AO19+AO20+AO21+AO22)}{DOWN (AO23-1)}~(GOTO)X41~(A7)(V7)(RV){IF AO24=0}{BRANCH B1300}
{S7}{BS}{DOWN (AO19+AO20+AO21+AO22+AO23)}{DOWN (AO24-1)}~(GOTO)X42~(A7)(V7)(RV){IF AO25=0}{BRANCH B1301}
{S7}{BS}{DOWN (AO19+AO20+AO21+AO22+AO23+AO24)}{DOWN (AO25-1)}~(GOTO)X43~(A7)(V7)(RV){IF AO26=0}{BRANCH B1302}
{S7}{BS}{DOWN (AO19+AO20+AO21+AO22+AO23+AO24+AO25)}{DOWN (AO26-1)}~(GOTO)X44~(A7)(V7)(RV)
{S8 (IF AQ19=0){BRANCH B1304}
{S8}{DOWN (AQ19-1)}~(GOTO)Z37~@AVG(ST8)-(RIGHT)@VAR(ST8)-(LEFT)(RV){IF AQ20=0}{BRANCH B1305}
{S8}{BACKSPACE}{DOWN (AQ19)}{DOWN (AQ20-1)}~(GOTO)Z38~(A8)(V8)(RV){IF AQ21=0}{BRANCH B1306}
{S8}{BS}{DOWN (AQ19+AQ20)}{DOWN (AQ21-1)}~(GOTO)Z39~(A8)(V8)(RV){IF AQ22=0}{BRANCH B1307}
{S8}{BS}{DOWN (AQ19+AQ20+AQ21)}{DOWN (AQ22-1)}~(GOTO)Z40~(A8)(V8)(RV){IF AQ23=0}{BRANCH B1308}
{S8}{BS}{DOWN (AQ19+AQ20+AQ21+AQ22)}{DOWN (AQ23-1)}~(GOTO)Z41~(A8)(V8)(RV){IF AQ24=0}{BRANCH B1309}
{S8}{BS}{DOWN (AQ19+AQ20+AQ21+AQ22+AQ23)}{DOWN (AQ24-1)}~(GOTO)Z42~(A8)(V8)(RV){IF AQ25=0}{BRANCH B1310}
```

```
(S8){B5}{DOWN (AQ19+AQ20+AQ21+AQ22+AQ23+AQ24)}{DOWN (AQ25-1)}-{GOTO}Z43-{A8}{V8}{RV}{IF AQ26=0}
(BRANCH B1311)
(S8){B5}{DOWN (AQ19+AQ20+AQ21+AQ22+AQ23+AQ24+AQ25)}{DOWN (AQ26-1)}-{GOTO}Z44-{A8}{V8}{RV}
(IF AQ27=0){BRANCH B1312}
(S8){B5}{DOWN (AO19+AQ20+AQ21+AO22+AQ23+AQ24+AO25+AQ26)}{DOWN (AQ27-1)}-{GOTO}Z45-{A8}{V8}{RV}
```

ID MACRO PARA OBTENER PONDERADORES

A1336

```
(GOTO)P50-"PONDERADORES-
(GOTO)L52-"Wh-(RIGHT)"WhSh-(RIGHT)"NhSh(RIGHT)"NhSh^2-(RIGHT)"Wh-(RIGHT)"WhSh-(RIGHT)"NhSh-
(RIGHT)"NhSh^2-"Wh-(RIGHT)"WhSh-(RIGHT)"NhSh(RIGHT)"NhSh(RIGHT)"NhSh^2-(RIGHT)"Wh-(RIGHT)"WhSh-(RIGHT)"NhSh-
(RIGHT)"NhSh^2-CL52.W52-(RIGHT)-
(GOTO)L54-/RFF3-L54.AZ65-
*AC19/$D$9-/C-(DOWN) (DOWN (L16-2))-(RIGHT)
+(LEFT)@SQRT(M37)-/C-(DOWN) (DOWN (L16-2))-(RIGHT)
*AC19@SQRT(M37)-/C-(DOWN) (DOWN (L16-2))-(RIGHT)
*AC19^M37-/C-(DOWN) (DOWN (L16-2))-(RIGHT)
*AE19/$D$9-/C-(DOWN) (DOWN (N16-2))-(RIGHT)
+(LEFT)@SQRT(O37)-/C-(DOWN) (DOWN (N16-2))-(RIGHT)
*AE19@SQRT(O37)-/C-(DOWN) (DOWN (N16-2))-(RIGHT)
*AE19^O37-/C-(DOWN) (DOWN (N16-2))-(RIGHT)
*AG19/$D$9-/C-(DOWN) (DOWN (P16-2))-(RIGHT)
+(LEFT)@SQRT(Q37)-/C-(DOWN) (DOWN (P16-2))-(RIGHT)
*AG19@SQRT(Q37)-/C-(DOWN) (DOWN (P16-2))-(RIGHT)
*AG19^Q37-/C-(DOWN) (DOWN (P16-2))-(RIGHT)
*AH19/$D$9-/C-(DOWN) (DOWN (R16-2))-(RIGHT)
+(LEFT)@SQRT(S37)-/C-(DOWN) (DOWN (R16-2))-(RIGHT)
*AH19@SQRT(S37)-/C-(DOWN) (DOWN (R16-2))-(RIGHT)
*AH19^S37-/C-(DOWN) (DOWN (R16-2))-(RIGHT)
*AK19/$D$9-/C-(DOWN) (DOWN (T16-2))-(RIGHT)
+(LEFT)@SQRT(U37)-/C-(DOWN) (DOWN (T16-2))-(RIGHT)
*AK19@SQRT(U37)-/C-(DOWN) (DOWN (T16-2))-(RIGHT)
*AK19^U37-/C-(DOWN) (DOWN (T16-2))-(RIGHT)
*AM19/$D$9-/C-(DOWN) (DOWN (V16-2))-(RIGHT)
+(LEFT)@SQRT(W37)-/C-(DOWN) (DOWN (V16-2))-(RIGHT)
*AM19@SQRT(W37)-/C-(DOWN) (DOWN (V16-2))-(RIGHT)
*AM19^W37-/C-(DOWN) (DOWN (V16-2))-(RIGHT)
*AO19/$D$9-/C-(DOWN) (DOWN (X16-2))-(RIGHT)
+(LEFT)@SQRT(Y37)-/C-(DOWN) (DOWN (X16-2))-(RIGHT)
*AO19@SQRT(Y37)-/C-(DOWN) (DOWN (X16-2))-(RIGHT)
*AO19^Y37-/C-(DOWN) (DOWN (X16-2))-(RIGHT)
*AQ19/$D$9-/C-(DOWN) (DOWN (Z16-2))-(RIGHT)
+(LEFT)@SQRT(AA37)-/C-(DOWN) (DOWN (Z16-2))-(RIGHT)
*AO19@SQRT(AA37)-/C-(DOWN) (DOWN (Z16-2))-(RIGHT)
*AO19^AA37-/C-(DOWN) (DOWN (Z16-2))-(RIGHT)
```

IE MACRO PARA CALCULAR ASIGNACION DE NEYMAN PARA CADA ESTRATO

A1381

```
(GOTO)M70-"TAMANO DE MUESTRA POR ESTRATO-
(GOTO)K72-"# DE ST-(RIGHT)+L16-(RIGHT)+N16-(RIGHT)+P16-(RIGHT)+R16-
(RIGHT)+T16-(RIGHT)+V16-(RIGHT)+X16-(RIGHT)+Z16-
(GOTO)K73-L-/C-(RIGHT) (RIGHT 0)-
(GOTO)K74-"n-
(RIGHT)@SUM(N54..N64)^2/(D8^2/D4^2+@SUM(O54..O64))-
(RIGHT)@SUM(R54..R64)^2/(D8^2/D4^2+@SUM(S54..S64))-
(RIGHT)@SUM(V54..V64)^2/(D8^2/D4^2+@SUM(W54..W64))-
(RIGHT)@SUM(Z54..Z64)^2/(D8^2/D4^2+@SUM(AA54..AA64))-
(RIGHT)@SUM(AD54..AD64)^2/(D8^2/D4^2+@SUM(AE54..AE64))-
(RIGHT)@SUM(AH54..AH64)^2/(D8^2/D4^2+@SUM(AI54..AI64))-
(RIGHT)@SUM(AL54..AL64)^2/(D8^2/D4^2+@SUM(AM54..AM64))-
(RIGHT)@SUM(AP54..AP64)^2/(D8^2/D4^2+@SUM(AQ54..AQ64))-
(GOTO)K75-L-/C-(RIGHT) (RIGHT 7)-
(GOTO)L77-+ $L$74^M54/@SUM($M$54..$M$64)-/C-(DOWN) (DOWN (L16-2))-
(GOTO)M77-+ $M$74^Q54/@SUM($Q$54..$Q$64)-/C-(DOWN) (DOWN (N16-2))-
(GOTO)N77-+ $N$74^U54/@SUM($U$54..$U$64)-/C-(DOWN) (DOWN (P16-2))-
```

```
{GOTO}O77-+{S}S74*Y54/@SUM({S}Y54 {S}Y564)-/C-{DOWN} {DOWN (R16-2)}-
{GOTO}P77-+{P}P74*AC54/@SUM({S}AC54 {S}AC564)-/C-{DOWN} {DOWN (T16-2)}-
{GOTO}Q77-+{Q}Q74*AG54/@SUM({S}AG54 {S}AG564)-/C-{DOWN} {DOWN (V16-2)}-
{GOTO}R77-+{R}R74*AK54/@SUM({S}AK54 {S}AK564)-/C-{DOWN} {DOWN (X16-2)}-
{GOTO}S77-+{S}S74*AO54/@SUM({S}AO54 {S}AO564)-/C-{DOWN} {DOWN (Z16-2)}-
```

IF MACRO QUE GENERA TABLA DE INTRA-VARIANZA E INTER-VARIANZA

A1412

```
{GOTO}AB35-~{X}H-XY*2*NI~/C-{RIGHT} {RIGHT 6}-
{GOTO}AB37-~{L37-SD55}*2*AC19~/C-{DOWN} {DOWN (L16-2)}-
{RIGHT}{N37-SD55}*2*AE19~/C-{DOWN} {DOWN (N16-2)}-
{RIGHT}{P37-SD55}*2*AG19~/C-{DOWN} {DOWN (P16-2)}-
{RIGHT}{R37-SD55}*2*AI19~/C-{DOWN} {DOWN (R16-2)}-
{RIGHT}{T37-SD55}*2*AK19~/C-{DOWN} {DOWN (T16-2)}-
{RIGHT}{V37-SD55}*2*AM19~/C-{DOWN} {DOWN (V16-2)}-
{RIGHT}{X37-SD55}*2*AO19~/C-{DOWN} {DOWN (X16-2)}-
{RIGHT}{Z37-SD55}*2*AQ19~/C-{DOWN} {DOWN (Z16-2)}-
{GOTO}N90-~TABLA DE VARIANZAS-
{GOTO}L91-~No. DE ESTRATOS-(RIGHT 2)*V(1)-(RIGHT)*V(2)-
{GOTO}L93-~L16-{DOWN}*N16-{DOWN}*P16-{DOWN}*R16-{DOWN}*T16-{DOWN}*V16-{DOWN}*X16-{DOWN}*Z16-
{GOTO}N93-@SUM(O54..O64)/SD59-(RIGHT)@SUM(AB37..AB47)/SD59-
{DOWN}{LEFT}@SUM(I54..I64)/SD59-(RIGHT)@SUM(AC37..AC47)/SD59-
{DOWN}{LEFT}@SUM(W54..W64)/SD59-(RIGHT)@SUM(AD37..AD47)/SD59-
{DOWN}{LEFT}@SUM(AA54..AA64)/SD59-(RIGHT)@SUM(AE37..AE47)/SD59-
{DOWN}{LEFT}@SUM(AE54..AE64)/SD59-(RIGHT)@SUM(AF37..AF47)/SD59-
{DOWN}{LEFT}@SUM(AI54..AI64)/SD59-(RIGHT)@SUM(AG37..AG47)/SD59-
{DOWN}{LEFT}@SUM(AM54..AM64)/SD59-(RIGHT)@SUM(AH37..AH47)/SD59-
{DOWN}{LEFT}@SUM(AO54..AO64)/SD59-(RIGHT)@SUM(AI37..AI47)/SD59-
```