



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

"ALTERNATIVA DE ESTRATIFICACION USANDO EL INDICE DE GINI, EN APLICACIONES DE MERCADO."

T E S I S

QUE PARA OBTENER EL TITULO DE:

A C T U A R I A

P R E S E N T A :

GABRIELA BECERRA TREJO



FACULTAD DE CIENCIAS UNAM

DIRECTORA DE TESIS: M. en A.P. MARIA DEL PILAR ALONSO REYES

2005

0349673





Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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



UNIVERSIDAD NACIONAL
AVENIDA DE
MEXICO

Autorizo a la Dirección General de Bibliotecas de la UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo recepcional.
NOMBRE: Gabriela Becerra Trejo

FECHA: 09 de Noviembre de 2005

FIRMA: [Firma]

ACT. MAURICIO AGUILAR GONZÁLEZ
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo escrito:
"Alternativa de estratificación usando el índice de Gini,
en aplicaciones de mercado."

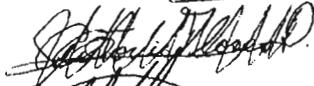
realizado por Becerra Trejo Gabriela

con número de cuenta 08711238-6 , quien cubrió los créditos de la carrera de: Actuaría

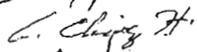
Dicho trabajo cuenta con nuestro voto aprobatorio.

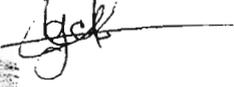
Atentamente

Director Propietario M. en A. P. María del Pilar Alonso Reyes 

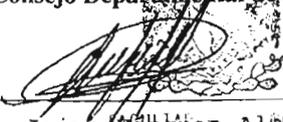
Propietario M. en C. José Antonio Flores Díaz 

Propietario Act. Jaime Vázquez Alamilla 

Suplente Act. Lucio Gerardo Chávez Heredia 

Suplente Dra. Guadalupe Carrasco Licea 

Consejo Departamental de Matemáticas


Act. Jaime Vázquez Alamilla

CONSEJO DEPARTAMENTAL

DE
MATEMÁTICAS

Índice

INTRODUCCIÓN	1
1. MÉTODOS DE ESTRATIFICACIÓN	5
1.1 Muestreo estratificado	7
1.1.1 Notación	8
1.1.2 Propiedades de las estimaciones	9
1.2 Condiciones que garantizan un diseño estratificado adecuado	14
1.3 Algunos métodos de estratificación	19
1.3.1 Estratificación óptima	21
1.3.2 Técnica de particiones sucesivas	27
1.3.3 Método de clases naturales	38
1.3.4 Técnicas de estratificación implícitas	45
1.3.5 Curvas de Lorenz e índice de Gini	50
2. ESTUDIOS DE MERCADO	65
2.1 Mezcla de marketing	66
2.2 Análisis de mercados	69
3. APLICACIONES	77
3.1 El México rural del fabricante X	77
3.1.1 Perfil geográfico de México	79
3.1.2 Metodología	87
3.2 Clasificación de tiendas de un fabricante licorero	95
3.2.1 Metodología	96
Anexo A	103
CONCLUSIONES	106
Bibliografía	108

Introducción

En la actualidad es necesario contar con información que contribuya a ayudar a entender mejor el mercado en el que los fabricantes están compitiendo o bien los segmentos del mercado en que desean ingresar.

Es frecuente el uso de herramientas que resultan de difícil interpretación o entendimiento para la toma de decisiones por parte del fabricante, o bien de largos y profundos análisis por parte del investigador que demandan gran cantidad de recursos y que por las características y dinámica del mercado, una vez que se tiene una conclusión, ésta tal vez ya no sea útil ni oportuna dado el entorno tan cambiante de los mercados.

Gran parte de los fabricantes que acuden a solicitar un estudio de mercado lo hacen para tener una visión adecuada para la toma de decisiones, aquellos que invierten en este tipo de estudios tienen la ventaja del conocimiento del segmento o segmentos en que están compitiendo. Como investigador, la mayor parte del trabajo corresponde a recolectar información para estratificar. En este trabajo de tesis se pretende presentar el índice de Gini como una alternativa de estratificación en aplicaciones de mercado y la curva de Lorenz como parte gráfica del proceso.

El índice de Gini (conocido también como coeficiente de concentración) fue creado por el estadístico italiano Corrado Gini y publicado por primera vez en 1912 en su

artículo "Variabilità e mutabilità", es frecuentemente usado para tratar de medir desigualdad de ingreso y niveles de pobreza en economía. La curva de Lorenz fue propuesta en 1905 con el propósito de ilustrar la desigualdad en la distribución de la salud y, desde su aparición, su uso se ha popularizado entre los estudiosos de la desigualdad económica. El índice de Gini se basa en la curva de Lorenz, que representa las frecuencias acumuladas y es posible comparar la distribución empírica de una variable con la igualdad o equidistribución del ingreso.

El índice de Gini es considerado el instrumento de uso más generalizado para medir y comparar la desigualdad distributiva de los ingresos. El Banco Mundial, por ejemplo, lo utiliza anualmente en sus Reportes de Desarrollo Mundial para medir la desigualdad distributiva del ingreso en los países del mundo y la Organización de las Naciones Unidas (ONU) lo utiliza para tratar de medir el Índice de Desarrollo Humano (IDH) de los países.

El primer capítulo de este trabajo da un acercamiento a los métodos de estratificación, exponiendo las razones del por qué estratificar, hasta los principales métodos que se emplean en la práctica, con ejemplos de cada caso. En este capítulo se establece la notación generalmente usada, propiedades de las estimaciones, condiciones que garantizan un diseño estratificado adecuado y algunos de los métodos de estratificación, concluyendo con la aplicación del índice de Gini y de la curva de Lorenz.

El segundo capítulo presenta algunos de los métodos para realizar estudios de mercado y las principales necesidades de los fabricantes en cuestión de información. Se presentan temas básicos de mercadotecnia como la mezcla de *marketing* o mercadeo y los puntos esenciales en el análisis de mercados. En este capítulo se presenta también un perfil del investigador o analista de mercado para satisfacer la función de facilitador en la comprensión del mercado por parte del fabricante o productor.

En el capítulo tres se presentan dos aplicaciones de mercado usando el índice de Gini y la curva de Lorenz como alternativa de estratificación; la primera responde a la necesidad por parte de un fabricante de productos de limpieza de establecer su *México rural*, es decir, delimitar el espacio accionable para sus esfuerzos de promoción, logística y ventas en el ámbito rural del país; la estratificación que ofrece el INEGI no le es útil dado que establece las localidades rurales como aquellas que tienen menos de 2500 habitantes y que no son cabeceras municipales y las localidades urbanas son, por complemento, aquellas con 2500 habitantes o más o que son cabeceras municipales, con base en el Censo de Población y Vivienda del año 2000 hay 199,125 localidades en la República Mexicana, en la parte urbana se tienen 2,896 localidades (1.5%) y en la parte rural 196,229 localidades (98.5%). Con la curva de Lorenz y el índice de Gini se establece una estratificación en cuatro cortes que resultó muy útil para el fabricante. La segunda aplicación ilustra la necesidad de un fabricante licorero de clasificar las tiendas a las que les surte productos para dirigir sus esfuerzos de promoción; a este respecto no hay una

estratificación previa de tiendas, por lo que se tuvo que recurrir a solicitar la colaboración de las tiendas para tratar de medir sus ventas de los productos del fabricante, con la curva de Lorenz e índice de Gini se determinan tres tipos de tienda: *grandes, medianas y pequeñas*.

Por último se presentan las conclusiones del presente trabajo, donde se expone la utilidad de la aplicación de estas dos técnicas.

A Víctor Manuel,

Mi tío, mi amigo, mi ejemplo
No habrá tiempo que borre
tus huellas en mi camino;
Ni libro que llene
de sabiduría mi tiempo.

Gracias por confiar en esa niña
por sostenerla en tus brazos
para que tocara las estrellas,
contigo jamás estuvieron lejos.

A María Luisa,

Mi madre, mi amiga, mi compañera
Gracias por tu espacio,
por tu espera,
por tu paciencia,
por ser esa maravillosa mujer
a la que espero parecerme un día.

A mis hermanas, Rebeca y Elizabeth,

Amigas constantes,
compañeras incansables,
gracias por estar a mi lado,
por entender mis oscuridades,
por ser parte de mi historia.

A Karla,

Gracias por llenar de luz
mi existencia, por colorear
mis días con tu sonrisa,
Eres un regalo de Dios.

A Pilar,
Por tu apoyo y confianza,
por ayudarme a convertir
este anhelo en realidad.
Muchas gracias.

A mis sinodales: Pita, Gerardo, Jaime y José Antonio,
Por su tiempo,
por sus consejos,
mi admiración y respeto.
Muchas gracias.

Capítulo 1

Métodos de Estratificación

El muestreo estratificado es un diseño en el cual se usa información acerca de la población para determinar grupos, denominados estratos, que son muestreados independientemente. Cada posible unidad de muestreo (son los elementos que son el objetivo del estudio y que se usan para seleccionarlos en la muestra) o miembro de la población debe pertenecer a sólo un estrato, no debe suceder que existan unidades de muestreo que no pertenezcan a ningún estrato ni unidades que pertenezcan a más de un estrato. Cuando se construyen apropiadamente los estratos con respecto a la variable a ser estimada, el diseño estratificado puede producir estimaciones de medias, proporciones, etc. de la población total, con una mayor precisión que un diseño de muestreo aleatorio simple.

En términos generales, el muestreo estratificado consiste en los siguientes pasos ¹

- La población completa de unidades de muestreo se divide en subpoblaciones distintas llamados estratos.
- Dentro de cada estrato se selecciona una muestra separada a partir de todas las unidades de muestreo que componen el estrato.

¹ Kish, Leslie, *Muestreo de encuestas*. Ed. Trillas, 1975

- De la muestra obtenida en cada estrato se calcula la media (se denota

$$\bar{y} = \frac{\sum y_i}{n},$$
 en donde y_i es el valor observado del elemento i -ésimo en la

muestra y n es el tamaño de muestra) o el error estándar (también conocido como desviación estándar, es la raíz cuadrada de la varianza de la

distribución de muestreo, se denota $S_y^2 = \frac{1}{N-1} \sum_1^N (Y_i - \bar{Y})^2$), u otra

estadística del estrato, estas estimaciones se ponderan adecuadamente para formar la estimación combinada para la población completa.

Un diseño muestral estratificado puede usarse para obtener estimaciones para subpoblaciones o para asegurar que estratos importantes tengan un número suficiente de unidades de muestreo en las muestras. Uno de los usos más comunes de la estratificación es para considerar variabilidad espacial por medio de la definición de estratos geográficos, resulta útil cuando los resultados del estudio necesitan ser reportados separadamente por áreas geográficas o regiones, como en los Censos de Población.

Existen tres razones principales por las que, comúnmente, se recurre a la estratificación:

1. La estratificación se utiliza para disminuir las varianzas de las estimaciones de la muestra.

2. Se pueden formar los estratos para utilizar diferentes diseños de muestreo y procedimientos de observación y de recolección de datos en diversas porciones de la población.
3. Las subpoblaciones dentro de los estratos pueden definirse como *dominios de estudio*, que se refieren a cualquier subdivisión acerca de la cual se planea un análisis o encuesta para proporcionar información numérica de precisión conocida. Por ejemplo, los resultados de encuestas nacionales se suelen publicar separadamente, según las regiones que componen al país; por tanto, es conveniente tratar a las regiones como estratos con selección separada en cada uno de ellos.

1.1 Muestreo estratificado

En el muestreo estratificado, la población de N unidades se divide en subpoblaciones de N_1, N_2, \dots, N_L unidades. Estas subpoblaciones no se traslapan y son exhaustivas, por lo tanto:

$$N_1 + N_2 + \dots + N_L = N$$

Estas subpoblaciones se denominan estratos y una vez determinados, se selecciona al menos una muestra de cada uno. Los valores de los N_h deben ser conocidos. Las extracciones deben realizarse de manera independiente en los diferentes estratos. Si la muestra en cada estrato (denotada como n_1, n_2, \dots, n_L) es seleccionada con muestreo aleatorio simple (MAS), el procedimiento total se describe como un *muestreo aleatorio estratificado* (MAE).

1.1.1 Notación

El subíndice h denota el estrato e i la i -ésima unidad dentro del estrato. Los símbolos siguientes se refieren, sin pérdida de generalidad, al estrato h .

N_h	Número total de unidades en el estrato h
n_h	Número de unidades en la muestra para el estrato h
Y_{hi}	Valor de la característica en estudio en la i -ésima unidad del estrato h
y_{hi}	Valor obtenido para la característica en estudio en la i -ésima unidad del estrato h
$W_h = \frac{N_h}{N}$	Ponderación o peso del estrato h $\left(\sum_{h=1}^L W_h = 1 \right)$
$f_h = \frac{n_h}{N_h}$	Fracción de muestreo en el estrato h
$Y_h = \frac{\sum_{i=1}^{N_h} Y_{hi}}{N_h}$	Media poblacional para la característica en estudio en el estrato h
$\bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$	Media muestral para la característica en estudio en el estrato h

$$S_h^2 = \frac{\sum (Y_{hi} - Y_h)^2}{N_h - 1}$$

Varianza poblacional para la característica en estudio en el estrato h

$$s_h^2 = \frac{\sum (y_{hi} - \bar{y}_h)^2}{n_h - 1}$$

Varianza muestral para la característica en estudio en el estrato h

1.1.2 Propiedades de las estimaciones

Con base en lo definido anteriormente sabemos que $Y = \sum_{h=1}^L \sum_{i=1}^{N_h} Y_{hi}$ de donde:

$$\frac{Y}{N} = \frac{\sum_{h=1}^L \sum_{i=1}^{N_h} Y_{hi}}{N} = \sum_{h=1}^L \frac{N_h}{N} \frac{\sum_{i=1}^{N_h} Y_{hi}}{N_h} = \sum_{h=1}^L W_h Y_h = Y_{st}$$

que es la expresión de la media

ponderada poblacional, se acostumbra denotar con el subíndice st (por stratified, en inglés). Para la parte muestral se calcula la media ponderada \bar{y}_{st} , como:

$$\bar{y}_{st} = \frac{\sum_{h=1}^L N_h \bar{y}_h}{N} = \sum_{h=1}^L W_h \bar{y}_h \text{ con } \sum_{h=1}^L N_h = N \text{ y } \sum_{h=1}^L W_h = 1$$

La varianza de esta media ponderada se obtiene como:

$$\text{var}(\bar{y}_{st}) = \sum W_h^2 \text{var}(\bar{y}_h)$$

Se tiene que tomar una muestra (que contenga al menos dos unidades de muestreo) de cada estrato para estimar su media correspondiente. El proceso de selección y el de estimación se llevan a cabo separada e independientemente.

La estimación \bar{y}_{st} en general no es la misma que la media muestral \bar{y} , que se escribe como:

$$\bar{y} = \frac{\sum_{h=1}^L n_h \bar{y}_h}{n}$$

La diferencia es que en \bar{y}_{st} las estimaciones a partir de estratos individuales reciben

sus ponderaciones correctas $\frac{N_h}{N}$. Cuando $\frac{n_h}{n} = \frac{N_h}{N}$ o $\frac{n_h}{N_h} = \frac{n}{N}$ entonces $\bar{y} = \bar{y}_{st}$,

(lo anterior significa que la fracción de muestreo es la misma en todos los estratos).

Esta estratificación se describe como estratificación con asignación proporcional de los números n_h y da lugar a una muestra autoponderada.

Si las ponderaciones W_h no representan una simple cuenta de los elementos, sino de otras unidades entonces se obtiene:

$$\bar{y}_{st} = \frac{1}{N} \sum N_h \bar{y}_h$$

$$\text{var}(y_{st}) = \frac{1}{N^2} \sum N_h^2 \text{var}(y_h)$$

Kish presenta varios ejemplos ilustrativos:

1. Se tomó una muestra de las madres en la ciudad A, y se registró el número de niños nacidos y se estimó separadamente en varias clases de edades de la madre; a estos números se les conoce como tasas de natalidad a una edad específica y para compararlas con otras ciudades, se puede asignar una distribución estándar de las edades de las madres como las ponderaciones (o pesos) para los estratos de edad, en lugar de la distribución de edades de la ciudad en la muestra.
2. El equipo completo de una compañía grande se distribuyó en estratos de diferentes tipos de equipo, de cada uno se obtuvo una muestra separada de elementos, mismos que se observaron y se les asignó a cada uno una medida de la razón de depreciación. La meta era producir una estimación única total de esta razón para el equipo completo de la compañía. Las ponderaciones de los estratos no se basaron en el número de máquinas que contenían (debido a la gran variación de valores monetarios de la maquinaria en los diferentes estratos), se basaron en los valores monetarios originales dentro de los estratos.
3. Se obtuvo una muestra de familias en una ciudad mediante un muestreo de manzanas (contenidas en un mapa, numeradas y divididas en estratos). Hay

que combinar las estimaciones de la muestra, hechas separadamente dentro de los estratos, para estimar las características de la ciudad completa. Las familias son los elementos de muestreo, pero su número en la población no se puede utilizar como ponderación porque no se conoce.

Si no se pueden encontrar buenas estimaciones para los tamaños de los estratos, se deberán utilizar las probabilidades conocidas de selección dentro de los estratos como ponderaciones: si f_h es la probabilidad de selección de cada elemento dentro del estrato h , entonces $\frac{1}{f_h}$ se puede asignar como ponderación en la muestra seleccionada en el estrato. Así, el tamaño (o peso) relativo del estrato h se estima como.

$$W_h = \frac{\frac{n_h}{\hat{f}_h}}{\sum \frac{n_h}{\hat{f}_h}} \quad y \quad \sum W_h = 1$$

en donde $\hat{f}_h = \frac{n_h}{N_h}$

Nótese que las ponderaciones no son aquí constantes conocidas, sino sólo estimaciones, sujetas a variación de muestreo. Estos problemas se presentan cuando no se conoce el número de elementos dentro de las unidades de muestreo.

En las *muestras estratificadas* en las que se hace una selección aleatoria individual y separadamente (mediante MAS) en cada estrato, la media simple de los elementos en el estrato h es:

$$\bar{y}_{h0} = \frac{1}{n_h} \sum_i^{n_h} y_{hi}$$

el subíndice 0 indica que la selección se realizó mediante un MAS. Para combinar los diferentes estratos se usará la definición de media ponderada y se obtiene como:

$$\bar{y}_{s10} = \sum_h^H W_h \bar{y}_{h0} = \sum_h^H W_h \frac{1}{n_h} \sum_i^{n_h} y_{hi}$$

La varianza de la muestra aleatoria simple de n_h elementos es:

$$\text{var}(\bar{y}_{h0}) = (1 - f_h) \frac{s_h^2}{n_h}, \quad \text{donde} \quad s_h^2 = \frac{1}{n_h - 1} \left(\sum_i^{n_h} y_{hi}^2 - \frac{y_h^2}{n_h} \right)$$

Se combinan las varianzas de las medias por estrato para obtener la varianza de la media de la muestra (\bar{y}_{s10}) como:

$$\text{var}(\bar{y}_{s10}) = \sum W_h^2 (1 - f_h) \frac{s_h^2}{n_h}$$

Si las ponderaciones se basan en las proporciones de los elementos de la población

en los estratos, entonces y puesto que $W_h = \frac{N_h}{N}$ y $n_h = f_h N_h$ se puede escribir

el estimador de la media de la muestra estratificada y la varianza como:

$$y_{r0} = \frac{1}{N} \sum_h^H N_h \frac{1}{n_h} \sum_i^{n_h} y_{hi} = \frac{1}{N} \sum_h^H \frac{1}{f_h} \sum_i^{n_h} y_{hi},$$

$$\text{var}(y_{r0}) = \frac{1}{N^2} \sum_h^H (1 - f_h) \frac{N_h^2}{n_h} s_h^2 = \frac{1}{N^2} \sum_h^H \frac{(1 - f_h)}{f_h} N_h s_h^2$$

1.2 Condiciones que garantizan un diseño estratificado adecuado

El método para definir los estratos depende del propósito de la estratificación. Si se desea calcular estimaciones para grupos particulares o regiones, cada uno puede ser asignado como un estrato. Si se tiene un conocimiento previo de la distribución espacial del área de estudio, los estratos deben ser definidos de manera tal que al interior de cada uno de ellos se tenga tanta homogeneidad como sea posible. Adicionalmente, puede usarse información de alguna variable auxiliar (para definir los estratos) que esté altamente correlacionada con la variable a ser estimada para incrementar la precisión, por ejemplo, si en un estudio previo se ha determinado que la cantidad de barro en el suelo está correlacionada con la cantidad de un cierto compuesto químico que puede encontrarse en el suelo, se puede entonces usar un mapa del área de estudio en el que se muestre la cantidad de barro en el piso para determinar los estratos que serán necesarios para estimar la concentración del químico y sus características.

Las condiciones de campo pueden requerir de distintos procedimientos de muestreo a ser aplicados en los grupos o estratos de la población con el objetivo de obtener mayor eficiencia (puede reflejarse en menor costo, mayor calidad del levantamiento o en ambas) y en esta característica radica una de las principales ventajas del diseño estratificado.

Los estratos pueden definirse para minimizar el costo del muestreo a fin de mantener un cierto nivel de precisión o bien para maximizar este nivel teniendo un costo predeterminado.

En el siguiente ejemplo se mostrará cómo el uso apropiado de la estratificación en un diseño muestral puede producir estimaciones con mayor precisión y, comparando contra un diseño aleatorio simple, como se requiere un tamaño de muestra menor (con el MAE) para obtener el mismo nivel de precisión.

Concentración de arsénico en la superficie del suelo². Se desea estimar la concentración promedio de arsénico en la superficie del suelo alrededor de una chimenea de un incinerador de basura para determinar si el suelo ha sido contaminado en un nivel por encima de las concentraciones normales de que se tiene registro en la región. Se tomaran muestras dentro de los 500 metros alrededor de la chimenea. La información recabada en estudios anteriores indica que la concentración de arsénico será mayor sobre el área en que prevalece la dirección

² Tomado de *Guidance for Choosing a Sampling Design for Environmental Data Collection*. United States Environmental Protection Agency. (<http://www.epa.gov/quality/qs-docs/g5s-final.pdf>)

del viento y que la variabilidad de la concentración del químico en el suelo será mayor para las áreas lodosas que para las arenosas. Con esta información sabemos que la precisión para la estimación puede incrementarse al dividir el área de estudio en estratos de acuerdo con la dirección del viento y el tipo de suelo.



Figura 1 Áreas de acuerdo con la dirección del viento y tipo de suelo

Por restricciones de presupuesto los investigadores tomaron 60 muestras del área alrededor de la chimenea. El área de estudio fue estratificada de acuerdo con la figura 1 y el método de asignación de Tschuprow-Neyman (en el que la muestra es

distribuida mediante la expresión: $n_h = n \frac{N_h S_h}{\sum_{h=1}^L N_h S_h}$) fue usado para determinar el

número de elementos a ser aleatoriamente seleccionados dentro de cada estrato.

El resumen de estadísticas para el muestreo estratificado para el ejemplo se muestran en la tabla 1.1:

	Estratos				MAE	MAS
	Viento bajo/Suelo lodoso	Viento bajo/Suelo Arenoso	Viento perpendicular/Suelo lodoso	Viento perpendicular/Suelo arenoso		
Tamaño de muestra (n)	43	5	10	2	60	60
Media	46.16	12.66	9.49	10.2	19.81	22.94
Error Estandar	9.99	4.63	2.28	3.12	0.48	4.35

Tabla 1.1 Resumen de estadísticas considerando un Muestreo Aleatorio Estratificado

Un MAS con 60 unidades de muestreo de suelo fueron tomadas del área de estudio, en la tabla 1.1 se observa que tomando el mismo tamaño de muestra para ambos diseños, se producen estimaciones para la concentración media de arsénico muy parecidas, pero el error estándar asociado con el MAE es mucho menor (es decir, se tiene una mejor precisión) que para el MAS.

En la tabla 1.2 se incluyen, para diferentes tamaños de muestra, los errores estándar calculados, debe notarse que usando el diseño estratificado es necesario un tamaño de muestra de 40 unidades para tener una precisión similar a la obtenida en el diseño aleatorio simple con un tamaño de muestra de 60.

	Muestreo Aleatorio Estratificado							Muestreo Aleatorio Simple
Tamaño de muestra	7	8	9	14	20	40	60	60
Error Estandar	10.59	9.73	9.06	7.57	6.41	4.51	3.68	4.35

Tabla 1.2 Tamaños de muestra y errores estándar

La estratificación puede ser útil cuando la implementación de diferentes diseños de muestreo reduce los costos asociados con la selección de la muestra. Los estratos pueden ser definidos para minimizar los costos asociados con el muestreo en varios lugares, que pueden definirse dentro de un mismo estrato cuando están cercanos o muy próximos para minimizar el tiempo de recorrido del equipo de campo que tomará las muestras en estos sitios. También puede suceder que sea más costoso levantar una muestra en una porción del lugar que para el resto, entonces esta parte puede asignarse como un sólo estrato para minimizar el costo de levantamiento de muestra.

El diseño estratificado también tiene algunas limitaciones, principalmente que se requiere de un conocimiento previo de la población para definir efectiva y apropiadamente los estratos y asignar los tamaños de muestra. Las ganancias en precisión o la reducción de costos dependen de la calidad de la información usada para la planeación y ejecución del diseño estratificado de muestreo. Cualquier

posible incremento en precisión depende particularmente de la fuerza de la correlación entre la(s) variable(s) de estratificación y aquella a ser estimada.

1.3 Algunos métodos de estratificación

Desde la óptica del muestreo es fundamental estratificar de manera tal que al interior de los estratos las características sean homogéneas, ya que esto reduce el número de selecciones para la muestra y contribuye a minimizar la varianza. Los métodos usados para definir los estratos dependen del objetivo de la estratificación.

Es muy importante tomar en cuenta que los criterios que se establezcan para determinar la estratificación sean variables permanentes en el tiempo, por ejemplo, cuando se necesita muestrear áreas geográficas y se determina que la variable tamaño de localidad (medida en población) es el criterio para estratificar, puede verse que esta variable varía muy poco de un censo de población a otro y puede ser un criterio válido por mucho tiempo, mismo que permitirá tener cierta *tendencia*. Otro aspecto importante a considerar es seleccionar límites conocidos, como por ejemplo los distritos electorales, dado que las publicaciones de interés basadas en encuestas son reportadas a esos niveles. No debe ignorarse que los recursos son el principal elemento para la planeación y ejecución de la estratificación.

A este respecto, se pueden agrupar los métodos de estratificación en 3 grandes grupos, según se basen en:

- **Criterios empíricos:** se utilizan cuando el investigador tiene suficiente conocimiento de las variables a estudiar y puede determinar aquellas que correlacionan con la variable a ser medida, por ejemplo, en el caso presentado de la chimenea, el investigador sabía que el tipo de suelo y la dirección del viento eran determinantes de la cantidad de arsénico en el piso. En general este conocimiento puede deberse a experiencias similares previas y es fundamental el sentido común y la experiencia del investigador y del muestrista.
- **Criterios administrativos:** son básicamente particiones geográficas, administrativas o por tamaño de población. La mayoría de las encuestas que se utilizan en análisis de mercados utilizan este tipo de criterios para determinar "zonas de influencia" para los establecimientos comerciales, productos de determinada marca o fabricante, así como también son los más utilizados en encuestas en donde se requiere obtener información acerca del consumo en hogares.
- **Métodos estadísticos:** se dirigen por características asociadas al fenómeno de estudio, como el nivel socioeconómico, de pobreza, etc. Estos métodos consisten en determinar los estratos mediante un criterio matemático, en donde la varianza al interior de los estratos sea minimizada y entre estratos sea la máxima.

En este trabajo se abordarán solamente algunos de los métodos del tercer grupo mencionado, es decir de los métodos estadísticos fundamentales o básicos, en el sentido de que la gran mayoría de los que no son básicos están conformados a partir de ellos.

Cabe mencionar que cuando los estratos se definen de acuerdo con una segunda variable que está correlacionada con aquella a ser estimada, como en el ejemplo de la chimenea en que se usaron la dirección del viento y el tipo de suelo, la definición óptima de los estratos consiste en conformarlos de tal forma que la población incluida en cada estrato sea tan homogénea como sea posible con respecto a esa segunda variable y tan heterogénea como sea posible entre estratos.

Enseguida se presentan los métodos básicos mencionados.

1.3.1 Estratificación óptima.

Los métodos para seleccionar las mejores fronteras para los estratos fueron denominados de *estratificación óptima* por Tore Elon Dalenius (1957, 1959). Dalenius habla de la determinación de los mejores puntos de frontera y_h ($h=1,2,3,\dots,H-1$) para crear H estratos. La muestra se debe afijar³

³ El término *Afijación* se refiere al reparto del tamaño de la muestra en los diferentes estratos o subpoblaciones. Cuando todos los estratos tienen el mismo número de elementos se le conoce como *afijación igual o simple*; cuando, por el contrario, el número de elementos en la muestra depende del tamaño del estrato se le conoce como *afijación proporcional*. La *afijación óptima* o de *Neyman* no sólo se toma en cuenta el tamaño del estrato sino la dispersión de los datos dentro de ellos.

óptimamente con $n_h = kW_h S_h$ y los puntos de frontera deben proporcionar la varianza mínima para un número fijo de estratos (H). Dalenius supone que la variable Y_i es continua. Las distribuciones de Y_i tienen, a menudo, colas largas. Se presentan los siguientes dos casos:

- si el recorrido de la distribución se divide en intervalos iguales, de manera que los recorridos $(y_h - y_{h-1})$ de los estratos sean iguales para toda h, entonces los tamaños relativos W_h de los estratos se hacen demasiado desiguales; los del extremo de la cola se hacen muy pequeños.
- si el área de la distribución de la población se divide en tamaños relativos W_h iguales para los H estratos, entonces los recorridos de los extremos se vuelven demasiado grandes. Por ejemplo, el estrato de elementos grandes en la cola de la derecha de una distribución asimétrica tendrá un recorrido largo $(y_h - y_{h-1})$. En estos estratos extremos la desviación estándar S_h será excesivamente grande.

Por lo anterior, hay que tomar un término medio entre proporciones W_h y recorridos $(y_h - y_{h-1})$ iguales. Es muy frecuente que se transija de manera semejante para presentar clases de datos económicos, demográficos y censales. Por ejemplo, las distribuciones de ingresos, de tamaños de ciudades (en millares), aparecen a menudo en clases con fronteras tales como 1, 2.5, 5, 10, 25, 50, 100, 250, 500,

1000, etc. Estas transacciones intuitivas pueden ayudar a comprender las fronteras que proporciona la teoría, que son términos medios de tipo similar.

La solución que da directamente la teoría es que la mejor frontera y_h , entre los estratos h y $h+1$, debe satisfacer la relación:

$$\frac{(y_h - \bar{Y}_h)^2 + S_h^2}{S_h} = \frac{(y_h - \bar{Y}_{h+1})^2 + S_{h+1}^2}{S_{h+1}}$$

esta solución óptima no es práctica porque los cinco parámetros en la ecuación dependen de y_h . Sin embargo, se han propuesto varias reglas para establecer estratos que permiten aproximarse al óptimo.

Los valores iguales de $W_h S_h$ para los estratos conducirán a fronteras que están cerca de las óptimas. Esta regla ($W_h S_h = W_{h+1} S_{h+1}$) no constituye un procedimiento simple para establecer directamente un conjunto de fronteras óptimas, porque sus parámetros dependen de ellas. Pero, cuando los datos se dan en clases, la regla puede ayudar a combinar algunas de ellas y quizá a dividir otras para obtener buenas fronteras. Igualdad de valores para $W_h S_h$ también significa igualdad en tamaños de muestra n_h en afijación óptima. En forma semejante, se ha demostrado que si se crean estratos con valores iguales de $W_h (y_h - y_{h-1})$ se obtienen buenas aproximaciones a las fronteras óptimas.

Las ponderaciones W_h para los estratos, representan proporciones de la curva de frecuencia relativa para la variable Y_i . Su altura es f_h en el valor y_h . Si la curva se

aproxima mediante un histograma de rectángulos de altura f_h y de anchura d_h , entonces $W_h = \sum f_h d_h$, sumados sobre el recorrido $(y_h - y_{h-1})$ para el estrato.

La siguiente regla ha resultado práctica y eficiente: primero calculando los valores de $\sqrt{f_h}$, y entonces acumulando $\sqrt{f_h}$, y dividiendo la acumulación en partes aproximadamente iguales. Los valores correspondientes de y_h serán las mejores fronteras.

En la práctica, la variable de estratificación se presenta en intervalos de clase y debe hacerse una modificación a la regla anterior que supone que las anchuras d_h son iguales para los intervalos, de la siguiente manera: deben acumularse los valores $\sqrt{d_h f_h}$ y dividir esta acumulación en el número de estratos deseados.

Enseguida se describe el procedimiento y se da un ejemplo con anchuras variables:

1. Formar un conjunto inicial de K intervalos que cubran el rango completo de valores observados y sea H el número de estratos deseado ($H < K$). Sean $[y_{h-1}, y_h]$ los puntos inicial y final del h-ésimo intervalo ($h=1,2,3,\dots,K-1$). Y contar el número de observaciones, f_h en cada intervalo.
2. Calcular $d_h = y_h - y_{h-1}$ y $T = \sqrt{f_h d_h}$.

3. Para cada intervalo h , calcular $C_h = \sum_{j=1}^h T_j$. Esto es una suma acumulativa

desde el primer intervalo hasta el h .

4. Calcular $Q = \frac{\sum_{j=1}^H T_j}{H}$. Para cada intervalo calcular $\frac{C_h}{Q}$ y redondearlo al

siguiente entero mayor, esta cantidad indica el número del estrato en que las observaciones del intervalo h serán clasificadas.

El siguiente ejemplo ilustra este procedimiento, en la tabla 1.3a se tienen los 10 municipios más importantes del estado de Chiapas, ordenados por la población, según Censo de Población y Vivienda del año 2000:

Entidad Federativa	Mun	Municipio	Pob_00
Chiapas	101	Tuxtla Gutiérrez	434,143
Chiapas	89	Tapachula	271,674
Chiapas	59	Ocosingo	146,696
Chiapas	78	San Cristóbal de las Casas	132,421
Chiapas	19	Comitán de Domínguez	105,210
Chiapas	52	Margaritas, Las	87,034
Chiapas	108	Villaflores	85,957
Chiapas	65	Palenque	85,464
Chiapas	97	Toniná	78,438
Chiapas	31	Chilón	77,686

Tabla 1.3a Los principales municipios del estado de Chiapas

Se construyen 5 intervalos iniciales y se desea tener $H=3$, en la tabla 1.3b se presentan los resultados:

Intervalo	d_h	f_h	T_h	C_h	C_h/Q	Estrato
77,686-87,500	9814	5	221.5	221.5	0.83	1
105,210-140,000	34790	2	263.8	485.3	1.81	2
146,696-150,000	3304	1	57.5	542.8	2.02	3
271,674-290,000	18326	1	135.4	678.2	2.53	3
434,000-450,000	16000	1	126.5	804.6	3.00	3
Total		10	804.6			

Tabla 1.3b Ejemplo del procedimiento de Dalenius-Hodges

El estrato 1 contiene los municipios de Chilón, Tonalá, Palenque, Villaflores y Las Margaritas, el estrato 2 comprende Comitán de Domínguez y San Cristóbal y el estrato 3 los municipios de Ocosingo, Tapachula y Tuxtla Gutiérrez.

Se ha demostrado que la regla de acumulación de $\sqrt{f_h}$ da resultados excelentes en una gran variedad de distribuciones. Cochran [5.A.7] presenta motivaciones convincentes para las reglas de aproximación y señala su buen comportamiento con datos empíricos. Sethi [1963, 1964] al investigar algunas distribuciones teóricas en detalle, también encontró exitosas las reglas y obtuvo fronteras óptimas para muestreo con afijación proporcional. Ambos autores encontraron que era más pobre el comportamiento de la regla simple de tamaños de estratos agregados iguales; es decir, valores iguales de $W_h Y_h$ en los estratos. Esto funciona bien, siempre y cuando

los coeficientes de variación $\frac{S_h}{Y_h}$ fueran aproximadamente iguales en los estratos.

Entonces $W_h \bar{Y}_h$ iguales implicarían $W_h Y_h$ iguales, lo cual da resultados próximos a la mejor solución.

Cabe resaltar que si bien las reglas de estratificación óptima pueden servir como una guía práctica para la formación de estratos, muy raras veces serán procedimientos precisos para obtener soluciones óptimas. Se necesita más evidencia para la situación común en la que la correlación entre las variables de estratificación y de estudio no es muy alta.

1.3.2 Técnica de particiones sucesivas.

Fue desarrollada por Francisco Sánchez Villareal y se basa en la idea de que las ganancias debidas a la estratificación se obtienen a medida que se logra maximizar la varianza entre estratos y, simultáneamente, minimizar la varianza dentro de los estratos. La variación total se puede descomponer en dos fuentes: la suma de cuadrados dentro de los estratos y entre ellos:

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

en donde $\sum_{h=1}^L \sum_{i=1}^{N_h} (y_{hi} - \bar{Y})^2$ corresponde a la variación total, $\sum_{h=1}^L \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2$ es la

suma de cuadrados dentro de los estratos y $\sum_{h=1}^L \sum_{i=1}^{N_h} (\bar{Y}_h - \bar{Y})^2$ es la suma de

cuadrados entre estratos.

La ganancia aportada por la estratificación se puede apreciar en la siguiente fórmula, que relaciona la varianza del estimador de la media obtenida por muestreo aleatorio simple con la varianza del estimador de la media obtenida por muestreo estratificado con afijación proporcional de la muestra:

$$\text{var}(\bar{y}_{\text{mas}}) = \text{var}(\bar{y}_{\text{prop}}) + \frac{\sum_{h=1}^L \sum_{i=1}^{N_h} (\bar{y}_h - \bar{Y})^2}{nN}$$

A medida que la varianza entre estratos sea mayor y, como consecuencia, la varianza al interior de los estratos sea menor, se obtendrá mayor eficiencia al utilizar muestreo estratificado. Maximizar la varianza entre estratos se reduce a minimizar la varianza dentro de los estratos, o sea, minimizar la siguiente suma de cuadrados:

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

Cuando se estratifica en forma univariada, la variable de estratificación presenta una alta correlación con la de estudio, por lo que el supuesto básico es que al lograr buenos resultados a partir de la variable auxiliar, se logrará un efecto similar con la de estudio.

En forma empírica, la minimización se logra a través del procedimiento de particiones sucesivas, que consiste en:

- Ordenar los datos en forma ascendente o descendente de los valores de la variable de estratificación.

- Inicialmente se tienen dos grupos o estratos, uno que incluye a los N elementos de la población y el otro que incluye cero elementos.
- Se calcula la aportación de cada uno de los dos estratos a la suma de cuadrados, en el paso inicial o final, uno de los dos estratos incluye la suma de cuadrados total y el otro tiene una contribución nula.

$$\sum_{i=1}^{N_k} (y_{hi} - \bar{Y}_h)^2$$

- Se excluyen, uno a la vez, los elementos del primer estrato y se incluyen en el segundo grupo. En el k -ésimo paso el primer estrato tiene $N - k$ elementos y el segundo tiene k . El proceso continúa hasta que el primer grupo tenga 0 y el segundo N .
- En cada paso se suman las aportaciones de los dos estratos a la suma de cuadrados:

$$\sum_{i=1}^{N_1} (y_{1i} - \bar{Y}_1)^2 + \sum_{i=1}^{N_2} (y_{2i} - \bar{Y}_2)^2$$

Para la suma anterior, en el paso k , $N_1 = N - k$ y $N_2 = k$

- Se observará que la suma de cuadrados disminuye desde los extremos hasta un punto, generalmente alejado del centro del grupo de observaciones, en el cual la suma de cuadrados es mínima.
- Este punto es la frontera para definir los estratos.

El investigador puede preferir la suma de cuadrados entre los estratos y tomar la mayor aportación a la suma correspondiente.

Como ejemplo para ilustrar el procedimiento se toman los datos de población total por municipio del estado de Chiapas reportados por el INEGI en el Censo de Población y Vivienda del año 2000. Chiapas está constituido por 118 municipios, presenta una distribución muy asimétrica pues hay municipios con más de 200 mil habitantes como **Tuxtla Gutiérrez** y **Tapachula** y municipios con menos de 3 mil habitantes como son los casos de **Sunuapa** y **Santiago del Pinar**. (Tabla 1.4 municipios del estado de Chiapas y población según censo 2000)

Entidad Federativa	Mun	Municipio	Pob_00
Chiapas	101	Tuxtla Gutiérrez	434,143
Chiapas	89	Tapachula	271,674
Chiapas	59	Ocosingo	146,696
Chiapas	78	San Cristóbal de las Casas	132,421
Chiapas	19	Comitán de Domínguez	105,210
Chiapas	52	Margaritas, Las	87,034
Chiapas	108	Villaflores	85,957
Chiapas	65	Palenque	85,464
Chiapas	97	Tonalá	78,438
Chiapas	31	Chilón	77,686
Chiapas	107	Villa Corzo	68,685
Chiapas	61	Ocozacoautla de Espinosa	65,673
Chiapas	17	Cintalapa	64,013
Chiapas	27	Chiapa de Corzo	60,620
Chiapas	57	Motozintla	59,875
Chiapas	99	Trinitaria, La	59,686
Chiapas	23	Chamula	59,005

Entidad Federativa	Mun	Municipio	Pob_00
Chiapas	96	Tila	58,153
Chiapas	106	Venustiano Carranza	52,833
Chiapas	34	Frontera Comalapa	52,168
Chiapas	77	Salto de Agua	49,300
Chiapas	40	Huixtla	48,476
Chiapas	69	Pijijiapan	46,949
Chiapas	20	Concordia, La	39,770
Chiapas	51	Mapastepec	39,055
Chiapas	15	Cacahoatán	39,033
Chiapas	92	Tecpatán	38,383
Chiapas	9	Arriaga	37,989
Chiapas	64	Oxchuc	37,887
Chiapas	46	Iquipilas	34,937
Chiapas	74	Reforma	34,809
Chiapas	102	Tuxtla Chico	33,467
Chiapas	93	Tenejapa	33,161
Chiapas	80	Siltepec	32,457
Chiapas	41	Independencia, La	32,245
Chiapas	81	Simojovel	31,615
Chiapas	37	Huehuetán	31,464
Chiapas	87	Suchiate	30,251
Chiapas	111	Zinacantán	29,754
Chiapas	68	Pichucalco	29,357
Chiapas	12	Berriozábal	28,719
Chiapas	32	Escuintla	28,064
Chiapas	26	Chenalhó	27,331
Chiapas	94	Teopisca	26,996
Chiapas	100	Tumbalá	26,866
Chiapas	71	Villa Comaltitlán	26,706
Chiapas	79	San Fernando	26,436
Chiapas	6	Amatenango de la Frontera	26,094
Chiapas	109	Yajalón	26,044
Chiapas	3	Acapetahua	25,154
Chiapas	30	Chicomuselo	24,994
Chiapas	2	Acala	24,754
Chiapas	72	Pueblo Nuevo Solistahuacán	24,405
Chiapas	54	Mazatán	24,079
Chiapas	103	Tuzantán	23,180
Chiapas	13	Bochil	22,722
Chiapas	4	Altamirano	21,948
Chiapas	8	Ángel Albino Corzo	21,848
Chiapas	76	Sabanilla	21,156

Entidad Federativa	Mun	Municipio	Pob_00
Chiapas	75	Rosas, Las	21,100
Chiapas	112	San Juan Cancuc	20,688
Chiapas	39	Huitlapán	20,041
Chiapas	48	Juárez	19,956
Chiapas	21	Copainalá	19,298
Chiapas	5	Amatán	18,778
Chiapas	38	Huixtán	18,630
Chiapas	44	Ixtapa	18,533
Chiapas	11	Bella Vista	18,205
Chiapas	62	Ostuacán	17,026
Chiapas	49	Larráinzar	16,538
Chiapas	66	Pantelhó	16,262
Chiapas	86	Suchiapa	15,890
Chiapas	16	Catazajá	15,709
Chiapas	83	Socoltenango	15,171
Chiapas	14	Bosque, El	14,993
Chiapas	114	Benemérito de las Américas	14,436
Chiapas	1	Acacoyagua	14,189
Chiapas	105	Unión Juárez	13,934
Chiapas	47	Jitotol	13,076
Chiapas	22	Chalchihuitán	12,256
Chiapas	104	Tzimol	11,925
Chiapas	70	Porvenir, El	11,641
Chiapas	35	Frontera Hidalgo	10,917
Chiapas	115	Maravilla Tenejapa	10,526
Chiapas	91	Tapilula	10,349
Chiapas	60	Ocoatepec	9,271
Chiapas	43	Ixtacomitán	9,143
Chiapas	42	Ixhuitán	8,877
Chiapas	116	Marqués de Comillas	8,580
Chiapas	67	Pantepec	8,566
Chiapas	82	Sitalá	7,987
Chiapas	84	Solosuchiapa	7,784
Chiapas	85	Soyaló	7,767
Chiapas	56	Mitontic	7,602
Chiapas	24	Chanal	7,568
Chiapas	18	Coapilla	7,217
Chiapas	53	Mazapa de Madero	7,180
Chiapas	25	Chapultenango	6,965
Chiapas	73	Rayón	6,870
Chiapas	10	Bejucal de Ocampo	6,673
Chiapas	7	Amatenango del Valle	6,559

Entidad Federativa	Mun	Municipio	Pob_00
Chiapas	36	Grandeza, La	5,969
Chiapas	110	San Lucas	5,673
Chiapas	98	Totolapa	5,513
Chiapas	50	Libertad, La	5,288
Chiapas	28	Chiapilla	5,242
Chiapas	33	Francisco León	5,236
Chiapas	117	Montecristo de Guerrero	5,086
Chiapas	55	Metapa	4,794
Chiapas	45	Ixtapangajoyá	4,707
Chiapas	29	Chicoasén	4,345
Chiapas	90	Tapalapa	3,639
Chiapas	113	Aldama	3,635
Chiapas	118	San Andrés Duraznal	3,423
Chiapas	58	Nicolás Ruíz	3,135
Chiapas	63	Osumacinta	3,132
Chiapas	119	Santiago el Pinar	2,174
Chiapas	88	Sunuapa	1,936

Tabla 1.4: Municipios del estado de Chiapas y población según Censo 2000

En la fase 1 resultan dos estratos. Ya que se tiene el ordenamiento de los municipios por la población de mayor, se procede a calcular las aportaciones de cada estrato a la suma de cuadrados dentro de estratos. La suma de cuadrados total es de 300,422 millones de unidades cuadráticas. Este valor se observa en la tabla 1.5, que es una fragmento de la tabla completa correspondiente al ejercicio. La suma de cuadrados desciende y su valor mínimo (92,506 millones) lo alcanza al incluir en el primer grupo a Tuxtla Gutiérrez y Tapachula. Los 116 municipios restantes forman el segundo grupo o estrato.

Mun	Municipio	Pob_00	Numeración		Suma de Cuadrados		
			Asc	Desc	Grupo 1	Grupo 2	Dentro Grupos
				119	0	300422111581	300422111581
101	Tuxtla Gutiérrez	434143	1	118	0	138315407183	138315407183
89	Tapachula	271674	2	117	13198087981	79308660190	92506748170
59	Ocosingo	146696	3	116	41547151418	65029361362	106576512780
78	San Cristóbal de las Casas	132421	4	115	58818198293	53750454000	112568652293
19	Comitán de Domínguez	105210	5	114	74728300335	47381331355	122109631690
52	Margaritas, Las	87034	6	113	89027998357	43505119854	132533118211
108	Villaflores	85957	7	112	99444607455	39693813608	139138421063
65	Palenque	85464	8	111	107338798933	35874666549	143213465482

Tabla 1.5 Primera Fase: dos estratos

El primer estrato aporta 13,198 millones y el segundo 79,308 millones.

En la fase 2 se incluyen los 116 municipios restantes del segundo estrato de la fase 1 y la suma de cuadrados total corresponde inicialmente a los 79,308 millones de la fase 1.

Al repetir el procedimiento se observa que la suma de cuadrados alcanza su valor mínimo en 35,595 millones y que corresponde a la posición del municipio de **Oxchuc**. Si en este punto se diera por terminado el proceso se tendrían tres estratos. El primero con 2 elementos, el segundo con 27 y el tercero con 89 elementos, como se observa en la tabla 1.6.

Mun	Municipio	Pob_00	Numeración		Suma de Cuadrados		
			Asc	Desc	Grupo1	Grupo 2	Dentro Grupos
101	Tuxtla Gutiérrez	434143	1	118			0
89	Tapachula	271674	2	117		79308660190	79308660190
59	Ocosingo	146696	3	116	12953285156	65029361362	77982646519
78	San Cristóbal de las Casas	132421	4	115	20056777054	53750454000	73807231054
19	Comitán de Domínguez	105210	5	114	24262221014	47381331355	71643552369
52	Margaritas, Las	87034	6	113	25621984944	43505119854	69127104798
108	Villaflores	85957	7	112	25704255170	39693813608	65398068778
65	Palenque	85464	8	111	25678942983	35874666549	61553609532
97	Tonalá	78438	9	110	25274031987	32815970356	58090002343
31	Chilón	77686	10	109	25224775011	29784394735	55009169746
107	Villa Corzo	68685	11	108	25127170686	27619448968	52746619654
61	Ocozacoautla de Espinosa	65673	12	107	24952376833	25688751300	50641128134
17	Cintalapa	64013	13	106	24856692465	23866458783	48723151248
27	Chiapa de Corzo	60620	14	105	24694051044	22291356779	46985407823
57	Motozintla	59875	15	104	24466249805	20745104808	45211354614
99	Trinitaria, La	59686	16	103	24224677652	19183580382	43408258034
23	Chamula	59005	17	102	23990352470	17645285965	41635638436
96	Tila	58153	18	101	23996545985	16143326914	40139872900
106	Venustiano Carranza	52833	19	100	23988337257	15001034026	38989371283
34	Frontera Comalapa	52168	20	99	24089731971	13880620857	37970352829
77	Salto de Agua	49300	21	98	24188074252	12923761385	37111835636
40	Huixtla	48476	22	97	24325097294	11998054962	36323152256
69	Pijijiapan	46949	23	96	24833543191	11144888833	35978432024
20	Concordia, La	39770	24	95	25314400661	10647553289	35961953950
51	Mapastepec	39055	25	94	25732513908	10171410003	35903923911
15	Cacahoatán	39033	26	93	26133858266	9685891038	35819749305
92	Tecpatán	38383	27	92	26508157916	9218390770	35726548686
9	Arriaga	37989	28	91	26844185785	8757666296	35601852080
64	Oxchuc	37887	29	90	27276736794	7934608598	35595316623
46	Jiquipilas	34937	30	89	27726736794	7934608598	35661345392
74	Reforma	34809	31	88	28183052202	7574892898	35757945100

Tabla 1.6 Segunda Fase: tres estratos

Los dos grupos resultantes de esta fase se subdividen a su vez. El grupo que originalmente incluye desde Ocosingo hasta Oxchuc minimiza su suma de cuadrados en el municipio de Chilón con 6,796 millones de unidades cuadráticas. Esto es, ahora este estrato se ha subdividido en un estrato con 8 elementos y otro con 19 elementos. Por otra parte, el estrato con 89 elementos que comprende desde Jiquipilas hasta Sunuapa minimiza su suma de cuadrados en el municipio de Bella Vista con 1,799 millones de unidades cuadráticas. En este punto se decide

terminar el proceso, aunque es obvio que puede continuarse a discreción, ver tablas

1.7 y 1.8

Mun	Municipio	Pob_00	Numeración		Suma de Cuadrados		
			Asc	Desc	Grupo 1	Grupo 2	Dentro Grupos
				28	0	19991337623	19991337623
59	Ocosingo	146696	1	27	0	13194598798	13194598798
78	San Cristóbal de las Casas	132421	2	26	101887813	8136702503	8238590316
19	Comitán de Domínguez	105210	3	25	888434114	5997568276	6886002390
52	Margaritas, Las	87034	4	24	2153800833	5118277341	7272078174
108	Villaflora	85957	5	23	2967034137	4226049912	7193084049
65	Palenque	85464	6	22	3530350137	3280141733	6810491870
97	Tonalá	78438	7	21	4235992988	2652275501	6888268489
31	Chilón	77686	8	20	4798084894	1998372061	6796456954
107	Villa Corzo	68685	9	19	5662158914	1684830068	7346988982
61	Ocozacoautla de Espinosa	65673	10	18	6511837282	1440832106	7952669388
17	Cintalapa	64013	11	17	7292996755	1220105118	8513101872
27	Chiapa de Corzo	60620	12	16	8120283337	1068517682	9188801019
57	Motozintla	59875	13	15	8858682847	914980201	9773663048
99	Trinitaria, La	59686	14	14	9500793602	743004240	10243797842
23	Chamula	59005	15	13	10088762776	561943291	10650706066
96	Tita	58153	16	12	10641339160	373204915	11014544075
106	Venustiano Carranza	52833	17	11	11383470436	283527984	11666998420
34	Frontera Comalapa	52168	18	10	12076757729	188108774	12264866503
77	Salto de Agua	49300	19	9	12843910521	126025550	12969936071
40	Huixtla	48476	20	8	13577199743	61168572	13638368314
69	Pijijapan	46949	21	7	14319640406	2647645	14322288051
20	Concordia, La	39770	22	6	15408222496	1238011	15409460507
51	Mapastepec	39055	23	5	16446726270	809352	16447535622
15	Cacahoatán	39033	24	4	17400017513	137219	17400154732
92	Tecpatán	38383	25	3	18315172273	5202	18315177475
9	Arriaga	37989	26	2	19182538021	0	19182538021
64	Oxchuc	37887	27	1	19991337623	0	19991337623

Tabla 1.7 Fase 3-1 : cuatro estratos

Mun	Municipio	Pob_00	Numeración		Suma de Cuadrados		
			Asc	Desc	Grupo 1	Grupo 2	Dentro Grupos
				90	0	8088588051	8088588051
46	Jiquipilas	34937	1	89	0	7738191240	7738191240
74	Reforma	34809	2	88	8192	7384504288	7384512480
102	Tuxtla Chico	33467	3	87	1326083	7072041171	7073367254
93	Tenejapa	33161	4	86	2485491	6763054694	6765540185
80	Siltepec	32457	5	85	4627997	6471318327	6475946323
41	Independencia, La	32245	6	84	6556371	6179795101	6186351472

Mun	Municipio	Pob_00	Numeración		Suma de Cuadrados		
			Asc	Desc	Grupo 1	Grupo 2	Dentro Grupos
81	Simojovel	31615	7	83	9643062	5902583639	5912226701
37	Huehuetán	31464	8	82	12407852	5623540867	5635948719
87	Suchiate	30251	9	81	19220208	5377238917	5396459124
111	Zinacantan	29754	10	80	27093812	5140271088	5167364900
68	Pichucalco	29357	11	79	35600613	4909449141	4945049754
12	Berriozábal	28719	12	78	46315474	4691909217	4738224691
32	Escuintla	28064	13	77	59567647	4487841463	4547409110
26	Chenalhó	27331	14	76	76186713	4299080613	4375267326
94	Teopisca	26996	15	75	93151180	4114428524	4207579703
100	Tumbalá	26866	16	74	108980845	3928236270	4037217116
71	Villa Comaltitán	26706	17	73	124132518	3741231491	3865364008
79	San Fernando	26436	18	72	139595431	3556352170	3695947601
	Amatenango de la						
6	Frontera	26094	19	71	156017808	3375497662	3531515470
109	Yajalón	26044	20	70	171175039	3190695307	3361870346
3	Acapetahua	25154	21	69	192075939	3024312183	3216388122
30	Chicomuseo	24994	22	68	212464002	2857082192	3069546193
2	Acala	24754	23	67	233159738	2690998290	2924158028
	Pueblo Nuevo						
72	Sojistahuacán	24405	24	66	255223741	2528806496	2784030237
54	Mazatán	24079	25	65	278502838	2369858720	2648361558
103	Tuzantán	23180	26	64	308941229	2228179192	2537120421
13	Bochil	22722	27	63	342098906	2092812259	2434911165
4	Altamirano	21948	28	62	381900634	1970785025	2352685659
8	Ángel Albino Corzo	21848	29	61	420163381	1846876069	2267039450
76	Sabanilla	21156	30	60	464469871	1734011705	2198481576
75	Rosas, Las	21100	31	59	506630246	1618404591	2125034836
112	San Juan Cancuc	20688	32	58	551418871	1507592698	2059011569
39	Huitiupán	20041	33	57	602164285	1406318182	2008482467
48	Juárez	19956	34	56	651089132	1303043168	1954132300
21	Copainalá	19298	35	55	706448322	1209182160	1915630482
5	Amatán	18778	36	54	766409685	1121801804	1888211489
38	Huixtán	18630	37	53	825350134	1033820257	1859170391
44	Ixtapa	18533	38	52	882628079	944183590	1826811669
62	Ostucán	17026	40	50	1016923820	787812831	1804736652
49	Larráinzar	16538	41	49	1096717488	723639689	1820357177
66	Pantelhó	16262	42	48	1177540222	661164404	1838704626

Tabla 1.8 Fase 3-2: cinco estratos

Al final del proceso han quedado cinco estratos con 2, 8, 19, 39 y 50 elementos respectivamente. Se tiene una suma de cuadrados dentro de estratos de 21,793

millones de unidades cuadráticas. Las aportaciones de cada estrato en unidades cuadráticas y porcentaje se pueden observar en la tabla 1.9.

Estrato	Límite de Estrato	Nh	Suma de Cuadrados dentro del Estrato	Porcentaje
1	Tapachula	2	13198087981	60.56%
2	Chilón	8	4798084894	22.02%
3	Oxchuc	19	1998372061	9.17%
4	Bella Vista	39	941846894.8	4.32%
5	Sunuapa	50	857179116.7	3.93%
Total		118	21793570946	100.00%

Tabla 1.9 Estratos Resultantes

1.3.3 Método de clases naturales.

Fue propuesto por Mineo entre los años 1978 y 1979, ha sido utilizado para construir las clases de una distribución continua cuando es usada para verificar la hipótesis de bondad de ajuste con la prueba χ^2 . Se ha demostrado que las clases naturales, teniendo mínima varianza permiten que la distribución multinomial de la prueba χ^2 sea un mejor ajuste para la distribución χ^{24} . Dado que son de mínima varianza, Giovanna Nicolini propuso en marzo del 2001 un método para determinar fronteras entre estratos. Este método se describe a continuación.

⁴ Según Giovanna Nicolini, 1987

Dada una población subdividida en H estratos con N_h elementos y $\sum_{h=1}^H N_h = N$ se

satisface la siguiente relación:

$$(N-1)S^2 = \sum_{h=1}^H (N_h - 1)S_h^2 + \sum_{h=1}^H N_h (\bar{Y}_h - \bar{Y})^2$$

Donde S^2 y \bar{Y} son la varianza y la media de la población, respectivamente. Si se

divide la ecuación anterior por $N-1$ y considerando que $\frac{(N_h - 1)}{(N-1)} \cong \frac{N_h}{(N-1)} \cong W_h$,

donde $W_h = \frac{N_h}{N}$, se tiene:

$$S^2 = \sum_{h=1}^H W_h S_h^2 + \sum_{h=1}^H W_h (\bar{Y}_h - \bar{Y})^2$$

Al multiplicar ambos lados de la ecuación anterior por el factor $\frac{(1-f)}{n}$ donde f es

la fracción de muestreo y n el tamaño de muestra, entonces $\frac{(1-f)}{n} S^2$ es el

estimador para la varianza en el MAS, $\frac{(1-f)}{n} \sum_{h=1}^H W_h S_h^2$ es el estimador de la

varianza en el muestreo estratificado con asignación proporcional, mientras que

$\frac{(1-f)}{n} \sum_{h=1}^H W_h (\bar{Y}_h - \bar{Y})^2$ es el incremento de la varianza en el MAS con respecto al estratificado.

Dado que la Estratificación es una alternativa de ganancia ante el MAS, es necesario formar estratos en los que el incremento de la varianza MAS vs. MAE sea *máximo* y el estimador de la varianza en el MAE con asignación proporcional sea *mínimo*. Este método se basa en medir las distancias entre los valores de la variable de estratificación con un *índice de expansión* particular. De entre las funciones de distancia posibles se elige la siguiente:

$$\frac{1}{N^2} \sum_{i=1}^K \sum_{j=1}^K (Y_i - Y_j)^2 N_i N_j$$

en donde $N_i \geq 1$ y $N_j \geq 1$ son las frecuencias asociadas con Y_i y Y_j , respectivamente.

En este método se considera, con base en la continuidad⁵ de la distribución, que la suma de las desviaciones entre los valores de la variable es la misma que de las observaciones con respecto del centro, representado por la media aritmética de las observaciones. Dado que el punto medio de la clase coincide con la media aritmética de los valores contenidos en ella se tiene que la variabilidad dentro de la clase es mínima.

⁵ Es común que la variable continua tome valores discontinuos pero, para respetar su naturaleza, esos valores pueden ser agrupados en intervalos pequeños, con la restricción de que la frecuencia en los intervalos sea mayor o igual a una unidad.

Por ejemplo, si la estatura de dos individuos es de 1675 mm. y 1683 mm., estos individuos pueden ser incluidos en la misma clase: 1675-1685, o también por aproximación, a ambos se les puede asignar una estatura de 1680 mm., que es el valor del punto medio de la clase. La distribución de una variable cuantitativa continua se representa con K intervalos de la forma: $[(Y_{il}, Y_{iL}), N_i]$, donde (Y_{il}, Y_{iL}) son los extremos del intervalo o clase i e N_i es la frecuencia ($i = 1, 2, \dots, K$). Sea Y_i la media de los extremos del intervalo, representada por: $\frac{(Y_{il} + Y_{iL})}{2}$, entonces la distribución está formada por los K pares ordenados (Y_i, N_i) ; se determinan H , $H < K$ clases naturales de la siguiente manera:

1. Se define la cantidad: $V_j = \frac{N_i N_{i+1}}{(N_i + N_{i+1})^2} (Y_{i+1} - Y_i)^2 \quad j = 1, 2, \dots, [K/2]$ (donde

$[K/2]$ es la parte entera de $K/2$) para valores contiguos X_i, X_{i+1}

2. Agrupar los valores Y_i, Y_{i+1} para los cuales se tenga el valor de V_j mínimo y

se forma una nueva clase con $Y_j = \frac{(Y_i N_i + Y_{i+1} N_{i+1})}{(N_i + N_{i+1})}$ con frecuencia

$$N_j = N_i + N_{i+1}$$

3. Repetir 1 y 2 hasta que se han obtenido H distintos estratos, en donde cada valor obtenido en el paso 2 representa la media de la clase natural y por la

forma de elegir los intervalos para formar cada clase natural, tiene varianza mínima al interior.

Este método determina una división en clases o estratos muy parecida a la distribución real de la variable.

Para ejemplificar el desempeño del método, se tomará la información del estado de Chiapas, del municipio de Tuxtla Gutiérrez a Chilón (10 municipios), determinando 5 clases naturales iniciales y estableciendo $H=3$, que se muestra en la tabla 1.10.

i	Clase	Pares
1	$[(77686, 87500), 5]$	$(82593, 5)$
2	$[(105210, 140000), 2]$	$(122605, 2)$
3	$[(146696, 150000), 1]$	$(148348, 1)$
4	$[(271674, 290000), 1]$	$(280837, 1)$
5	$[(434000, 450000), 1]$	$(442000, 1)$

Tabla 1.10 Cinco clases naturales iniciales

$$K = 5 \text{ y } \left\lceil \frac{K}{2} \right\rceil = 2 \text{ y se calculan los valores para el paso 1:}$$

$$V_1 = \frac{N_1 N_2}{(N_1 + N_2)^2} (Y_2 - Y_1)^2 = \frac{5(2)}{(5+2)^2} (122605 - 82593)^2 = 326726560$$

$$V_2 = \frac{N_3 N_4}{(N_3 + N_4)^2} (Y_4 - Y_3)^2 = \frac{1(1)}{(1+1)^2} (280837 - 148348)^2 = 4388333780$$

El valor mínimo corresponde a las clases 1 y 2, entonces se aplica el paso 2 y se forma la nueva clase con valor:

$$Y_1 = \frac{(Y_1 N_1 + Y_2 N_2)}{(N_1 + N_2)} = \frac{(82593 * 5 + 122605 * 2)}{(5 + 2)} = 94025 \quad \text{con} \quad \text{frecuencia}$$

$N_1 = N_1 + N_2 = 5 + 2 = 7$, quedando cuatro clases naturales, que se muestran en la tabla 1.11:

i	Clase	Pares
1	[(77686,140000),7]	(94025,7)
2	[(146696,150000),1]	(148348,1)
3	[(271674,290000),1]	(280837,1)
4	[(434000,450000),1]	(442000,1)

Tabla 1.11 Clases naturales resultantes

Como aún no tenemos en número de estratos deseado ($H=3$), se repiten los pasos 1 y 2:

$$V_1 = \frac{N_1 N_2}{(N_1 + N_2)^2} (Y_2 - Y_1)^2 = \frac{7(1)}{(7+1)^2} (148348 - 94025)^2 = 322764348$$

$$V_2 = \frac{N_3 N_4}{(N_3 + N_4)^2} (Y_4 - Y_3)^2 = \frac{1(1)}{(1+1)^2} (442000 - 280837)^2 = 6493378142$$

El valor mínimo corresponde a las clases 1 y 2, entonces se aplica el paso 2 y se forma la nueva clase con valor:

$$Y_1 = \frac{(Y_1 N_1 + Y_2 N_2)}{(N_1 + N_2)} = \frac{(94025 * 7 + 148348 * 1)}{(7 + 1)} = 100815 \quad \text{con frecuencia } N_1 = 8,$$

quedando tres clases, que se muestran en la tabla 1.12:

i	Clase	Pares
1	[(77686,150000),8]	(100815,8)
2	[(271674,290000),1]	(280837,1)
3	[(434000,450000),1]	(442000,1)

Tabla 1.12 Clases naturales finales

Y ya tenemos los 3 estratos, el primero considerando los municipios de Ocosingo, San Cristóbal de las Casas, Comitán de Domínguez, Las Margaritas, Villaflores, Palenque, Tonalá y Chilón. El segundo constituido por el municipio de Tapachula y el tercero con Tuxtla Gutiérrez.

1.3.4 Técnicas de estratificación implícita

El Muestreo Sistemático es uno de los procedimientos de selección de muestra que se conoce más ampliamente porque es muy sencillo de aplicar, consiste en tomar cada unidad k -ésima de muestreo después de un arranque aleatorio, a menudo se usa conjuntamente con la estratificación. A continuación se describe el procedimiento.

Supóngase que el tamaño de la población N es múltiplo entero del tamaño de muestra n , entonces el intervalo de muestreo $k = N/n$ es también un número entero. Si se selecciona al azar un número de 1 a k , este arranque aleatorio r para el intervalo determinará la unidad que se va a seleccionar en cada uno de los n estratos implícitos o zonas de k unidades cada una. Dentro de cada zona se selecciona una unidad, que ocupa la misma localización en todas las zonas. Puesto que el arranque se selecciona al azar de 1 a k , cada unidad recibe la misma probabilidad $1/k$ de selección.

El orden de la selección $1, 2, \dots, (n-1), n$ refleja el orden de numeración de los elementos de la población.

El símbolo k denota comúnmente el intervalo de selección en muestreo sistemático. Se usará $f = 1/F$ para las fracciones de selección y se tiene que $k = F = N/n$. Con el intervalo $k = F$, se selecciona una unidad de cada zona de $k = F$ unidades. Y

como se mencionó anteriormente, el arranque aleatorio r de 1 a k imparte a cada unidad la probabilidad de selección de $1/k = 1/F = f$. Este método es de fácil aplicación y es una de sus principales ventajas. Otra ventaja es que produce fácilmente una muestra proporcional y si se toma en cuenta que la muestra se reparte uniformemente en la población mediante un ordenamiento correspondiente de los elementos de la población. Por ejemplo, considérese una muestra de una lista de nombres en orden alfabético, tras la aplicación del muestreo sistemático se producirá aproximadamente la misma proporción de nombres de cada letra.

En la figura 2 se muestra un ejemplo del muestreo sistemático :

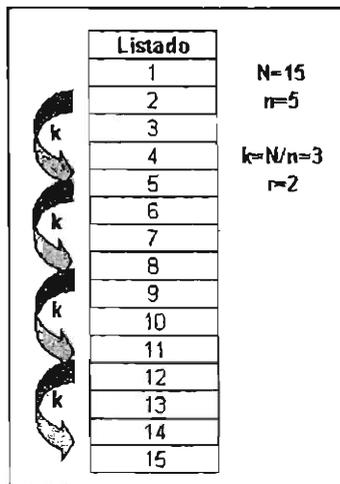


Figura 2 Ejemplo del muestreo sistemático

Se tiene un listado de 15 elementos, se requiere seleccionar 5 elementos, con el arranque $r = 2$ y $k=3$.

Los elementos seleccionados así son: 2, 5, 8, 11 y 14.

La muestra reflejará cualquier estratificación que exista en el ordenamiento de la lista de la población. Los límites específicos de la cantidad de estratificación quedan definidos mediante la selección de una unidad de cada conjunto de k unidades consecutivas. Las fracciones de k unidades producen una o cero selecciones, y en la proporcionalidad hay un error absoluto de menos de una unidad para cualquier estrato. En un estrato de ck unidades el error relativo es menor que $\frac{1}{c}$; es por tanto pequeño cuando los estratos son grandes. También se infiere que si se ordena con flexibilidad la lista de la población, se puede obtener estratos de k unidades relativamente homogéneos al interior. O bien, se puede hacer que cada conjunto de $2k$ unidades sea homogéneo.

Si el tamaño de la población N no es múltiplo entero de n , se presenta un problema y el muestrista debe seleccionar la forma más conveniente de resolverlo:

1. Permitir que el tamaño de la muestra sea n o bien $(n+1)$. Entonces se escoge k de manera que $nk < N < (n+1)k$. El arranque aleatorio es el que determina que el tamaño de muestra sea n o bien $(n+1)$.
2. Eliminar con el método de selección con igual probabilidad de selección (MESIP) suficientes unidades para reducir los listados a exactamente

nk elementos antes de seleccionar con el intervalo k . La probabilidad de selección sobre los dos procedimientos es $\frac{n}{N}$. En lugar de la eliminación de elementos puede resultar más conveniente seleccionar algunos listados con MESIP y agregar estos duplicados al final de la lista.

3. Considerar la lista como si fuera circular, es decir, que la última unidad será seguida por la primera. Se selecciona un arranque aleatorio de 1 a N , se agregan los intervalos k hasta seleccionar exactamente n elementos; al llegar al final de la lista se continúa con el principio. Cualquier intervalo conveniente k resultará en un *MESIP* de n elementos seleccionados con probabilidad de $\frac{n}{N}$. Por lo general, el entero k más próximo a la razón

$\frac{N}{n}$ será el más conveniente. Esta forma de solucionar el problema es la más

flexible y la más comúnmente usada.

4. El uso de intervalos fraccionarios es sencillo con una fracción decimal. Por ejemplo, supóngase que para seleccionar una muestra de $n = 100$ elementos de una población de $N = 920$ unidades se aplica el intervalo $k = \frac{N}{n} = \frac{920}{100} = 9.2$. Se selecciona un arranque aleatorio de 1 a 92 y se agrega el intervalo 92 sucesivamente, entonces se redondean estos números eliminando los últimos dígitos y se obtiene exactamente $n = 100$.

En muchas situaciones, además de una gran variación irregular, hay una tendencia de las unidades próximas entre sí a parecerse una a la otra, y a desviarse en forma semejante de la media de la población. Por ejemplo, los archivos que se ordenan cronológicamente están sujetos a una variación con respecto a la dimensión del tiempo. Una lista de empleados puede reflejar homogeneidades parciales de los grupos de trabajo. Las listas de direcciones de miembros de alguna suscripción pueden estar en orden geográfico. La numeración de las manzanas de una ciudad, distrito tras distrito, establece una lista con divisiones geográficas.

Este modelo supone la existencia de una división de la población en estratos y de una mezcla de las unidades de la muestra dentro de los estratos.

Existe el **Modelo de selecciones pareadas**, el fundamento de este modelo se basa en suponer que cada par sucesivo de selecciones fue tomado al azar, dos de cada estrato implícito o zona. Este modelo supone que dentro de cada zona los $2k$ elementos se hayan mezclado completamente antes de seleccionar dos de ellos. Los pares de selecciones que se comparan se determinan por el orden de la selección; la primera selección se contrasta con la segunda; la tercera con la cuarta, y así hasta la $(n - 1)$ con la n . El número de las zonas es $H = \frac{n}{2}$. Los pares pueden designarse con los subíndices ha y hb , y h toma los valores $1, 2, \dots, n/2$.

Estos son algunos de los métodos básicos de estratificación de uso más frecuente, en la siguiente subsección presentamos un método alternativo que ha sido aplicado

en problemas de mercadeo y con el cual se han obtenido resultados que han permitido dirigir certeramente estrategias de promoción.

1.3.5 Curva de Lorenz e índice de Gini.

La Curva de Lorenz ha sido usada entre otros autores por Atkinson (1970) y Shorrocks (1983) para tratar de medir la desigualdad y bienestar en diversas distribuciones de ingreso. La gran utilidad de esta curva radica en que muestra criterios intuitivos que pueden usarse en distribuciones con cierto ordenamiento de datos o en un *ranking* sin suposiciones paramétricas de la forma específica del índice de desigualdad o de la función de bienestar social⁶ (el nivel de bienestar de una comunidad se evalúa en función de indicadores socioeconómicos y tomando como base otra comunidad de referencia, entre los cuales están la esperanza de vida al nacer, relacionada estrechamente con el nivel de salud y de nutrición, así como el ingreso familiar; la tasa de mortalidad infantil, vinculada a condiciones de vida y acceso a servicios de salud, y la tasa de analfabetismo, ligada a la calidad de vida, el empleo, el ingreso y la salud).

⁶ *Por bienestar social se puede entender la satisfacción de necesidades cualitativas y cuantitativas de un grupo de personas en una sociedad, o bien el acceso y el consumo de bienes y servicios, las relaciones sociales y la calidad del ambiente. Un bienestar social mínimo se alcanza cuando se cubren las necesidades básicas de los individuos, a saber: alimentación, salud, educación y acceso al suelo para la vivienda*

La importancia de esta curva para la medición de desigualdad se destaca en el artículo que escribió Atkinson⁷ en 1970, en donde establece que la curva representa la distribución acumulada creciente de los ingresos según la acumulación creciente de la población que los percibe. Actualmente la Curva es utilizada para tratar de medir desigualdad en campos como la medicina, la educación y la economía, entre otros.

En los estudio sobre la desigualdad, se dispone de diversos métodos para describir la forma en que se distribuye el ingreso entre los diferentes grupos de individuos en una sociedad, entre otros: los diagramas de dispersión, los indicadores de desigualdad y los ordenamientos de la información.

Elaborar una diagrama para visualizar la distribución del ingreso resulta una opción extremadamente útil para el análisis de la desigualdad, ya que permite identificar ciertos aspectos de la forma de la distribución que de otra manera no sería posible apreciar.

Se contemplan al menos cuatro posibles opciones para generar ordenamientos de datos: las distribuciones de frecuencias, la curva de Lorenz, los denominados diagramas de desfile propuestos por Pen y la transformación logarítmica.

⁷ Atkinson, A.B. "On the measurement of inequality". *Journal of economic theory*, 1970.

La distribución de frecuencias es la manera más intuitiva de presentar las observaciones, aunque se pierde información, al agrupar el ingreso de individuos en diferentes intervalos de clase y no considerar la concentración de observaciones que se forman al interior de cada uno de ellos. Adicionalmente, las observaciones que se agrupan al interior de los intervalos quedan representadas por el punto medio o marca de clase.

En la medida de que el número de grupos es pequeño, la pérdida de información es mayor para este caso, este tipo de gráficos no aporta mucho al conocimiento de la forma en que se distribuyen las observaciones. Por el contrario, hacer uso de esta opción supone formar un número suficiente de intervalos de tal manera que se observen las características relevantes de la población asociadas a sus principales medidas de tendencia central y a la dispersión que existe en torno a ellas.

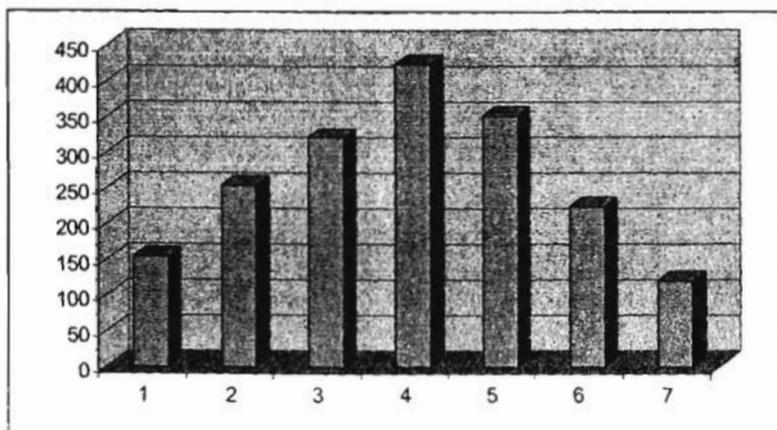


Figura 3 Distribución de frecuencias ingreso

La forma más habitual de representar la desigualdad es a partir de la Curva de Lorenz. Esta medida fue propuesta en 1905 con el propósito de ilustrar la desigualdad en la distribución de la salud y desde su aparición, su uso se ha popularizado entre los economistas.

En términos simples, la curva de Lorenz representa el porcentaje acumulado de ingreso ($\%Y_i$) recibido por un determinado grupo de la población ($\%P_i$) ordenado en forma ascendente de acuerdo con la cuantía de los ingresos ($y_1 \leq y_2 \leq \dots \leq y_n$).

La construcción de esta curva se efectúa como se muestra a continuación: Supóngase que se tienen n individuos ordenados en forma creciente respecto al valor de sus ingresos $y_1 \leq y_2 \leq \dots \leq y_n$ y se forman g grupos de igual tamaño, llamado percentiles⁸ (cabe mencionar que los quintiles y los deciles son las formas más comunes de ordenar a la población). La curva de Lorenz se define como la relación que existe entre las proporciones acumuladas de población ($\%P_i$) y las proporciones acumuladas de ingreso ($\%Y_i$).

De esta manera, en caso de que a cada porcentaje de la población le corresponda el mismo porcentaje de ingresos ($P_i = Y_i; \forall i$) se forma una línea de 45°, la cual divide en dos partes iguales el cuadrado de lado uno que se forma al graficar las proporciones acumuladas de personas en el eje horizontal (P_i) y de ingresos en la

⁸ Son valores que dividen a un conjunto de datos u observaciones en 100 partes iguales.

vertical (Y_i) . Dicha diagonal corresponde a lo que Lorenz definió como la **línea de equidad perfecta** y denota la ausencia de desigualdad.

En la figura 4 se observan tres elementos: la línea de equidad perfecta (que indica que al 10% de la población le corresponde el 10% de ingreso, al 20% de la población el 20%, etc.), la curva correspondiente a la distribución empírica formada a partir de las parejas (P_i, Y_i) , o curva de Lorenz, y el área entre las dos líneas, denominada área de concentración. A partir de la relación entre el área de concentración y la línea de igualdad perfecta, es posible derivar diversos indicadores que se utilizan para evaluar la concentración del ingreso, por ejemplo: mientras más pequeña sea esta área mejor distribuido está el ingreso, si la curva de Lorenz coincide con la línea de igualdad perfecta, se le conoce como *equidistribución*.

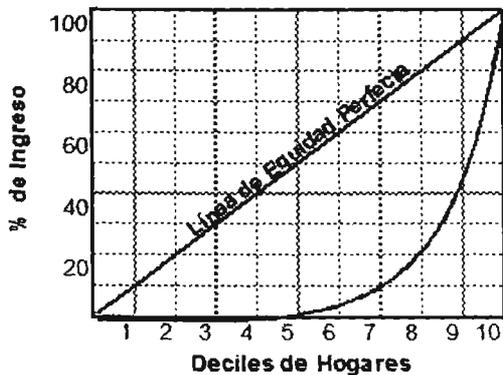


Figura 4 Curva de Lorenz

Desde el punto de vista empírico, la curva de Lorenz se construye graficando los porcentajes acumulados de ingresos que reciben los distintos grupos de la población, con la única condición de que éstos se definan con la misma amplitud, a efectos de evitar problemas asociados al número de observaciones agrupadas en cada intervalo. En el ámbito del análisis de la distribución del ingreso de las familias, es común que la curva de Lorenz se construya a partir de datos agrupados en subconjuntos iguales de tamaño 10%, denominados deciles de hogares. Por ejemplo, en la tabla 1.13 se presenta la información de la Encuesta Nacional de Ingreso Gasto en los Hogares (ENIGH)

**Distribución del ingreso corriente total trimestral de los hogares, 1992-2002
(precios de 2002)**

Decil	1992	1994	1996	1998	2000	2002
I	1.6	1.6	1.8	1.5	1.5	1.6
II	2.7	2.8	3.0	2.7	2.6	2.9
III	3.7	3.7	4.0	3.6	3.6	3.9
IV	4.7	4.6	4.9	4.7	4.6	4.9
V	5.7	5.7	6.0	5.8	5.7	6.1
VI	7.1	7.1	7.4	7.2	7.1	7.4
VII	8.9	8.7	9.0	8.9	8.8	9.3
VIII	11.4	11.3	11.5	11.5	11.2	11.9
IX	16.0	16.1	16.0	16.0	16.1	16.4
X	38.2	38.4	36.4	38.1	38.7	35.6
TOTAL	100.0	100.0	100.0	100.0	99.9	100.0

Fuente: INEGI, Encuesta de Ingreso y Gasto de los Hogares, Tercer Trimestre 1992, 1994, 1996, 1998, 2000 y 2002

Tabla 1.13 Información por deciles de hogares, ENIGH.

Si se grafican dos curvas de Lorenz y una queda totalmente contenida en la otra, salvo quizá en los valores extremos, véase figura 5, puede afirmarse que aquella que se ubica más cerca de la diagonal presenta una distribución más igualitaria, en

cuyo caso se dice que domina en el orden de Lorenz. En términos algebraicos lo

anterior se expresa diciendo que la curva $X \left(P_i, \sum_{i=1}^n y_i^x \right)$ domina a $Y \left(P_i, \sum_{i=1}^n y_i^y \right)$ en

el sentido u orden de Lorenz, si se cumple que:

$$\sum_{i=1}^n y_i^x \geq \sum_{i=1}^n y_i^y \text{ donde } y_i \text{ representa el ingreso de la familia } i.$$

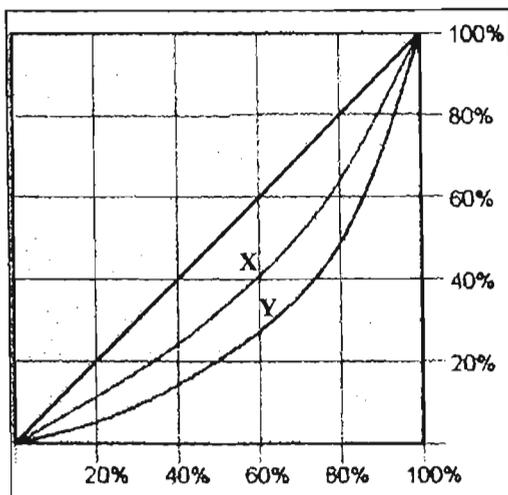


Figura 5 Dominancia en curvas de Lorenz

Sin lugar a dudas, uno de los índices más utilizados en el estudio de la desigualdad es el **Coefficiente de Concentración de Gini** que definió Gini en 1912 y que es posible también deducirlo desarrollando un procedimiento geométrico a partir de la curva de Lorenz, de la siguiente manera: el índice es igual a la proporción del área

de la curva de concentración con relación al total del área bajo la línea de equidad perfecta, como puede observarse en la figura 6.

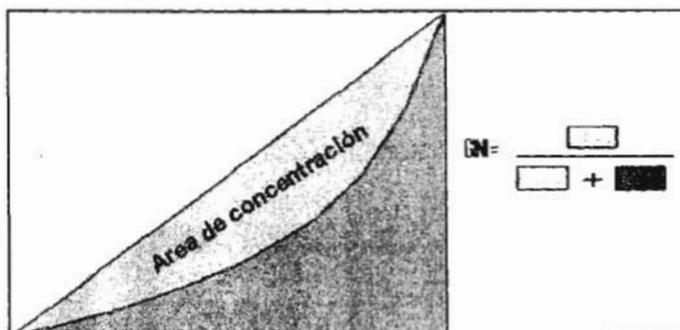


Figura 6 Índice de Gini a partir de la curva de Lorenz

Se define en los términos siguientes.

$$Gini = \frac{1}{2\mu} \left[\frac{\sum_{i=1}^n \sum_{j=1}^n |y_i - y_j|}{n(n-1)} \right] = \frac{1}{2\mu} \Delta$$

en donde Δ representa la media aritmética de las $n(n-1)$ diferencias absolutas de las observaciones y 2μ es el valor máximo que asume Δ cuando un individuo concentra todo el ingreso.

En 1914 Gini propuso un nuevo indicador de concentración que se define como uno menos dos veces el área bajo la curva de Lorenz, y demostró que era equivalente a

la fórmula $Gini = \frac{1}{2\mu} \Delta$.

El rango del índice de Gini es $[0,1]$, asume el valor de cero cuando se tiene la equidad perfecta y uno cuando todo el ingreso está concentrado en un individuo.

Entonces $Gini = 1 - (2F(y))$ donde $F(y)$ representa la curva de Lorenz; es decir, la proporción de individuos, familias, hogares, tiendas, etc. que tienen ingresos acumulados menores o iguales a y , en términos de estadística, la distribución de ingreso, $F(y)$ representa la probabilidad de que la variable aleatoria Y asuma un valor menor o igual a y , $[F(Y) = P(Y \leq y)]$.

A diferencia de las medidas estadísticas que tradicionalmente son utilizadas para estudiar la dispersión de un conjunto de datos, donde la media es el parámetro de localización que se utiliza como referencia, el índice de Gini se construye al comparar la distribución empírica que se forma con los datos observados y la línea de equidad perfecta que supone la distribución teórica derivada de la curva de Lorenz.

Nuevamente, supóngase que se tiene información sobre n individuos ordenados en forma creciente respecto al valor de sus ingresos: $y_1 \leq y_2 \leq \dots \leq y_n$. Además, se construye la distribución de frecuencias relativas simple y acumulada de la población en estudio (p_i, P_i) , así como de la variable a distribuir (en este caso el ingreso del hogar y_i e Y_i). Nótese que, debido al ordenamiento de los datos se cumple que $Y_i \leq Y_{i+1} \forall i$. Conforme a lo anterior, el índice de Gini se basa en la suma de las diferencias $(P_i - Y_i)$.

Con el propósito de estandarizar su recorrido al intervalo $[0,1]$, la expresión anterior se divide entre $\sum_{i=1}^{n-1} P_i$, dando paso a una de las fórmulas que comúnmente se utilizan para calcular el **índice de Gini** cuando se dispone de datos no agrupados, entonces otra forma de estimarlo es:

$$\text{Índice de Gini} = \frac{\sum_{i=1}^{n-1} (P_i - Y_i)}{\sum_{i=1}^{n-1} P_i}$$

El valor mínimo del indicador es cero y se obtiene cuando $P_i = Y_i, \forall i$ mientras que asume el valor 1 cuando $Y_1 = Y_2 = \dots = Y_{n-1} = 0$, ya que en este caso el individuo n es quien concentra todo el ingreso Y_n .

En el trabajo empírico es habitual que el cálculo de los índices de desigualdad se efectúe a partir de las observaciones agrupadas, que generalmente son los únicos datos disponibles.

Sin embargo, esta manera de proceder conduce necesariamente a la pérdida de información, debido a que los valores individuales de las observaciones serán sustituidos por algún valor que represente al intervalo en que se encuentran agrupados, por ejemplo, la marca de clase. De acuerdo a lo anterior, el procedimiento que se utiliza para el cálculo del índice de Gini para datos agrupados es la siguiente:

1. Ordenar los individuos, hogares, tiendas, etc., en forma descendente conforme a su ingreso.
2. Definir intervalos de igual tamaño (por ejemplo deciles, quintiles).
3. Construir la distribución de frecuencias relativas, simple y acumulada, de la variable a distribuir (ingreso), así como de la población. Si se opta por la formación de deciles de individuos, cada grupo deberá concentrar el 10% de las observaciones.
4. Calcular el índice de Gini conforme a alguna de las expresiones que existen para el caso en que se trabaja con datos agrupados.

Por ejemplo: $Gini = 1 - \sum_{i=1}^n x_i (Y_i + Y_{i+1})$ o bien $Gini = 1 - \sum_{i=1}^n \frac{(X_{i+1} - X_i)}{(Y_i + Y_{i+1})}$ en

donde n representa el número de grupos, x_i la proporción (o porcentaje) de población en el grupo i ; X_i la proporción acumulada de la población en el grupo i ; Y_i el ingreso acumulado en el grupo i , ver ejemplo en la tabla 1.14:

Decil	% de Hogares (P _i)	% de Ingreso (y _i)	% acumulado de Hogares (P)	% acumulado de Ingreso (Y _i)
1	10	y ₁	10	Y ₁ = y ₁
2	10	y ₂	20	Y ₂ = Y ₁ + y ₂
3	10	y ₃	30	Y ₃ = Y ₂ + y ₃
4	10	y ₄	40	Y ₄ = Y ₃ + y ₄
5	10	y ₅	50	Y ₅ = Y ₄ + y ₅
6	10	y ₆	60	Y ₆ = Y ₅ + y ₆
7	10	y ₇	70	Y ₇ = Y ₆ + y ₇
8	10	y ₈	80	Y ₈ = Y ₇ + y ₈
9	10	y ₉	90	Y ₉ = Y ₈ + y ₉
10	10	y ₁₀	100	Y ₁₀ = Y ₉ + y ₁₀

Tabla 1.14 Índice de Gini con datos agrupados

Para ejemplificar el caso de datos no agrupados, en la tabla 1.15 se presenta por hogar el ingreso total, su tamaño, representado por el número de habitantes y el ingreso per cápita⁹:

Hogar	Ingreso total	Tamaño del Hogar	Ingreso Per cápita
1	25	7	3.57
2	29	7	4.14
3	38	6	6.33
4	49	6	8.17
5	50	5	10.00
6	128	5	25.60
7	155	7	22.14
8	159	7	22.71
9	258	4	64.50
10	369	4	92.25
11	423	4	105.75
12	536	3	178.67
13	569	5	113.80
14	639	4	159.75
15	698	6	116.33
16	719	7	102.71
17	789	4	197.25
18	1259	3	419.67
19	1278	4	319.50
20	1295	5	259.00
21	1459	4	364.75
22	1594	3	531.33
23	2587	4	646.75
24	2589	4	647.25
25	3574	3	1191.33
26	3697	3	1232.33
27	4225	5	845.00
28	12369	4	3092.25
29	15632	5	3126.40
30	69845	4	17461.25
Totales	127036	142	31370.48
Media	4234.53		

Tabla 1.15 Ingreso total por hogar. Índice de Gini datos no agrupados

⁹ El ingreso per cápita se define como la relación entre el ingreso bruto interno y la población total.

Considerando el Ingreso Total como variable a distribuir, se tienen los siguientes resultados para las diferencias absolutas de las observaciones:

Hogares	Suma de diferencias absolutas
1	126286
2	126174
3	125940
4	125676
5	125654
6	124094
7	123608
8	123544
9	122158
10	120826
11	120286
12	119382
13	119184
14	118904
15	118786
16	118786
17	118926
18	120806
19	120920
20	121056
21	122696
22	124316
23	138218
24	138250
25	155980
26	158440
27	170056
28	365512
29	450350
30	1968314
Totales	6233128

Para aplicar la fórmula del Índice de Gini = $\frac{1}{2\mu} \left[\frac{\sum_{i=1}^n \sum_{j=1}^n |y_i - y_j|}{n(n-1)} \right] = \frac{1}{2\mu} \Delta$ falta

calcular el denominador completo, por lo que $2\mu n(n-1) = 7368088$.

Finalmente, se tiene:

$$Gini = \frac{6233128}{7368088} = 0.8460$$

que por ser un valor muy cercano a uno nos indica que el ingreso está distribuido muy inequitativamente.

El brasileño Lourival Camara¹⁰ en 1949 propuso una clasificación de la concentración con base en el Índice de Gini en estudios de las tierras en Brasil, que nos será útil para determinar cuando conviene separar la información en donde se concentra la información para estratificar, la clasificación se muestra en la tabla 1.16:

Índice de Gini	Clase	Separar
0.000-0.100	Concentración Nula Absoluta	No
0.101-0.250	Concentración Nula a Baja	No
0.251-0.500	Concentración Baja a Media	No
0.501-0.700	Concentración Media a Alta	Si
0.701-0.900	Concentración Alta a Fuerte	Si
0.901-1.000	Concentración Fuerte Absoluta	Si

Tabla 1.16 Clasificación de valores del índice de Gini

De esta manera cuando para el Índice de Gini se presente valores superiores a 0.500 se debe determinar el punto en el que se tiene la máxima distancia vertical

¹⁰ Camara, Lourival. *A concentracao da propriedade agraria no Brasil. Boletim Geográfico, Rio de Janeiro, 1949.*

entre la curva de Lorenz y la línea de equidad perfecta, que tomaremos como punto de corte¹¹, a esta distancia máxima se le conoce como el índice Robin Hood, dado que el valor acumulado en ese punto representa el total de ingreso que debe ser transferido de los individuos que se encuentran por encima de la media a aquellos que se encuentran por debajo para tener igualdad en la distribución de ingresos, ver la figura 7 en donde el punto D corresponde al punto de corte:

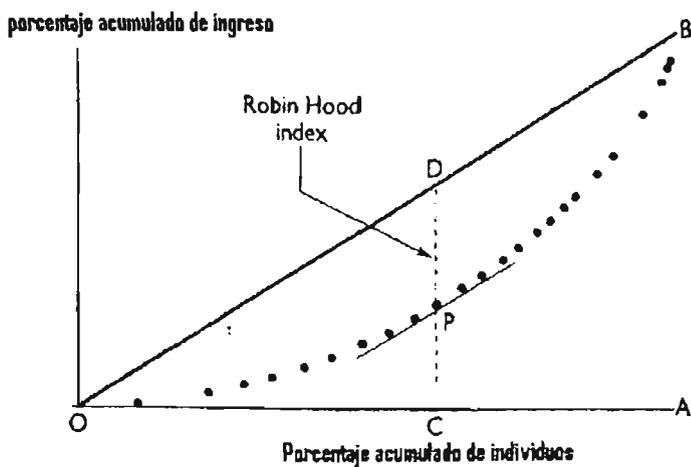


Figura 7 Índice Robin Hood

¹¹ Este punto se denomina punto de corte o turning point dado que es el punto después del cual la acumulación en las abscisas no representa una mayor concentración en las ordenadas.

Capítulo 2

Estudios de Mercado

La *American Marketing Association* define la Investigación de Mercados como la función que relaciona al consumidor, al cliente y al público con el especialista de investigación de mercados a través de la información, esta información se utiliza para identificar y definir las oportunidades y los problemas de mercadeo o *marketing*; generar, perfeccionar y evaluar las acciones de marketing; monitorear el desempeño del marketing; y mejorar la comprensión del marketing como un proceso.

Cada vez es más común que las organizaciones vean a la Investigación de Mercados como un proceso de investigación, este concepto sugiere que los recursos y las actividades de la organización sean enfocadas, de una manera integrada, en las *necesidades y deseos del consumidor*, contrario quizá a las necesidades y deseos de la organización. El problema central que enfrentan estas organizaciones es cómo monitorear las necesidades del mercado y anticiparse al futuro. Es muy útil, por ejemplo, saber en qué momento un producto determinado está teniendo fuerte demanda y, por lo tanto, altos agotamientos; información como ésta ayuda no sólo a tener un seguimiento del desempeño del producto, sino que facilita el proceso de toma de decisiones. Entonces se puede entender la

investigación de mercados como el conjunto de técnicas y metodologías necesarias para proporcionar información que permita conocer y comprender mejor el mercado.

2.1 Mezcla de *marketing*.

En este concepto están involucradas distintas variables, que se entienden como una propiedad que adquiere distintos valores en diferentes tiempos. Por ejemplo, una organización puede variar su presupuesto publicitario, o bien el precio de venta, o el color de la envoltura, o el gramaje. Una *variable independiente* es la causa de una *variable dependiente*, la cual es el efecto esperado. Las variables independientes se clasifican de acuerdo con la facilidad de manipulación o control. Las variables que se pueden controlar se identifican como los componentes de la mezcla de marketing del **producto, precio, plaza y promoción**. Estas variables son conocidas en su conjunto como las cuatro P de la Mercadotecnia y pueden tener, entre otras características, la siguiente composición:

Variable	Composición
<i>Producto</i>	Calidad, Características, Opciones, Estilo, Nombre y Marca, Embalaje, Gramaje, Garantías, Ganancias
<i>Precio</i>	Precio de lista, Descuentos, Facilidades, Plazos para pago, Términos de Crédito
<i>Plaza</i>	Canales, Cobertura, Ubicación, Inventario, Transporte
<i>Promoción</i>	Anuncios, Medios de Promoción, Muestras gratis, Publicidad

Hay un conjunto de variables independientes que no están bajo el control de la organización, son llamados factores situacionales y constituyen el estado de naturaleza al cual se debe adaptar la organización de ventas para planear e implantar un programa de marketing. El estado de naturaleza se compone de factores tales como la disponibilidad de energía, acciones competitivas, ambiente económico, tendencias del mercado y regulaciones gubernamentales.

Ambos conjuntos de variables independientes, la mezcla de marketing y factores situacionales, se combinan para determinar una respuesta de comportamiento tal como una compra, intenciones de compra, preferencias, actitudes, recompra y *switcheo*, que se refiere al cambio de una marca por otra de un mismo satisfactor o producto. Esta respuesta de comportamiento es la *variable dependiente* o el *efecto esperado*.

La respuesta de comportamiento está determinada por las consecuencias del comportamiento pasado, además de la influencia inmediata de las variables independientes. Por lo tanto, es necesario conocer la historia del desempeño del producto, así como tener un monitoreo constante de los efectos que puede tener el cambio en el color del empaque, o la presencia de demostradoras, o una inversión en promoción. El mercado es un entorno dinámico, que crea necesidades constantemente, por ejemplo comprar un shampoo en México hace aproximadamente diez años en un autoservicio se limitaba a determinar si se necesitaba para cabello normal o anticasca, además de seleccionar la marca de

preferencia; el mercado ha evolucionado principalmente para satisfacer las necesidades de las nuevas generaciones: hoy importa si se usa el cabello liso o chino, además si se mantiene su color con algún colorante artificial o no, si es largo o corto, etcétera. Y qué decir también acerca de los colorantes para cabello, que antiguamente estaban dirigidos al segmento de señoras mayores, hoy hay colores electrizantes para jóvenes, colorantes para hombre y para mujer, que no sólo son un colorante sino hasta un símbolo de status.

El siguiente ejemplo muestra la necesidad de la investigación de mercados y las repercusiones que puede traer para una marca una decisión tomada no sobre las necesidades y deseos del consumidor:

Las consecuencias que trae a una marca una decisión sobre las variables conocidas como las cuatro P de la Mercadotecnia que no esté fundamentada en un estudio de mercado se gestó en las oficinas corporativas de *The Coca-Cola Company* en Atlanta, el 3 de enero de 1985, la noticia fue que la antigua Coca-Cola dejaría de existir y en su lugar nacería la "New Coke", nadie sospechó que el producto, después de llegar a la meta excepcional de un siglo de fabricación bajo la misma fórmula, cambiaría de pronto los componentes de la fórmula que llevó a la compañía a su riqueza.

No tomaron en cuenta que para los estadounidenses, la bebida familiarmente conocida como *Coke* es un símbolo Nacional. ¿Por qué la Coca-Cola, con ventas de

7 mil 100 millones de dólares y una utilidad neta de 629 millones de dólares en 1984 decidió cambiar su sabor?. La nueva Coke resultó ser más dulce, menor efervescente y con un sabor semejante al de la Pepsi. Sus consumidores se sintieron decepcionados y una encuesta realizada por *USA Today* demostró que 59% de los consumidores preferían la antigua Coca-Cola, 25% la Pepsi y sólo 13% la nueva Coke. La reacción de los consumidores y la consecuente baja en las ventas provocó que *The Coca-Cola Company* regresara a su fórmula original renombrando al producto original como Coca-Cola Clásica. Ahora un proverbio de la Bolsa de Valores de Estados Unidos es "jamás cambien la fórmula de la Coca-Cola".

Peter Ducker¹² dice que la finalidad de la mercadotecnia es hacer que el vender resulte superfluo. El objetivo es conocer y comprender tan bien al cliente, que el producto o servicio sea una necesidad y se venda por sí solo. El concepto de Mercadotecnia sostiene que para lograr los objetivos de la organización es indispensable determinar las necesidades y los deseos del mercado meta y proporcionar las satisfacciones con mayor efectividad y eficiencia que los competidores.

2.2 Análisis de mercados

Una compañía que fabrica productos de consumo puede hacer uso del Análisis de Mercados para diferentes necesidades, por ejemplo:

¹² *Profesor y consultor especializado en estrategias de negocios.*

1. **Conocimiento del Mercado.** A este respecto es importante saber en qué canales debe colocarse el producto para que llegue a los consumidores meta; mediante el conocimiento de algunos productos sustitutos¹³ existentes ya en el mercado o mediante estudios con grupos de consumidores.

2. **Penetración en el Mercado.** Los estudios para conocer la penetración de un producto en el mercado abarca desde saber en qué canales se encuentra presente el producto, qué mecanismos de promoción están siendo adecuados, análisis de precios contra productos similares o sustitutos, tasa de compra y recompra, etc. Y los resultados proporcionan una guía para establecer mecanismos de promoción más adecuados, rebajas necesarias de precio, aumento en la publicidad, introducir el producto en más tiendas o lograr que queden mejor colocados en ellas con mayores espacios en anaquel, presencia en islas, cabeceras o botaderos, etc.

3. **Desarrollo de Mercados.** Estos análisis ayudan a identificar y desarrollar nuevos mercados para los productos actuales. Al analizar mercados demográficos -bebés, preescolares, adolescentes, adultos jóvenes, adultos y adultos mayores- se puede establecer impulsar a alguno de ellos para que compre o bien para que compre más de los productos de la marca. Los mercados institucionales -restaurantes, servicios alimentarios, hospitales, escuelas- para ver si podrían extenderse las ventas a este tipo de

¹³ *Un producto sustituto es aquel que ofrece características similares al producto en cuestión, de tal forma que podría ser una opción de sustitución para el consumidor.*

compradores. Y también existen los mercados demográficos -México, Guatemala, Japón- que pueden analizarse para establecer la posibilidad de posicionamiento en alguno de ellos.

4. **Diversificación.** Mediante este tipo de análisis es posible conocer las posibilidades de comenzar en un negocio totalmente ajeno a los productos o mercados actuales. Algunos fabricantes exploran mercados con necesidades diferentes a las que ellos atienden, toman en cuenta su capacidad instalada, su facilidad de posicionamiento, de distribución y de fabricación. Por ejemplo, en México The Coca Cola Company se diversificó, mediante una alianza estratégica con Nestlé, para la producción de leche saborizada y de té listo para beber.

La figura 8 ilustra la ayuda del Análisis de Mercados para la identificación de oportunidades:

	Productos Existentes	Nuevos Productos
	CONOCIMIENTO DEL MERCADO	
Mercados Existentes	Penetración de Mercado	Desarrollo de Productos
Nuevos Mercados	Desarrollo de Mercados	Diversificación

Figura 8 Detección de oportunidades en el mercado

La Mercadotecnia estudia las necesidades del consumidor y la capacidad de la compañía para satisfacerlas. Los consumidores constituyen el mercado total, el fabricante lo divide en segmentos más pequeños (que son grupos de clientes con necesidades, características o comportamientos propios) selecciona aquellos que más prometen y se centra en atenderlos y satisfacerlos por medio de sus productos, entonces mediante estudios y análisis de mercados diseña una mezcla de mercadotecnia compuesta de factores que están bajo su control: producto, precio, plaza y promoción.

Gran parte del éxito de un producto en el mercado responde al análisis cuidadoso de los consumidores. Si una compañía está estudiando mercados posibles para un nuevo producto primero debe calcular el tamaño presente y futuro del mercado y de sus diversos segmentos. Para calcular el tamaño del mercado debe identificar todos los productos sustitutos que son competidores, calcular la saturación del mercado con los productos existentes, calcular su nivel de ventas y decidir si el mercado permite detectar necesidades aún no satisfechas que pudieran redundar en una ventaja competitiva para el nuevo producto.

El potencial de crecimiento puede depender de la tasa de crecimiento de ciertos grupos de edad, nivel de ingreso, nivel de escolaridad, etc. que representen a seguros consumidores del producto.

También puede estar relacionado con cambios dentro del medio, como las condiciones económicas, los avances tecnológicos y el estilo de vida. El pronóstico o predicción de los que es más probable que hagan los consumidores bajo un conjunto de circunstancias resulta difícil, pero debe hacerse para tomar decisiones sobre el mercado. Si el pronóstico se ve favorable la compañía debe decidir cómo entrar al mercado. No pueden satisfacer a todos los consumidores de un mercado, o al menos, no a todos de la misma forma. Así, cada compañía debe dividir al mercado total, elegir los mejores segmentos y diseñar estrategias que le permitan atenderlos mejor que sus competidores.

Los consumidores pueden agruparse de diversas formas, según: factores geográficos (regiones o ciudades), demográficos (sexo, edad, nivel de ingreso o educación), psicográficos (clases sociales, estilos de vida) y de comportamiento (circunstancias de compra, beneficios que se buscan o tasas de utilización). A este proceso de clasificar a los clientes en grupos según sus diferentes necesidades, características o comportamientos es lo que se conoce como **Segmentación del Mercado**. No todas las formas de segmentar un mercado son igualmente útiles. Por ejemplo, a Tylenol no le serviría de nada distinguir entre consumidores femeninos y masculinos de analgésicos, si ambos responden de la misma manera al estímulo del mercado. En el mercado de automóviles, por ejemplo, los consumidores que eligen el auto más grande y el más lujoso sin importarles el precio forman un segmento de mercado; otro sería el de aquellos que se preocupan ante todo por el precio y la

funcionalidad. Una vez que una compañía ha definido sus segmentos de mercado, debe determinar a qué segmento(s) del mercado desea penetrar.

La selección del mercado meta implica la evaluación del atractivo que ofrece cada uno de los segmentos establecidos y la selección de alguno(s) de ellos. La mayor parte de las compañías penetran en un nuevo mercado atendiendo a un solo segmento, y si éste tiene éxito, le añaden otros; las compañías grandes, por su parte, suelen tratar de cubrir la mayoría de los segmentos posibles.

Una vez que se ha resuelto en qué segmentos se quiere penetrar, la compañía debe establecer sus estrategias de penetración y posicionamiento del mercado. La posición de un producto es el lugar que éste ocupa en la mente de los consumidores en comparación con los de la competencia. Si el producto nuevo parece ser exactamente igual a uno ya existente en el mercado, los consumidores no encontrarán una razón para comprar el nuevo en lugar al que ya venían usando. El posicionamiento consiste en lograr que un producto ocupe un lugar claro, distinto y superior en la mente de los consumidores meta en comparación con los demás productos sustitutos. Por ejemplo, BIC es una marca posicionada como "No Sabe Fallar", BIMBO como la compañía que fabrica pan "Con el Cariño de Siempre", FORD como los autos "Nacidos Fuertes". Estas afirmaciones aparentemente sencillas son la columna vertebral de la estrategia de mercadotecnia para posicionar un producto o una marca.

Ya que la compañía estableció sus estrategias de penetración y posicionamiento está lista para detallar su mezcla de mercadotecnia. Después la tarea se concreta en hacer planes a corto, mediano y largo plazo y accionarlos y establecer un control que mida resultados, evalúa resultados y permita tomar medidas correctivas.

La Información de Mercadotecnia está constituida por datos cotidianos sobre el desarrollo de los productos en el mercado, que ayuda a preparar y ajustar los planes. La compañía puede comprar información sobre participación de marcas, precio al detalle, porcentaje de tiendas en que está presente la marca, el share o participación del producto o la marca, etc. para un periodo determinado de tiempo; también existe información acerca de agotamientos de producto, de actividades promocionales de todas las marcas de una misma *Clase de Producto*¹⁴ o de un conjunto cerrado de marcas, que constituyan a las principales competidoras, además de información más específica sobre el consumidor por medio de un servicio de Panel de Hogares, en donde puede obtenerse su perfil, sus hábitos de consumo, los lugares en donde efectúa sus compras, así como las preferencias de compra, tasa de compra y recompra, lealtad a la marca, switcheo de marcas, etc.

Los fabricantes toman decisiones que están encaminadas a conseguir un mejor posicionamiento de sus marcas, mejorar los canales de distribución, establecer estrategias de promoción y/o reducciones temporales de precios, etc. que les proporcione ventajas competitivas en el mercado. La labor de quienes realizan estudios y análisis de mercados es ayudar a que los fabricantes entiendan mejor

¹⁴ Se usa para referirse al tipo de producto, por ejemplo Pañales Infantiles es la Clase de Producto y sus marcas Pampers, Huggies, Kleen Bebe, Kiddies, etc.

sus mercados, frecuentemente los estudios que solicitan tienen que ver con la delimitación de áreas accionables que les redunden en una ganancia significativa, por ejemplo, el primer caso que se presenta en el siguiente capítulo describe la necesidad de conocer cuales son las localidades de la República Mexicana en las que invertir en promoción, por su número de habitantes y volumen de ventas de la canasta de productos, significará un incremento en las ventas y descartar aquellas localidades que son muchísimas, que están muy dispersas y que no tienen una contribución significativa en las ventas de la canasta.

Se necesita una estratificación cuyo objetivo sea el foco en las localidades accionables; el fabricante tendrá entonces una clasificación que le permita la adecuada toma de decisiones.

Capítulo 3

Aplicaciones

Se seleccionó el Método de Índice de Gini y Curvas de Lorenz para hacer dos aplicaciones de mercado en las que se requiere de una estratificación. La primera aplicación es delimitar el espacio accionable para una compañía fabricante de artículos de consumo masivo como Shampoo, Jabón de Tocado, Detergente, etc. en relación con las zonas geográficas de la República Mexicana. La segunda aplicación es establecer una clasificación de tiendas de acuerdo con sus niveles de venta de las marcas de un fabricante licorero.

3.1 El México rural del fabricante X

La actividad promocional hecha por el fabricante X se ha dirigido a grandes centros urbanos. Ahora el fabricante X desea extender su promoción a localidades más pequeñas, por lo que se debe definir qué significa el México Rural para su negocio, con el fin de delimitar su espacio accionable, con ello puede responder a las preguntas: ¿Hasta dónde debe llegar la implantación de promociones?, ¿qué tipo de localidades deben ser parte de su programa?, ¿qué es el México Rural para su negocio?

Los objetivos específicos que se persiguen con esta delimitación son:

- Definir en términos del tamaño de población, número de habitantes, el tipo de localidades pequeñas que pueden representar un potencial de negocio para el fabricante X.
- Definir en términos del tamaño de población, el tipo de localidades pequeñas que NO representan un potencial de negocio para el fabricante X, las cuales quedarán excluidas de la definición anterior.
- Cuantificar las localidades urbanas y rurales que hay en México, según la definición del primer punto.
- Establecer las características demográficas y socioeconómicas de las localidades definidas como rurales para las necesidades del fabricante X que se tenga del Censo de Población y Vivienda del Instituto Nacional de Estadística Geografía e Informática (INEGI).

Se cuenta con información acerca del estimado del Universo de establecimientos comerciales que son el negocio del fabricante X, este estimado se obtuvo mediante una muestra nacional de localidades, la cual se denomina Muestra Maestra de Establecimientos. Para estimar las ventas de los establecimientos se cuenta con los datos de venta del fabricante X de algunas de las tiendas y se realiza una estimación de las ventas de la canasta conformado por las categorías: *detergente* (ropa, trastes), *pañales*, *papel higiénico*, *protección sanitaria*, *shampoo* y *jabón de tocador*.

3.1.1 Perfil geográfico de México

La extensión territorial de México, es de 1 964 375 km² de los cuales 1 959 248 km² son superficie continental y 5 127 km² corresponden a superficie insular.

El territorio de la República Mexicana está compuesto por 32 entidades federativas o estados, que a su vez están compuestas por 2, 443 Municipios, formados por 199,408 localidades. Para efectos de análisis, se realizó la agrupación del conjunto de localidades que forman un área metropolitana para formar una sola localidad. Por ejemplo: el área metropolitana de la Ciudad de México se considera como una localidad compuesta por las 16 delegaciones del Distrito Federal y 34 municipios conurbados del estado de México. Lo anterior implica trabajar sólo con 199,125 localidades. Según el XII Censo General de Población y Vivienda 2000 publicado por el INEGI hay un total de 21'954,733 viviendas y en ellas 97'483,412 habitantes.

El INEGI establece los estratos de localidades como: **Rurales**, aquellas con menos de 2500 habitantes y que no son cabeceras municipales y **Urbanas**, aquellas con 2500 habitantes o más o que son cabeceras municipales independientemente de su número de habitantes.

En la tabla 3.1 se presenta la distribución por tipo de localidad de las 199,408 localidades de la República Mexicana:

**ESTA TESIS NO SALE
DE LA BIBLIOTECA**

Tamaño de Localidad	Localidades	%	%Acum	
1-49 Hab.	133,175	66.79	66.79	RURALES
50-99 Hab.	15,415	7.73	74.52	
100-499 Hab.	33,781	16.94	91.46	
500-999 Hab.	8,700	4.36	95.82	
1000-1999 Hab.	4,482	2.25	98.07	
2000-2499 Hab.	814	0.41	98.47	
2500-4999 Hab.	1,580	0.79	99.27	
5000-9999 Hab.	711	0.36	99.62	URBANAS
10000-14999 Hab.	237	0.12	99.74	
15000-19999 Hab.	109	0.05	99.80	
20000-49999 Hab.	226	0.11	99.91	
50000-99999 Hab.	64	0.03	99.94	
100000-499999 Hab.	84	0.04	99.98	
500000-999999 Hab.	20	0.01	99.99	
1000000 y más Hab.	10	0.01	100.00	
TOTAL	199408	100		

Tabla 3.1 Distribución de Localidades según tamaño de localidad, porcentaje del total y porcentaje acumulado. Fuente: INEGI XII Censo General de Población y Vivienda 2000

En la tabla 3.2 se presenta la agrupación de localidades para conformar las áreas metropolitanas:

Total de localidades	Áreas Metropolitanas	Localidades Agrupadas	Localidades Resultantes
199,408	87	370	199,125

Tabla 3.2 Agrupación en áreas metropolitanas

Se tienen 87 áreas metropolitanas conformadas por 370 localidades; el número de localidades resultante es $199,408 - 370 + 87 = 199,125$.

En la tabla 3.3 se muestra la conformación del área metropolitana del Distrito Federal, que está conformada por las 16 Delegaciones del D.F. más 35 municipios conurbanos del estado de México, en total agrupa a 94 localidades:

Área Metropolitana del DF					
ESTADO	MUNICIPIO	MUNICIPIO	LOCAL	NLOCAL	
9	DF	2	AZCAPOTZALCO	1	AZCAPOTZALCO
9	DF	3	COYOACAN	1	COYOACAN
9	DF	4	CUAJIMALPA DE MORELOS	1	CUAJIMALPA DE MORELOS
9	DF	5	GUSTAVO A. MADERO	1	GUSTAVO A. MADERO
9	DF	6	IZTACALCO	1	IZTACALCO
9	DF	7	IZTAPALAPA	1	IZTAPALAPA
9	DF	8	MAGDALENA CONTRERAS, LA	1	MAGDALENA CONTRERAS, LA
9	DF	9	MILPA ALTA	1	VILLA MILPA ALTA
9	DF	9	MILPA ALTA	11	SAN ANTONIO TECOMITL
9	DF	9	MILPA ALTA	15	SAN BARTOLOME XICOMULCO
9	DF	9	MILPA ALTA	17	SAN FRANCISCO TECOXPA
9	DF	9	MILPA ALTA	24	SAN PABLO OZTOTEPEC
9	DF	9	MILPA ALTA	29	SAN PEDRO ATOCPAN
9	DF	9	MILPA ALTA	33	SAN SALVADOR CUAUHTENCO
9	DF	9	MILPA ALTA	36	SANTA ANA TLACOTENCO
9	DF	9	MILPA ALTA	152	SAN LORENZO TLACDYUCAN
9	DF	10	ALVARO OBREGON	1	ALVARO OBREGON
9	DF	11	TLAHUAC	1	TLAHUAC
9	DF	11	TLAHUAC	11	SAN ANDRES MIXQUIC
9	DF	11	TLAHUAC	21	SAN JUAN IXTAYOPAN
9	DF	11	TLAHUAC	24	SAN NICOLAS TETELCO
9	DF	11	TLAHUAC	26	SANTA CATARINA YECAHUITZOTL
9	DF	12	TLALPAN	1	TLALPAN
9	DF	12	TLALPAN	26	SAN MIGUEL AJUSCO
9	DF	12	TLALPAN	27	SAN MIGUEL TOPILEJO
9	DF	13	XOCHIMILCO	1	XOCHIMILCO
9	DF	14	BENITO JUAREZ	1	BENITO JUAREZ
9	DF	15	CUAUHTEMOC	1	CUAUHTEMOC
9	DF	16	MIGUEL HIDALGO	1	MIGUEL HIDALGO
9	DF	17	VENUSTIANO CARRANZA	1	VENUSTIANO CARRANZA
15	MEXICO	2	ACOLMAN	1	ACOLMAN DE NEZAHUALCOYOTL
15	MEXICO	2	ACOLMAN	5	SAN BARTOLO ACOLMAN
15	MEXICO	2	ACOLMAN	12	SANTA CATARINA
15	MEXICO	2	ACOLMAN	15	TEPEPAN
15	MEXICO	11	ATENCO	1	SAN SALVADOR ATENCO
15	MEXICO	11	ATENCO	2	SAN CRISTOBAL NEXQUIPAYAC
15	MEXICO	11	ATENCO	4	SANTA ISABEL IXTAPAN
15	MEXICO	13	ATIZAPAN DE ZARAGOZA	1	CIUDAD LOPEZ MATEOS
15	MEXICO	20	COACALCO DE BERRIOZABAL	1	SAN FRANCISCO COACALCO

ESTADO	NEDO	MUN	NMUN	LOCAL	NLOCAL
15	MEXICO	24	CUAUTITLAN	1	CUAUTITLAN
15	MEXICO	24	CUAUTITLAN	88	SAN MATEO IXTACALCO
15	MEXICO	25	CHALCO	1	CHALCO DE DIAZ COVARRUBIAS
15	MEXICO	28	CHIAUTLA	1	CHIAUTLA
15	MEXICO	29	CHICOLOAPAN	1	CHICOLOAPAN DE JUAREZ
15	MEXICO	30	CHICONCUAC	1	CHICONCUAC DE JUAREZ
15	MEXICO	31	CHIMALHUACAN	1	CHIMALHUACAN
15	MEXICO	33	ECATEPEC DE MORELOS	1	ECATEPEC DE MORELOS
15	MEXICO	37	HUIXQUILUCAN	1	HUIXQUILUCAN DE DEGOLLADO
15	MEXICO	37	HUIXQUILUCAN	5	DOS RIOS
15	MEXICO	37	HUIXQUILUCAN	9	JESUS DEL MONTE
15	MEXICO	37	HUIXQUILUCAN	26	ZACAMULPA
15	MEXICO	37	HUIXQUILUCAN	71	NAUCALPAN DE JUAREZ
15	MEXICO	39	IXTAPALUCA	1	IXTAPALUCA
15	MEXICO	44	JALTENCO	1	JALTENCO
15	MEXICO	44	JALTENCO	2	SANTA MARIA TONANITLA
15	MEXICO	44	JALTENCO	20	ALBORADA JALTENCO
15	MEXICO	53	MELCHOR OCAMPO	1	MELCHOR OCAMPO
15	MEXICO	57	NAUCALPAN DE JUAREZ	1	NAUCALPAN DE JUAREZ
15	MEXICO	58	NEZAHUALCOYOTL	1	CIUDAD NEZAHUALCOYOTL
15	MEXICO	59	NEXTLALPAN	1	SANTA ANA NEXTLALPAN
15	MEXICO	60	NICOLAS ROMERO	1	VILLA NICOLAS ROMERO
15	MEXICO	69	PAPALOTLA	1	PAPALOTLA
15	MEXICO	70	PAZ, LA	1	REYES ACAQUILPAN, LOS
15	MEXICO	81	TECAMAC	1	TECAMAC DE FELIPE VILLANUEVA
15	MEXICO	81	TECAMAC	9	SAN PABLO TECALCO
15	MEXICO	81	TECAMAC	19	OJO DE AGUA
15	MEXICO	81	TECAMAC	25	SAN MARTIN AZCATEPEC
15	MEXICO	91	TEOLOYUCAN	1	TEOLOYUCAN
15	MEXICO	91	TEOLOYUCAN	10	SAN BARTOLO
15	MEXICO	92	TEOTIHUACAN	1	TEOTIHUACAN DE ARISTA
15	MEXICO	92	TEOTIHUACAN	2	ATLATONGO
15	MEXICO	92	TEOTIHUACAN	19	SAN LORENZO TLALMIMILPAN
15	MEXICO	93	TEPETLAOXTOC	1	TEPETLAOXTOC DE HIDALGO
15	MEXICO	93	TEPETLAOXTOC	3	CONCEPCION JOLALPAN
15	MEXICO	95	TEPOTZOTLAN	1	TEPOTZOTLAN
15	MEXICO	95	TEPOTZOTLAN	21	SAN MATEO XOLOC
15	MEXICO	99	TEXCOCO	1	TEXCOCO DE MORA
15	MEXICO	99	TEXCOCO	29	SAN MIGUEL COATLINCHAN (COATLINCHAN)
15	MEXICO	99	TEXCOCO	42	SANTIAGO CUAUTLALPAN
15	MEXICO	99	TEXCOCO	45	TULANTONGO
15	MEXICO	100	TEZOYUCA	1	TEZOYUCA
15	MEXICO	100	TEZOYUCA	2	TEQUISISTLAN
15	MEXICO	104	TLALNEPANTLA DE BAZ	1	TLALNEPANTLA
15	MEXICO	108	TULTEPEC	1	TULTEPEC
15	MEXICO	108	TULTEPEC	14	SANTIAGO TEYAHUALCO
15	MEXICO	109	TULTITLAN	1	TULTITLAN DE MARIANO ESCOBEDO
15	MEXICO	109	TULTITLAN	3	BUENAVISTA

ESTADO	NEDO	MUN	NMUN	LOCAL	NLOCAL
15	MEXICO	109	TULTITLAN	25	SAN PABLO DE LAS SALINAS
15	MEXICO	109	TULTITLAN	68	FUENTES DEL VALLE
15	MEXICO	120	ZUMPANGO	1	ZUMPANGO DE OCAMPO
15	MEXICO	120	ZUMPANGO	54	SAN SEBASTIAN
15	MEXICO	121	CUAUTITLAN IZCALLI	1	CUAUTITLAN IZCALLI
15	MEXICO	121	CUAUTITLAN IZCALLI	20	HUILANGO
15	MEXICO	122	VALLE DE CHALCO SOLIDARIDAD	1	XICO

Tabla 3.3 Localidades que conforman el área metropolitana del Distrito Federal

En la tabla 3.4 se muestran los acumulados de localidades y población para rural y urbana, según el ámbito del INEGI en términos absolutos y relativos considerando sólo las 199,125 localidades:

Ámbito INEGI	LOCALIDADES		POBLACIÓN	
	Abs.	%	Abs.	%
Urbana (>=2500 hab.)	2,896	1.5	75,062,227	77
Rural (<2500 hab.)	196,229	98.5	22,421,185	23
Total	199,125	100.0	97,483,412	100

Tabla 3.4 Acumulados de localidades y población según ámbito del INEGI.

El Marco Geoestadístico Nacional¹⁵ está conformado por áreas geoestadísticas divididas en 3 niveles de desagregación:

¹⁵ El Marco Geoestadístico Nacional es un sistema que permite relacionar la información estadística con el espacio geográfico correspondiente, divide al territorio nacional en áreas de fácil identificación en campo y es adecuado para las actividades de captación de información. El Marco Geoestadístico ha evolucionado desde su creación, en el año de 1978 y en la actualidad ha alcanzado sus objetivos básicos: ser un instrumento único, de carácter nacional, cuya expresión gráfica, la cartografía censal, garantiza la cobertura y referencia geográfica de la información estadística.

- **Área Geoestadística Estatal (AGEE).** Es el área geográfica que contiene todos los municipios pertenecientes a una entidad federativa. **Conformación:** México está conformado por 32 entidades federativas (32 AGEE), a cada una de ellas se les ha asignado una clave consecutiva, formada por dos números, de acuerdo con el orden alfabético de sus nombres oficiales. **Límites:** Sus límites se adaptan en lo posible a los político-administrativos y están trazados sobre rasgos físicos fácilmente identificables en campo. La tabla 3.5 contiene las entidades federativas con su clave:

Clave	AGEE
01	Aguascalientes
02	Baja California
03	Baja California Sur
04	Campeche
05	Coahuila de Zaragoza
06	Colima
07	Chiapas
08	Chihuahua
09	Distrito Federal
10	Durango
11	Guanajuato
12	Guerrero
13	Hidalgo
14	Jalisco
15	México
16	Michoacán de Ocampo
17	Morelos
18	Nayarit
19	Nuevo León
20	Oaxaca
21	Puebla
22	Querétaro de Arteaga
23	Quintana Roo

Clave	AGEE
24	San Luis Potosí
25	Sinaloa
26	Sonora
27	Tabasco
28	Tamaulipas
29	Tlaxcala
30	Veracruz de Ignacio de la Llave
31	Yucatán
32	Zacatecas

Tabla 3.5 Entidades federativas con su clave

- Área Geoestadística Municipal (AGEM).** Es el área geográfica que corresponde a la división municipal de cada entidad federativa, el número total de los AGEM por estado es igual al total de sus municipios, y en el caso de Distrito Federal, al total de las delegaciones políticas. **Conformación:** Está conformada por todas las localidades rurales y urbanas que pertenecen a cada uno de los municipios y delegaciones, actualmente existen en el país 2,443 municipios. **Límites:** Sus límites son de fácil identificación en el terreno y permanentes, coincidiendo en la medida de lo posible con los político-administrativos.
- Área Geoestadística Básica (AGEB).** Área geográfica que corresponde a la subdivisión de la AGEM. Constituye la unidad básica del Marco Geoestadístico Nacional y dependiendo de las características que presenta (principalmente en densidad de población y uso de suelo), se clasifican en dos tipos: **Urbana** y **Rural**. La **Urbana** es el área geográfica ocupada por un

conjunto de manzanas¹⁶ que generalmente son de 1 a 50, perfectamente delimitadas por calles, avenidas, etc. Este tipo de AGEB se asigna en áreas geográficas que tengan una población de 2,500 habitantes o más y cuyo uso del suelo sea principalmente habitacional, industrial, de servicios, comercial, etc. La **Rural** es una extensión territorial que puede llegar a tener hasta 10,000 hectáreas y contener un conjunto de localidades con menos de 2,500 habitantes cada una, asentadas en terreno de uso generalmente agropecuario o forestal; en su interior se puede encontrar localidades urbanas, rurales y extensiones naturales como pantanos, lagos, desiertos, etc. , sus límites están constituidos por elementos perdurables y de fácil identificación en el terreno, pudiendo ser rasgos naturales como ríos, arroyos, barrancas, etc. y culturales como vías de ferrocarril, líneas de conducción eléctrica, carreteras, brechas, veredas, ductos, límites prediales, etc.

Para el fabricante X querer llegar a todas las localidades rurales, en los términos definidos por el INEGI, resulta muy difícil y casi imposible pues representan el 98.5% del total de localidades y concentran el 23% de la población total; el gasto de penetración sería muy costoso y con abarcar sólo las localidades urbanas estaría llegando con sus productos al 77% de la población. El fabricante X acude a solicitar un estudio de mercado que le permita conocer cuáles son las localidades que de

¹⁶ Es el espacio geográfico que está constituido por un grupo de viviendas y/o edificios, predios, lotes o terrenos destinados a diversos usos, tales como: habitacional, comercial, industrial, de servicios, etc.; se considera como la unidad mínima estadística del trabajo operativo para censos y encuestas. Está delimitada por calles, andadores, brechas, veredas, cercas, arroyos, límites de parcelas, etc. y que pueden rodearse caminando.

ese 98.5% son accionables, es decir, que no le significaran un gasto excesivo y que al mismo tiempo le redundará en un incremento significativo en la cobertura en términos de cobertura de población y ventas.

El fabricante X quiere que se le ayude a definir su México Rural, en la siguiente sección se presenta la metodología empleada, así como algunos de los principales resultados.

3.1.2 Metodología

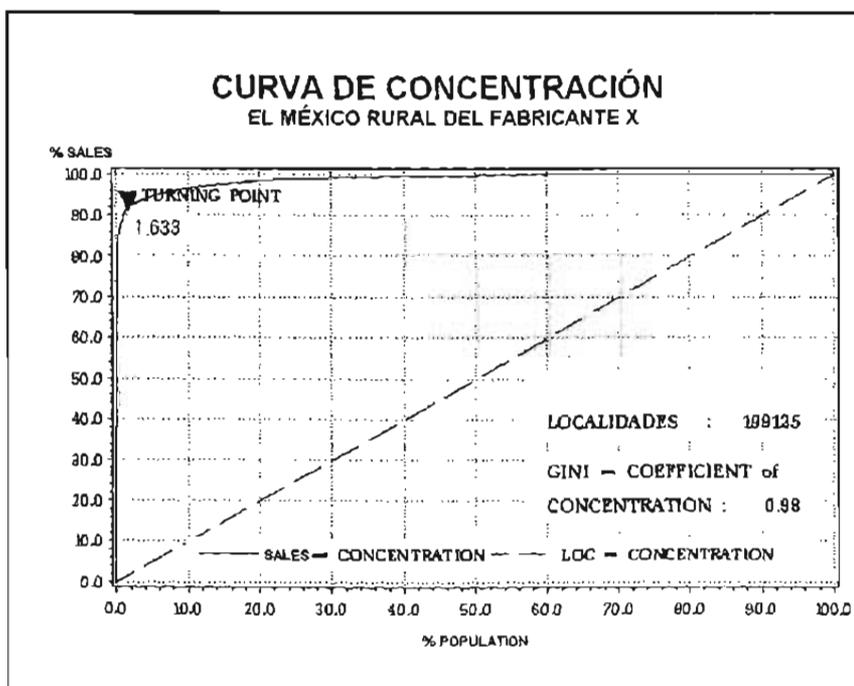
Se programó en el paquete estadístico SAS el método para obtener las curvas de concentración, los cortes en la curva y el índice de Gini y el objetivo es identificar, mediante una serie iterativa de curvas de Lorenz de las ventas de la canasta, los puntos de corte que separan las localidades con la mayor concentración, de aquellas localidades con poca concentración de ventas.

En cada gráfica se ordenan las localidades en forma descendente por número de habitantes en el eje de las x y por esta forma de ordenamiento se presenta la curva de Lorenz por encima de la línea de equidad perfecta y no por debajo como se presenta usualmente. Por otro lado, se acumuló el valor estimado de las ventas de la canasta en el eje de las y .

Para recordar la clasificación propuesta por Camara de los valores del índice de Gini, se presenta la tabla 1.16:

Índice de Gini	Clase	Separar
0.000-0.100	Concentración Nula Absoluta	No
0.101-0.250	Concentración Nula a Baja	No
0.251-0.500	Concentración Baja a Media	No
0.501-0.700	Concentración Media a Alta	Si
0.701-0.900	Concentración Alta a Fuerte	Si
0.901-1.000	Concentración Fuerte Absoluta	Si

Tabla 1.16 Clasificación de valores del índice de Gini

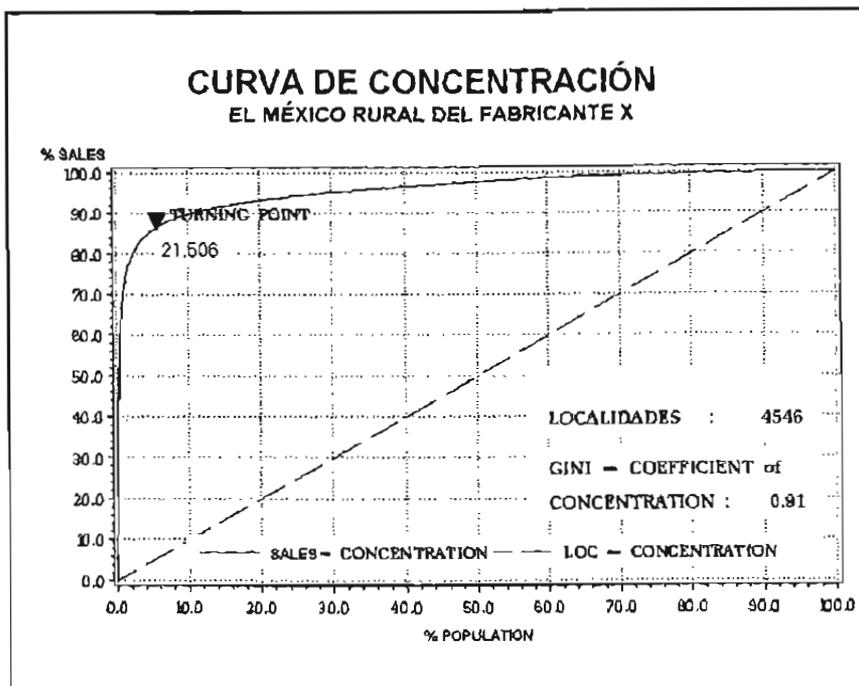


Gráfica 1

La Gráfica 1, es la primera que acumula las ventas de la canasta de las 199,125 localidades y marca el punto de corte en una localidad de 1,633 habitantes. Las localidades a la izquierda de este punto son el 2.3% del total y acumulan el 98% de

las ventas de la canasta. Desde el punto de vista operativo, el 2.3% de localidades, o sea 4,546 localidades, resulta un número grande y geográficamente disperso en el que se incluye gran cantidad de localidades pequeñas, dado que la clasificación de este valor, según Camara, es de Concentración Fuerte Absoluta, debemos separar las localidades y construir la siguiente gráfica eliminando las localidades a la derecha del punto de corte para generar un nuevo índice de Gini.

En la Gráfica 2 se consideran las 4,546 más grandes y se encontró un nuevo punto de corte en una localidad de tamaño de 21,506 habitantes. Un total de 292 localidades se encuentran al lado izquierdo del punto, mismas que representan un 87% de las ventas de la canasta entre las 4,546 más grandes, 91% en las ventas de la canasta nacional. El valor del índice de Gini sigue correspondiendo a una Concentración Fuerte Absoluta, por lo que debemos separar las localidades a partir del turning point.



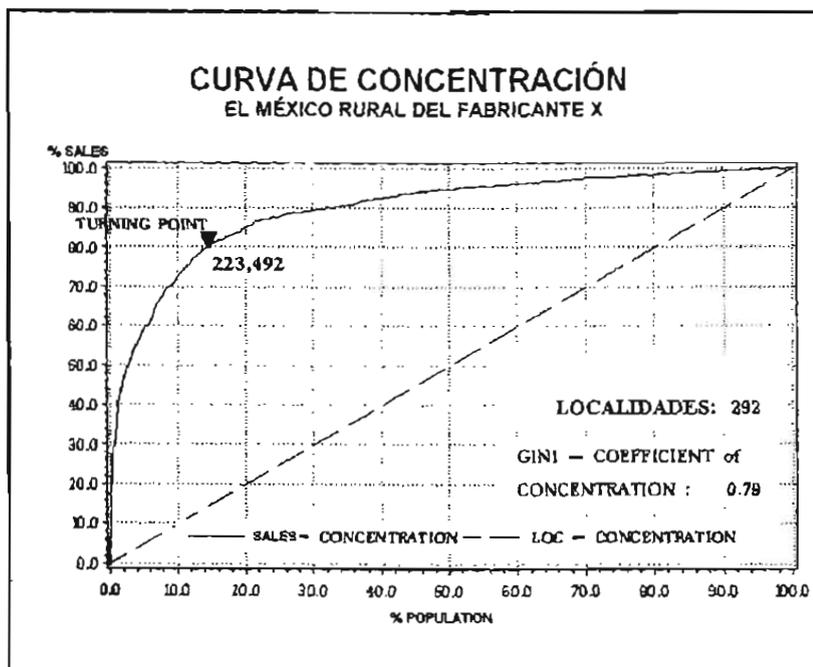
Gráfica 2

De manera sucesiva se continuó con el procedimiento hasta llegar al punto de corte de una localidades con 1'148,681 habitantes. Gráficas 3 y 4 .

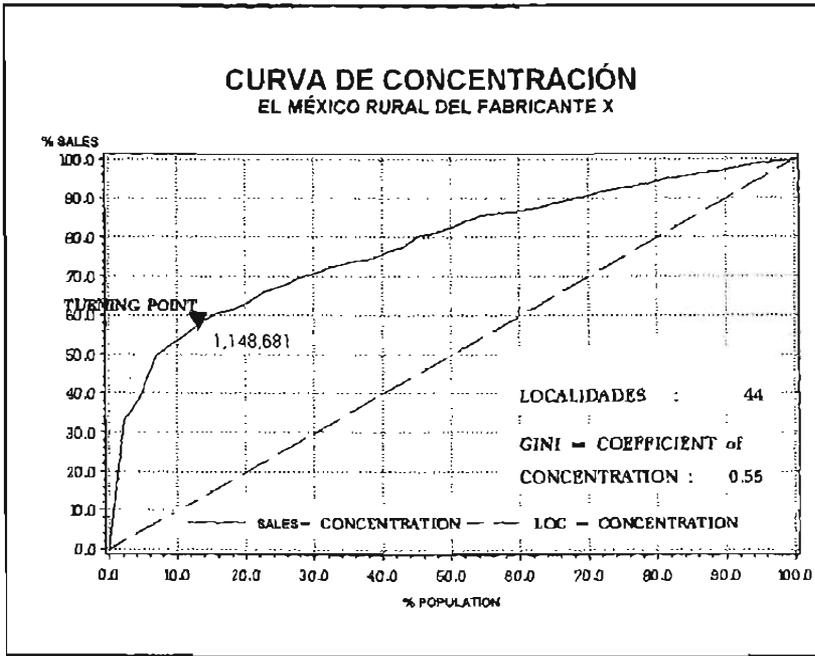
En la primera curva se detectaron las localidades con escasa aportación a la concentración de las ventas; se eliminaron dichas localidades para elaborar la siguiente curva de concentración.

En la segunda curva se hizo un proceso análogo y así sucesivamente hasta decidir cuál sería el último corte. En cada curva se calculó en Índice de Gini para medir el grado de concentración de las ventas de la canasta. Así mismo se ubicó el punto de

corle o tamaño de localidad a partir de la cual el crecimiento de las ventas de la canasta era menor al crecimiento del número de localidades.



Gráfica 3



Gráfica 4

Del ejercicio de curvas de concentración descrito anteriormente, se tomaron los primeros seis cortes para analizar la importancia de cada uno, de acuerdo con el peso en población, dispersión de localidades y al porcentaje de ventas de la canasta que representan, el resumen se presenta en la tabla 3.6:

Cortes	Rango de Habitantes	Número de Localidades	POBLACIÓN		%Acum de Población	Ventas del Canasto %	%Acum de Ventas del Canasto
			Número de Habitantes	%			
1	1'000,000 ó más	7	28,743,297	29.5%	29.5%	38.1%	38.1%
2	250,000 a 999,999	34	16,180,078	16.6%	46.1%	30.2%	68.2%
3	150,000 a 249,999	17	3,242,487	3.3%	49.4%	7.7%	76.0%
4	50,000 a 149,999	65	5,244,834	5.4%	54.8%	8.3%	84.3%
5	15,000 a 49,999	273	7,104,016	7.3%	62.1%	11.2%	95.5%
6	1 a 14,999	198,729	36,972,278	37.9%	100.0%	4.5%	100.0%
		199,125	97,486,990	100%			

Tabla 3.6: Resumen de Cortes

El corte 6 es el que representa a las localidades más pequeñas, acumula el mayor número de localidades y por tanto la mayor dispersión en el territorio del país. Aún cuando representa un 37.9% de la población total, significa sólo el 4.5% de las ventas de la canasta. Por lo anterior y por características que se verán más adelante, se establece que estas localidades sean las **NO ACCIONABLES**, dado que son muchísimas y no representan una ganancia significativa en las ventas de la canasta.

Se estableció el corte 5 (de 15,000 a 49,999 habitantes) como el grupo de localidades más pequeñas para el análisis.

Por su bajo peso en población, los cortes 3 y 4 se unieron en un sólo grupo de análisis y por el número de localidades que acumulan los cortes 1 y 2 también se unieron en un sólo grupo de análisis.

En resumen, los grupos seleccionados para el análisis comparativo son los cuatro que se presentan en la tabla 3.7:

Grupos	Perfil de Análisis	Rango de Habitantes	Número de Localidades	POBLACIÓN		% Acum de Población	Ventas del Canasto %	% Acum de Ventas del Canasto
				Número de Habitantes	%			
1	SD y C	250,000 ó más	41	44,923,375	46.1%	46.1%	68.2%	68.2%
2	SD y C	50,000 a 249,999	82	8,487,321	8.7%	54.8%	16.1%	84.3%
3	SD y C	15,000 a 49,999	273	7,104,016	7.3%	62.1%	11.2%	95.5%
4	SD	1 a 14,999	198,729	36,972,278	37.9%	100.0%	4.5%	100.0%
			199,125	97,486,990	100%		100%	

SD: Sociodemográfico
C: Comercial

Tabla 3.7: Resumen de Grupos de Análisis

El perfil sociodemográfico se analizó para los cuatro grupos de localidades y se presenta un resumen en el anexo A. Por lo que toca al perfil comercial por categorías de producto, se analizará principalmente en los tres primeros grupos.

Estos resultados fueron aplicados en la estrategia de penetración del fabricante X inicialmente y han servido también para dirigir sus estrategias promocionales de manera efectiva. La retroalimentación que se recibió por parte del fabricante es que ahora conoce mejor su mercado, mediante estratos claros y de fácil visualización, para los cuales, además, se publica información de consumidores de manera oficial por diversas fuentes.

3.2 Clasificación de tiendas de un fabricante licorero.

La actividad promocional que lleva a cabo un fabricante licorero, como por ejemplo la pintura de las fachadas, la colocación de anuncios luminosos de sus marcas, la colocación de estantes especiales para sus productos, etc. debe estar relacionada con el tamaño de las tiendas de acuerdo con el nivel de ventas de sus productos; el fabricante licorero requiere de una clasificación de las tiendas que le permita establecer paquetes promocionales adecuados para cada tipo de tienda, que ha querido establecer como: grandes, medianas y pequeñas. Para ello solicita un estudio de mercado que le genere los resultados esperados: el grupo de tiendas que corresponden con cada uno de los tamaños.

Para llevar a cabo este estudio se cuenta con las ventas que el fabricante tiene en cada una de las tiendas (la cantidad en ventas en botellas, en litros y en pesos para cada marca); así como la auditoría de ventas correspondiente a una muestra de tiendas que nos reporta las botellas vendidas en cada tienda, el precio por botella y el inventario al inicio y al final del periodo de auditoría (que se llevó a cabo de manera semanal en cada tienda con reporte mensual durante ocho meses). El conjunto de tiendas que conforman el dominio de estudio son tiendas especializadas o semiespecializadas en la venta de licores, como licorerías o tiendas de abarrotes (incluyendo las tienditas de la esquina, los minisupers y las tiendas de venta al público las 24 horas) con venta de licor específicamente en el Distrito Federal y la zona conurbada de la Ciudad de México.

3.2.1 Metodología

Al realizar una auditoría de ventas en las tiendas resulta muy complicado tener a una persona encargada que nos reporte qué se ha vendido y cuánto; se podría preguntar al tendero o a la persona encargada de las ventas, sin embargo la información que pudiera proporcionarnos depende de situaciones como el humor del tendero, su buena memoria o la cantidad de clientes que lleguen a estar en la tienda durante la visita del auditor; información obtenida de esta manera no es confiable y no sirve para la correcta clasificación de las tiendas.

Se ofreció un paquete de artículos promocionales como pluma, gorra y camiseta por parte del fabricante para facilitar la colaboración de las tiendas. Se diseñó un plan de visita a las tiendas de acuerdo con los días de la semana, con el objetivo de captar las ventas normales de entre semana y las ventas especiales de fin de semana (motivadas por algún partido de fútbol, de reuniones familiares, etc.) y se estableció un rol de visitas para que todas las tiendas tuvieran la visita en un día diferente de semana a semana, mostrado en la tabla 3.8.

TIENDA	SEM1	SEM2	SEM3	SEM4	SEM5	SEM6	SEM7	SEM8	SEM9	SEM10	SEM11	SEM12	SEM13
1001	MIE	JUE	VIE	SAP	DOM	MAR	MIE	JUE	VIE	SAB	DOM	MAR	MIE
1003	SAB	DOM	MAR	MIE	JUE	VIE	SAB	DOM	MAR	MIE	JUE	VIE	SAB
1006	DOM	MAR	MIE	JUE	VIE	SAB	DOM	MAR	MIE	JUE	VIE	Sap	DOM
1023	MAR	MIE	JUE	VIE	SAB	DOM	MAR	MIE	JUE	VIE	SAB	DOM	MAR
1026	DOM	MAR	MIE	JUE	VIE	SAB	DOM	MAR	MIE	JUE	VIE	Sap	DOM
1028	MIE	JUE	VIE	SAP	DOM	MAR	MIE	JUE	VIE	SAB	DOM	MAR	MIE
1029	DOM	MAR	MIE	JUE	VIE	SAB	DOM	MAR	MIE	JUE	VIE	SAP	DOM
1031	VIE	SAB	DOM	MAR	MIE	JUE	VIE	SAB	DOM	MAR	MIE	JUE	VIE
1036	MAR	MIE	JUE	VIE	SAB	DOM	MAR	MIE	JUE	VIE	SAB	DOM	MAR

TIENDA	SEM1	SEM2	SEM3	SEM4	SEM5	SEM6	SEM7	SEM8	SEM9	SEM10	SEM11	SEM12	SEM13
1048	VIE	SAB	DOM	MAR	MIE	JUE	VIE	SAB	DOM	MAR	MIE	JUE	VIE
1062	JUE	VIE	SAB	DOM	MAR	MIE	JUE	VIE	SAB	DOM	MAR	MIE	JUE
1068	SAB	DOM	MAR	MIE	JUE	VIE	SAB	DOM	MAR	MIE	JUE	VIE	SAB
1073	MIE	JUE	VIE	SAB	DOM	MAR	MIE	JUE	VIE	SAB	DOM	MAR	MIE
1086	JUE	VIE	SAB	DOM	MAR	MIE	JUE	VIE	SAB	DOM	MAR	MIE	JUE
1087	VIE	SAB	DOM	MAR	MIE	JUE	VIE	SAB	DOM	MAR	MIE	JUE	VIE

Tabla 3.8 Roles de visita a tiendas

En la primera visita se levanta en las tiendas el inventario (que se denominará inventario inicial) de las marcas de licor en cuestión, del fabricante se tienen las notas de compra de la tienda y en cada visita semanal el auditor levantará el inventario correspondiente a la semana. Los datos con los que se cuenta para la semana i , son: Inventario Inicial (II_i), Compras (C_i) e Inventario Final (IF_i), cabe señalar que el IF de la semana i será considerado el II de la semana $i+1$, con la siguiente notación: $IF_i = II_{i+1}$

La primera semana de visita sólo se levantará el II y se considerará esta semana como la semana cero, las ventas de la primera semana se calculan como:

$$Ventas_1 = II_0 + C_1 - IF_1$$

de esta forma se pueden calcular las ventas en botellas que se realizaron en la primer semana, de tal forma que si $II_0 = 15$ botellas, se tiene el registro de compras

de la tienda con $C_1 = 12$ botellas y el resultado de la auditoría fue $IF_1 = 5$ botellas , entonces sustituyendo los valores en la fórmula se tiene:

$$Ventas_1 = 15 + 12 - 5 = 22 \text{ botellas}$$

En la auditoría se presenta un nivel de detalle semejante al que presenta la nota de venta del licorero a la tienda, es decir, la marca del licor, el gramaje y el precio. De tal forma que se pueda tener el registro de litros vendidos, así como el total de venta en pesos. Para ello se calcula un factor especial para convertir a litros las botellas, por ejemplo si una botella es de litro y medio, entonces su factor será 1.5, si el contenido corresponde a 360 ml el factor es de 0.350, etc. Una vez que se cuenta con las ventas convertidas, o ventas en litros y con el precio por litro se puede establecer las ventas en pesos de las marcas del licorero.

Por ejemplo, se tiene la venta en botellas de la marca X con las siguientes características:

Gramaje	Botellas	Precio por litro
350 ml.	12	\$22.00
750 ml.	15	\$22.00
1000 ml.	21	\$22.00
1500 ml.	10	\$22.00

El factor para cada gramaje es:

350 ml.	0.35
750 ml.	0.75
1000 ml.	1.00
1500 ml.	1.50

Para tener el número total de litros de la marca X se multiplica el número de botellas de cada gramaje por su factor y se suman los resultados:

$$\text{Total Litros} = (12 * 0.35) + (15 * 0.75) + (21 * 1.00) + (10 * 1.50)$$

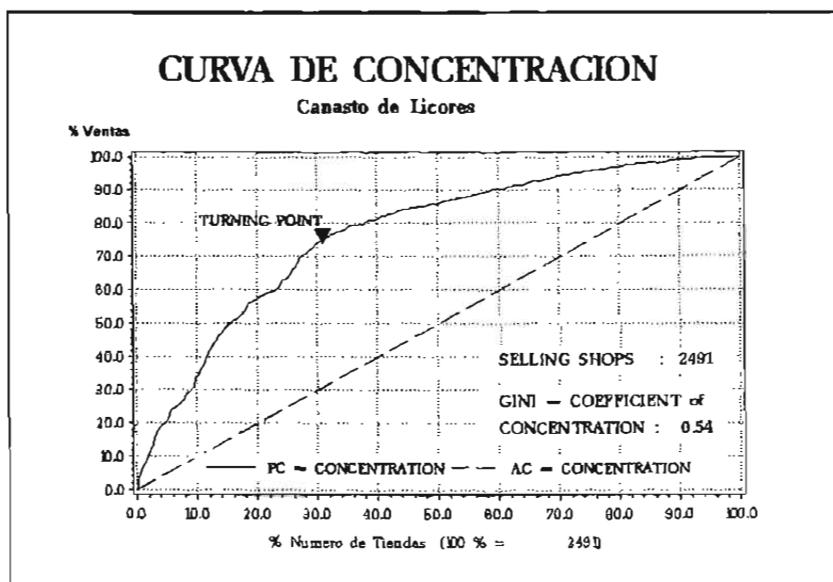
$$\text{Total Litros} = 4.20 + 11.25 + 21 + 15 = 51.45 \text{ litros}$$

Para tener las ventas en pesos que representa esa venta de botellas, se multiplica el Total Litros por el precio por litro:

$$\text{Ventas}_x = \text{Total Litros} * \text{Precio por litro} = 51.45 * 22 = \$1131.9$$

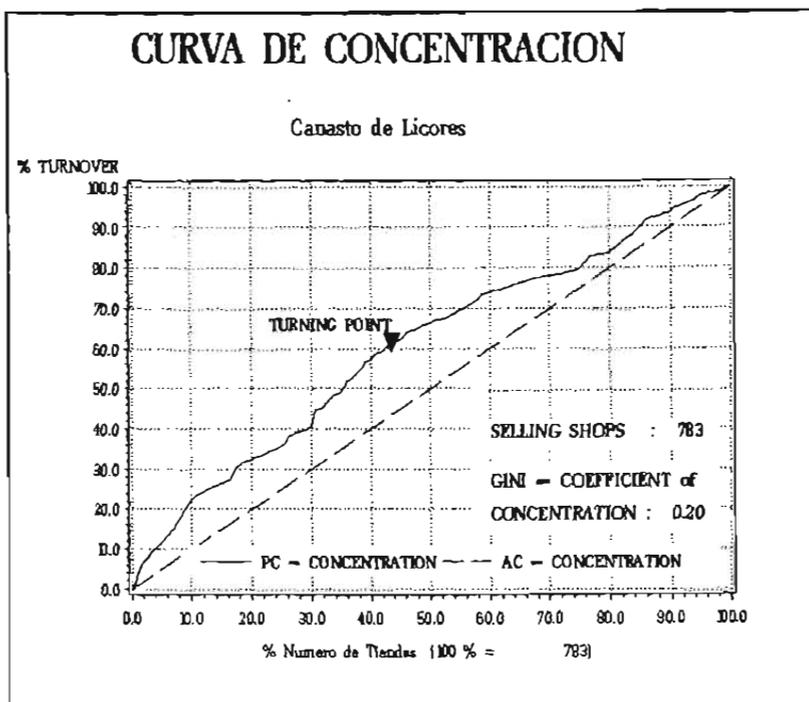
Este cálculo se debe hacer para cada tienda/marca por semana y hacer el acumulado para todas las marcas del fabricante para las cuatro semanas para tener el reporte mensual; para cada tienda se deberá calcular el promedio mensual de ventas de los ocho meses de duración del estudio y las gráficas de concentración se realizarán con base en el promedio mensual de ventas de las tiendas.

Se realizó el análisis sobre 2491 tiendas y con las ventas de las tiendas se construye la gráfica de concentración, que se muestra en la gráfica 5:



Gráfica 5

En la gráfica 5 el punto de corte (turning point) es en una tienda con ventas de \$21,962.3, las tiendas a la izquierda de este punto son el 31.43% del total y acumulan el 54% de las ventas de licores (de las ventas de la muestra concentran el 88.9%). Dado que el valor del índice de Gini es 0.54 y en la clasificación de Camara corresponde a una Concentración media a alta y si debemos separar las tiendas a la derecha del punto de corte, 1708 tiendas, y generando una nueva curva de concentración se obtiene la gráfica 6:



Gráfica 6

En la gráfica 6 el punto de corte se da en una tienda que vende \$70,157.2 , las tiendas a la izquierda de este punto son el 30% de las 783 tiendas y acumulan el 20% de las ventas del canasto de licores (de las ventas de la muestra un 66.4%).

Como el índice de Gini tiene un valor de 0.20 y corresponde a una concentración nula a baja y tenemos indicado no separar, esta iteración es la última.

En la tabla 3.9 se muestran los tres grupos de tiendas que se obtuvieron:

Grupo	Clasificación de Ventas	# Tiendas	% Tiendas
Grandes	de \$70,157.2 y más	235	9.43%
Medianas	de \$21,962.3 y menores de \$70,157.2	548	22.00%
Pequeñas	menores de \$21,962.3	1708	68.57%
		2491	100.00%

Tabla 3.9 Resumen de Grupos

Esta clasificación fue aplicada a las tiendas según sus promedios de venta y resultó muy útil no sólo para la aplicación de la actividad promocional, sino también para establecer la mezcla de productos adecuada a cada tipo de tienda.

Como pudo observarse en ambos ejemplos, la clasificación de valores del índice de Gini resulta ser una guía para determinar si el conjunto de observaciones puede separarse a partir del punto de corte, formando así un nuevo estrato. Esta alternativa de estratificación ha sido aplicada en algunos problemas como los que se han presentado, obteniendo resultados prácticos para las necesidades del mercado.

Anexo A

Perfil Comercial

Localidades con al menos un autoservicio	<p>Un indicador del perfil comercial, es el porcentaje de localidades del estrato que cuentan con al menos un autoservicio de cadena, ya sea nacional, regional o local. En la siguiente tabla se observa que el estrato de localidades más grande cuenta con el 100% de localidades con al menos un autoservicio, mientras que tres cuartas partes del estrato dos lo tienen. A medida que disminuye el tamaño de una localidad (en número de habitantes), la presencia de autoservicios baja, hasta ser casi nula en el estrato de localidades con menos de 15,000 habitantes.</p> <table border="1" data-bbox="373 760 990 986"> <thead> <tr> <th rowspan="2">Estrato</th> <th rowspan="2">Habitantes</th> <th rowspan="2">Número de localidades</th> <th colspan="2">Localidades con al menos un autoservicio de cadena</th> </tr> <tr> <th>Abs</th> <th>%</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>250,000 y más</td> <td>41</td> <td>41</td> <td>100.0%</td> </tr> <tr> <td>2</td> <td>50,000 a 249,999</td> <td>82</td> <td>63</td> <td>76.8%</td> </tr> <tr> <td>3</td> <td>15,000 a 49,999</td> <td>273</td> <td>38</td> <td>13.9%</td> </tr> <tr> <td>4</td> <td>1 a 14,999</td> <td>198,729</td> <td>7</td> <td>0.0%</td> </tr> </tbody> </table>	Estrato	Habitantes	Número de localidades	Localidades con al menos un autoservicio de cadena		Abs	%	1	250,000 y más	41	41	100.0%	2	50,000 a 249,999	82	63	76.8%	3	15,000 a 49,999	273	38	13.9%	4	1 a 14,999	198,729	7	0.0%
Estrato	Habitantes				Número de localidades	Localidades con al menos un autoservicio de cadena																						
		Abs	%																									
1	250,000 y más	41	41	100.0%																								
2	50,000 a 249,999	82	63	76.8%																								
3	15,000 a 49,999	273	38	13.9%																								
4	1 a 14,999	198,729	7	0.0%																								
Comercios manejantes de abarrotes	<p>Otro indicador del perfil comercial es el número de tiendas manejantes de abarrotes (grupo al que pertenecen las categorías del fabricante X: <i>detergente (ropa y trastes), pañales, papel higiénico, protección sanitaria, shampoo y jabón de tocador</i>); para el estrato 1 el 75% de los comercios son manejantes, el 17.8% en el estrato 2, los grupos 3 y 4 agrupan el 13.9% y el 32.2%, respectivamente.</p>																											

Perfil socio-demográfico

Distribución de la población	<p>En México existen localidades de muy variados tamaños; mientras que 41 grandes centros urbanos agrupan el 46.1% de la población del país (estrato 1), existen 198,729 localidades pequeñas dispersas en todo el territorio nacional (estrato 4) que agrupan un 37.9%.</p>
-------------------------------------	--

	<p>Además del potencial comercial que representa un grupo de localidades por concentración de ventas, el espacio accionable del fabricante X podrá ser definido con base en la dispersión y al tamaño de ellas. El esfuerzo de llegar a pequeñas comunidades no siempre bien comunicadas es mucho mayor que a los centros urbanos. El estrato o grupo recomendado como <i>NO ACCIONABLE</i> es el de las localidades menores a 15,000 habitantes por su altísima dispersión.</p>
Nivel de instrucción	<p>En el estrato de localidades más grande, 96 de cada 100 habitantes saben leer y escribir, mientras que en las localidades más pequeñas (estrato 4), 82 de cada 100 personas son consideradas analfabetas.</p> <p>La población que solo habla lengua indígena representa el 1.14% en las localidades pequeñas. El porcentaje es menor conforme la localidad de residencia es más grande.</p> <p>El nivel de escolaridad alcanzado por la población es mayor según aumenta el tamaño de localidad. Las condiciones menos favorecidas de escolaridad se observan en personas que radican en las localidades pequeñas (estrato 4) con un 18.3% de la población que no asistió a la escuela. A ello se añade un 28.5% de personas de 15 años o más que no completó la educación primaria.</p>
Ocupación	<p>La población de 12 años o más que realiza una actividad económica es mayor en las localidades del estrato 4 y desciende conforme se hace menor el tamaño de localidad donde reside.</p> <p>La población de las localidades de los grupos 1 a 3, se ocupa principalmente en actividades económicas del sector terciario como el comercio y el servicio, después en actividades del sector secundario como la industria de manufactura y transformación y finalmente, en el sector primario de agricultura, ganadería o pesca. La tendencia se invierte en las localidades del estrato 4, donde la mayoría de los trabajadores se clasifica en el sector primario de actividad económica.</p>
Material de construcción de las viviendas	<p>Una de cada diez viviendas de localidades pequeñas (estrato 4) tiene el piso de material de desecho y lámina de cartón en el techo. A mayor tamaño de localidad, menor porcentaje de</p>

	<p>construcciones con este material.</p> <p>Un piso de tierra en la vivienda es un indicador de desarrollo desfavorecido de la población, tres de cada diez viviendas tienen piso de tierra en las localidades del estrato 4, mientras que solo una lo tiene en las localidades del estrato 3, las proporciones son menores en los estratos 1 y 2.</p>
Número de habitaciones en la vivienda	<p>Contar con un cuarto específico para cocinar, sin que se utilice como dormitorio, es una condición que favorece el desarrollo y salud de la población. En las localidades de los estratos 2 a 4, cerca de una de cada diez viviendas cuenta con un solo cuarto. Esta proporción es del 3.7% en las localidades grandes.</p>
Combustible para cocinar	<p>Casi la totalidad de las viviendas de los grupos 1 a 3 usan gas para cocinar. En la mitad de las viviendas de las localidades pequeñas se usa gas y la otra mitad usa leña como combustible en la cocina.</p>
Disponibilidad de agua entubada, luz y drenaje	<p>Nueve de cada diez viviendas del estrato 1 de localidades cuenta con energía eléctrica, agua entubada y drenaje; mientras que sólo cuatro de cada diez viviendas del estrato 4 tienen los servicios básicos.</p>
Bienes domésticos en la vivienda	<p>La disponibilidad de bienes materiales en la vivienda que proporcionen comodidad, acceso a algunos medios masivos de comunicación o aligeren las tareas domésticas, puede ser un indicador de cómo hacer la promoción, hacia dónde orientar el esfuerzo y qué tipo de productos enfocar en las localidades.</p> <p>Al menos ocho de cada diez viviendas en localidades de los estratos 1 a 3 cuentan con radio o radiograbadora; nueve de cada diez cuenta con televisor, una tercera parte con videocasetera, atributo que alcanza a la mitad de las viviendas del estrato 1.</p> <p>En los estratos 1 a 3, que representan el espacio accionable del fabricante X, al menos siete de cada diez viviendas cuentan con refrigerador, poco más de la mitad tienen lavadora, una tercera parte tiene teléfono.</p>

Conclusiones

El objetivo de este trabajo de tesis fue logrado al proponer una alternativa de estratificación que es ampliamente usada en Economía y Ciencias de la Salud. Esta alternativa ofrece una parte gráfica que además de ser un excelente apoyo visual para conocer la concentración de la variable que se esté utilizando también resulta ser una herramienta rápida para generar resultados.

De una manera natural, el punto de corte (turning point) generado en la curva de Lorenz resulta ser una frontera del estrato y este punto de corte resulta ser el lugar en donde se genera la distancia máxima entre la curva de Lorenz y la recta de 45° (de equidad perfecta), es decir, el punto después del cual una mayor concentración de individuos no redundará en una concentración significativa de ingreso.

En la primera aplicación se usó como variable de concentración las ventas de la canasta de productos del fabricante X y de esta forma se obtuvieron estratos homogéneos al interior y heterogéneos entre sí, como indica la teoría del muestreo estratificado. Las iteraciones por lo general se acuerdan con el usuario (con base en las necesidades del mismo y la experiencia del investigador), pues es posible aumentar o disminuir el número de cortes resultantes.

En el **segundo** ejemplo se utilizó la misma técnica usando como variable de **concentración** las ventas de la canasta de licores, generando nuevamente estratos útiles para **los fines** del fabricante.

Las **herramientas** propuestas tienen un trasfondo matemático muy interesante que las hace **fuertes** candidatas a ser incluidas en el material estándar de muestreo dada su **utilidad** en estudios de mercado.

Bibliografía

- Atkinson, A. B., On the measurement of inequality. *Jornal of economic theory*, 1970.
- Atkinson, A. B., *The Economics of Inequality*. Second Edition, Oxford:Clarendon Press 1983.
- Boltvinik, Julio, *Revista electrónica de comercio exterior*, “*Opciones metodológicas para medir la pobreza en México*”, México, 2001.
- Boltvinik, Julio, *Revista electrónica de comercio exterior*, “*Tipología de los métodos de medición de la pobreza*”. México, 2003.
- Camara, Lourival. A concentracao da propriedade agraria no Brasil. *Boletim Geográfico*, Río de Janeiro, 1949.
- Castillo-Salgado, Carlos, “*Medición de desigualdades en salud: coeficiente de Gini e Índice de concentración*”. Documento electrónico, Sin país, Sin año.
- Cochran, William, *Técnicas de Muestreo*. Ed. CECSA, México, 1980.

- Kinnear, Thomas, *Investigación de Mercados un enfoque aplicado*, Ed. McGraw Hill, Colombia, 1981.
- Kish, Leslie, *Muestreo de Encuestas*. Ed. Trillas, México, 1975.
- Koolman, Xander, "*On the interpretation of a concentration index of inequality*". Documento electrónico. Holanda, 2003.
- Kotler, Philip, *Fundamentos de Mercadotecnia*, Prentice Hall, México 1991.
- MacArthur, Catherine et al, "Income Inequality", Documento electrónico www.macses.ucsf.edu/research/social%20environment/notebook/inequality.html
- Medina, Fernando, "*Consideraciones sobre el índice de Gini para medir la concentración del ingreso*", Comisión Económica para América Latina y el Caribe, documento electrónico, Chile, 2001.
- Medina, Fernando, "*Marcos Maestros de Muestreo en el contexto de un sistema integrado de encuestas de hogares*". Comisión Económica para América Latina y el Caribe, documento electrónico, Chile, Sin año.

- Nicolini, Giovanna, "A method to define strata boundaries". Documento electrónico, Italia, Sin año.
- Sánchez Villareal, Francisco, "Procedimiento para estratificación mediante particiones sucesivas en función de sumas de cuadrados dentro de estratos". Apuntes Congreso Nacional de la Sociedad Matemática Mexicana, México, 1999.
- Stanton, William, *Fundamentos de Marketing*, McGraw Hill, México, 1996.
- The United States Environmental Protection Agency, "Guidance for choosing a sampling design for environmental data collection". Documento electrónico, Estados Unidos, 2000. <http://www.epa.gov/quality/qs-docs/g5s-final.pdf>
- Winkler, William, "Strata boundary determination". Documento electrónico, Sin país, Sin año.

Páginas electrónicas:

- www.inegi.org.mx (Censo Población y Vivienda y datos de población del estado de Chiapas)
- <http://revistas-bancomext.gob.mx> (Artículos de la revista electrónica de Comercio Exterior)