

# Técnicas de Muestreo para una Investigación de Mercadeo

Administración de Empresas

Este material de autoestudio fue creado en el año 2006 para el programa Administración de Empresas y ha sido autorizada su publicación por el (los) autor (es), en el Banco de Objetos Institucional de la Universidad Pedagógica y Tecnológica de Colombia.



TECNICAS DE MUESTREO PARA UNA INVESTIGACION DE  
MERCADERO

JOSE ANTONIO ACOSTA JIMENEZ

TUNJA

1994

## CONTENIDO

	pág.
INTRODUCCION	5
INSTRUCCIONES GENERALES PARA USAR EL MODULO	6
OBJETIVOS GENERALES	7
UNIDAD 1. TERMINOS BASICOS	8
A. OBJETIVOS ESPECIFICOS	8
B. MUESTREO	8
C. CENSO	8
D. UNIVERSO O POBLACION	8
E. MUESTRA	9
F. ELEMENTO	9
G. UNIDAD DE MUESTREO	9
H. MARCO MUESTRAL	10
I. VARIABLE	10
J. MEDIR	11
K. ERROR DE MUESTREO	11
L. ERROR NO MUESTRAL	12
LL. CONFIANZA	12
M. ESTADIGRAFO	12
N. PARAMETRO	12
Ñ. INFERIR	13
O. EJERCICIO DE AUTOEVALUACION	13
UNIDAD 2. EL MUESTREO DENTRO DE UN PROCESO DE	

INVESTIGACION FORMAL	14
A. OBJETIVOS ESPECIFICOS	14
B. RELACION ENTRE DECISION, INFORMACION E INVESTIGACION	14
C. PARTES BASICAS DE UNA INVESTIGACION FORMAL	16
D. EJERCICIO DE AUTOEVALUACION	18
UNIDAD 3. MUESTREO ALEATORIO SIMPLE	19
A. OBJETIVOS ESPECIFICOS	19
B. DEFINICION	19
C. CALCULO DEL TAMAÑO DE LA MUESTRA	19
D. ELECCION DEL TAMAÑO DE LA MUESTRA	24
E. DISEÑO DEL FORMATO PARA REGISTRAR LA INFORMACION	25
F. ELABORACION DE LOS CUADROS DE TABULACION Y LOS CUADROS ESTADISTICOS	28
G. INFERENCIA	29
H. EJERCICIO DE AUTOEVALUACION	37
UNIDAD 4. MUESTREO ALEATORIO ESTRATIFICADO	38
A. OBJETIVOS ESPECIFICOS	38
B. DEFINICION	38
C. CALCULO DEL TAMAÑO DE LA MUESTRA	39
D. ELECCION DEL TAMAÑO DE MUESTRA	46
E. DISEÑO DEL FORMATO PARA REGISTRAR LA INFORMACION	47
F. ELABORACION DE LOS CUADROS DE TABULACION Y DE LOS CUADROS ESTADISTICOS	48
G. INFERENCIA	49
H. EJERCICIO DE AUTOEVALUACION	55
UNIDAD 5. OTROS METODOS ALEATORIOS O AL AZAR	

PARA ELEGIR UNA MUESTRA	56
A. OBJETIVOS ESPECIFICOS	56
B. GENERALIDADES	56
C. MUESTREO SISTEMATICO	56
D. MUESTREO POR CONGLOMERADOS	59
E. MUESTREO POR ETAPAS	64
F. MUESTREO POR FASES	65
G. METODOS MIXTOS	66
H. EJERCICIO DE AUTOEVALUACION	66
UNIDAD 6. METODOS DE MUESTREO NO ALEATORIOS O NO AL AZAR	67
A. OBJETIVOS ESPECIFICOS	67
B. CONCEPTO DE MUESTREO NO ALEATORIO O NO AL AZAR	67
C. MUESTREO DIRIGIDO	67
D. MUESTREO POR CUOTAS	67
E. MUESTREO POR CONVENIENCIA	68
F. EJERCICIO DE AUTOEVALUACION	68
BIBLIOGRAFIA	70
ANEXO # 1. CASO PARA ILUSTRAR LA APLICACION DE LA TEORIA	71
ANEXO # 2. NUMEROS ALEATORIOS	78
ANEXO # 3. DISTRIBUCION Z	79
ANEXO # 4. DISTRIBUCION T	80
ANEXO # 5. TABLA 8 - 6	81

## INTRODUCCION

Este Módulo guiará al estudiante en la estructuración de la parte de muestreo, dentro de la metodología en un plan de investigación formal.

El caso usado como ejemplo para la aplicación de la teoría es de Mercadeo, pero la teoría expuesta acá es aplicable, con sus ajustes respectivos, a cualquier investigación formal.

El contenido de éste Módulo es producto de las investigaciones bibliográficas hechas por el autor y la experiencia en la docencia presencial en las cátedras de Investigación de Mercadeo y Muestreo Administrativo.

Lo planteado en el Módulo es lo que en forma básica y común se usa en una investigación formal. La bibliografía citada permitirá, al lector, profundizar y refinar sus conocimientos sobre el tema.

El Módulo está compuesto de 6 unidades. La primera, se ocupa de definir los términos más comunes de Muestreo. La segunda, ubica el Muestreo dentro del proceso de investigación formal. De la tercera a la sexta, se plantean los diferentes métodos de muestreo para cálculo y elección de una muestra.

## INSTRUCCIONES GENERALES PARA USAR EL MODULO

1. Al comenzar cada unidad los objetivos específicos le indican lo que aprenderá. Dentro de cada unidad hay contenidos desarrollados que le permitirán lograr o alcanzar cada objetivo.
2. Fijese bien como se aplica la teoría al caso que se usa a lo largo de todas las unidades.
3. Use el mismo caso para hacer las autoevaluaciones. Esto permitirá ver, fácilmente, las diferencias de cada método.
4. Comente los resultados de cada ejercicio de autoevaluación con su tutor y demás compañeros.
5. Los textos citados en la bibliografía le permitirán profundizar y refinar sus conocimientos sobre el tema.
6. Al aplicar la teoría hay que ser creativos, pues los casos de la vida real son distintos.

## OBJETIVOS GENERALES

Al finalizar éste Módulo, el estudiante estará en capacidad de:

1. Calcular un tamaño de muestra por los métodos: aleatorio simple, aleatorio estratificado y no aleatorio.
2. Elegir un tamaño de muestra por los métodos: aleatorio simple, aleatorio estratificado, sistemático, polietapico, por conglomerados, por fases y no aleatorio.
3. Diseñar formatos para registrar información.
4. Tabular cuestionarios y estructurar los cuadros estadísticos respectivos.
5. Hacer inferencia por el método de estimación y por el método de prueba de hipótesis.



## UNIDAD 1

## TERMINOS BASICOS

## A. OBJETIVOS ESPECIFICOS.

Al finalizar esta unidad, el estudiante estará en capacidad de:

1. Definir cada uno de los términos básicos usados en muestreo.
2. Aplicar los términos a la resolución de un caso.
3. Diferenciar los términos entre si.

## B. MUESTREO.

Proceso para medir una serie de variables en una parte de un universo o población.

En comparación con un censo, el muestreo ahorra tiempo y dinero. Pero, existe la posibilidad de que se presente el error de muestreo, debido al hecho de trabajar con parte del universo.

## C. CENSO.

Proceso para medir una serie de variables en la totalidad del universo o población.

En comparación con el muestreo, el censo es más demorado y cuesta más. Pero, se elimina la posibilidad de que se presente el error de muestreo.

## D. UNIVERSO O POBLACION.

Conjunto de elementos de interés para una investigación.

Los elementos pueden ser personas, animales, vegetales, cosas, etc.

En el caso, los elementos que componen el universo son TARJETAS.

En cuanto al número de elementos, un universo puede ser de dos clases: Finito o infinito.

Un universo es o se considera FINITO, cuando se conoce el número de elementos y éste es menor de 100.000. En el caso, el universo se considera FINITO porque el número de elementos es 160.

Un universo es o se considera INFINITO, cuando no se conoce el número de elementos o éste es mayor de 100.000.

Un universo se define en términos de: Elementos, área geográfica y tiempo.

#### E. MUESTRA.

Parte de un universo o población que se calcula y se selecciona através de un proceso de muestreo.

#### F. ELEMENTO.

Es la persona, animal, vegetal, cosa, etc., de la cual se toma la información que se necesita. En el caso, el elemento es una cosa llamada TARJETA.

Si se esta trabajando en un proceso de muestreo el elemento se llama: Elemento muestral.

Si se esta trabajando en un proceso de censo el elemento se llama: Elemento censal.

En el caso, el elemento es cada una de las tarjetas que contiene información de cada uno de los socios.

#### G. UNIDAD DE MUESTREO.

Es la persona, animal, vegetal, cosa, etc., que se encuentra disponible para ser elegida en cada etapa de un proceso de muestreo.

Cuando el proceso de elección de la muestra tiene una sola etapa, el elemento y la unidad de muestreo son lo mismo.

En el caso, se podría elegir la muestra en dos etapas: En la primera, la unidad muestral sería cada una de las planas que contienen la información de las 160

tarjetas. En la segunda, la unidad muestral ( o elemento muestral ) sería cada una de las tarjetas.

H. MARCO MUESTRAL.

Es el listado de elementos y/o unidades de donde se puede elegir una muestra.

En el caso, el marco muestral son todas las páginas donde esta contenida información de las 160 tarjetas.

I. VARIABLE.

Es una característica de un elemento. La edad, el ingreso anual de la unidad familiar, el gasto semanal por comida, el número de personas de la unidad familiar, etc., son variables.

Generalmente, una variable tiene varios valores o alternativas. En el caso, la variable "existen personas entre 6 y 18 años en la unidad familiar" tiene dos valores o alternativas: Si y no.

Según la forma en que se expresan las alternativas de una variable, ésta puede ser de dos clases: Cuantitativa o cualitativa.

1. Variable cuantitativa.

Es aquella cuyos valores o alternativas son CIFRAS (de uno o más dígitos). En el caso, son cuantitativas las siguientes variables: Ingreso anual de la unidad familiar, gasto semanal por comida, número de personas de la unidad familiar y edad del cabeza de la unidad familiar.

Segun la clase de cifras (enteras o enteras y decimales) que puede tener cada alternativa, una variable puede ser: Discreta o continua.

Una variable es discreta cuando cada alternativa solo puede tener cifras enteras. Por ejemplo, el número de personas de la unidad familiar.

Es lógico que hayan 2, o 3, o 4 personas, pero no es lógico que hayan 1,5 o 1,85 personas.

Una variable es continua cuando cada alternativa puede tener cifras enteras y decimales. Por ejemplo, el gasto semanal por comida. Es lógico que una persona gaste \$ 9.870,55 semanalmente en comida.

Si las alternativas de una variable cuantitativa se

pueden verificarse se considera que es un hecho y si no se pueden verificar se considera una opinión. En el caso, todas las variables cuantitativas son hechos.

## 2. Variable cualitativa.

Es aquella cuyos valores o alternativas están compuestas de una o más palabras.

En el caso, son cualitativas las siguientes variables: Nivel de educación del jefe de la unidad familiar, existen niños por debajo de 6 años en la unidad familiar y existen personas entre 6 y 18 años en la unidad familiar.

Si las alternativas de una variable cualitativa se pueden verificar se considera que es un hecho y si no se pueden verificar se considera opinión o motivo.

La opinión y el motivo son subjetivos. La diferencia está en que la opinión es un juicio acerca de un hecho y el motivo es la razón o causa de un hecho o una opinión. En el caso, todas las variables cualitativas son hechos.

## J. MEDIR.

Es determinar cuántas veces cierta unidad de medida, llamada patrón de medida, cabe en cada característica del elemento a medir.

Medir el gasto semanal por comida de cada socio, es averiguar cuántas veces la unidad de medida peso es gastada en una semana.

## K. ERROR DE MUESTREO.

Es la diferencia entre la medida tomada de una muestra y la medida tomada de una población o universo, debida al hecho de haber sido investigada una de las tantas muestras que se pueden tomar de una población dada.

Para ver si éste error se dió, en una investigación, habría que comparar la medida de la muestra (estadígrafo) con la medida del universo (parámetro).

Como es un error que se estima cometer, éste lo da el investigador a su juicio. Se acostumbra identificarlo con la letra e o d.

#### L. ERROR NO MUESTRAL.

Es la diferencia entre la medida tomada de una muestra y la medida tomada de una población o universo, debido a fallas en el proceso de investigación: Compilación de la información, manipulación de la información, instrumento usado para registrar la información, etc.

Para ver si éste error se dió, en una investigación, habría que hacerle una investigación al proceso de ésta.

Este es un error que se dá si hay fallas en el proceso de investigación, de lo contrario no se dá.

#### LL. CONFIANZA.

Es el porcentaje de veces que se espera sucederá lo que se afirma. También se le llama margen de confiabilidad. Se identifica con la letra Z.

Como es algo que se espera que se de, la cifra la da el investigador a su juicio.

Para su uso en el tamaño de muestra, inferencia, etc., la cifra dada se debe convertir a desviaciones estandar por medio de una tabla.

Este proceso se verá en unidades posteriores.

#### M. ESTADIGRAFO.

Es la medida de una característica de los elementos de una muestra. Por ejemplo, si de una muestra de 30 socios se encuentra que el gasto semanal por comida promedio es de \$ 9.000, esto es un estadígrafo.

Para indicar un estadígrafo, se acostumbra usar letras minúsculas.

#### N. PARAMETRO.

Es la medida de una característica de los elementos de un universo. Por ejemplo, si se averigua el gasto semanal por comida promedio a los 160 socios y éste da \$ 10.500, ésto es un parámetro.

Para indicar un parámetro, se acostumbra usar letras mayúsculas o letras griegas.

#### Ñ. INFERIR.

Es concluir cuál puede ser la característica de un universo, con base en lo que se halló en una muestra acerca de la misma característica. Por ejemplo, si el estadígrafo (de una muestra de 30 socios) del gasto semanal por comida promedio es de \$ 9.000, inferir consistiría en determinar si el gasto semanal por comida promedio en los 160 socios es igual a \$ 9.000, menor o mayor.

Para indicar que se está haciendo una inferencia, se acostumbra usar letras minúsculas con un triángulo sin base encima de estas.

#### O. EJERCICIO DE AUTO EVALUACION.

En un texto de investigación de mercadeo, ubique un caso que haga referencia a una investigación por hacerse o ya realizada, lealo hasta entenderlo e identifique las partes de éste que hacen alusión a cada uno de los terminos vistos en esta unidad.

## UNIDAD 2

## EL MUESTREO DENTRO DE UN PROCESO DE INVESTIGACION FORMAL

## A. OBJETIVOS ESPECIFICOS.

Al finalizar esta unidad, el estudiante estará en capacidad de:

1. Indicar y aplicar la relación entre decisión, información e investigación.
2. Señalar y aplicar lo que es muestreo dentro de un proceso de investigación formal.

## B. RELACION ENTRE DECISION, INFORMACION E INVESTIGACION.

La toma de decisiones es algo que hace parte de la vida cotidiana de todas las personas.

Cuando se tiene información a la mano (en la mente o en algún documento), la decisión se puede tomar inmediatamente.

Cuando no se tiene la información necesaria, hay que averiguarla através de un proceso de investigación formal o informal (ver gráfico # 1).

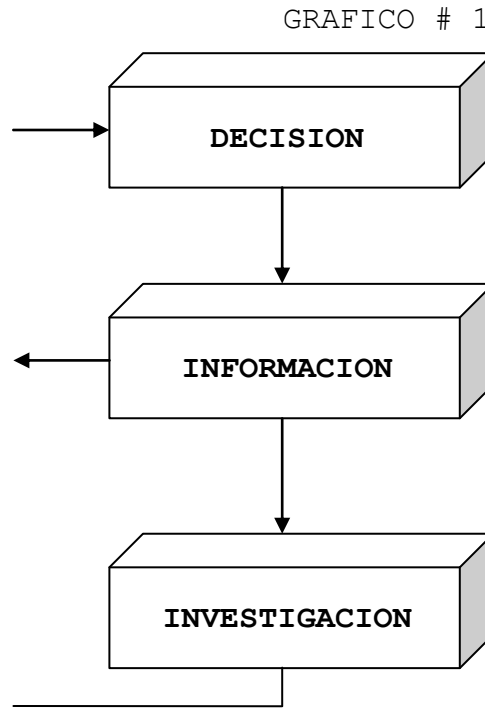
En el caso de la CMA la decisión que esta pendiente de tomarse, por parte de la gerente, es acerca de: "Planificar mejor la mezcla y cantidad de productos que la CMA ofrecia".

La gerente considera que para tomar ésta decisión, necesita información acerca de: "Los habitos de compra y características demográficas de los socios".

Como la anterior información no esta compilada o acopiada, se hace necesario averiguarla através de un proceso de investigación.

Al entrar a planear y desarrollar una investigación acerca de: "Habitos y características demográficas ...", es necesario DESMENUZAR los términos anteriores

en variables, para que se puedan medir a nivel de cada socio.



RELACION ENTRE DECISION, INFORMACION E INVESTIGACION

Una línea vertical descendente conecta este texto con el diagrama anterior.

Fuente: El autor.

Al entrar a desmenuzar, se han detectado 7 variables que se ajustan a los términos: "Habitos y



características demográficas ...", ellas son:

- Gasto semanal por comida de la familia.
- Ingreso anual de la familia.
- Número de personas de la familia.
- Educación del jefe de la familia.
- Edad del jefe de la familia.
- Si hay niños menores de 6 años en la familia.
- Si hay personas entre 6 y 18 años en la familia.

La primera variable se considera un hábito y las 6 restantes son características demográficas.

#### C. PARTES BASICAS DE UNA INVESTIGACION FORMAL.

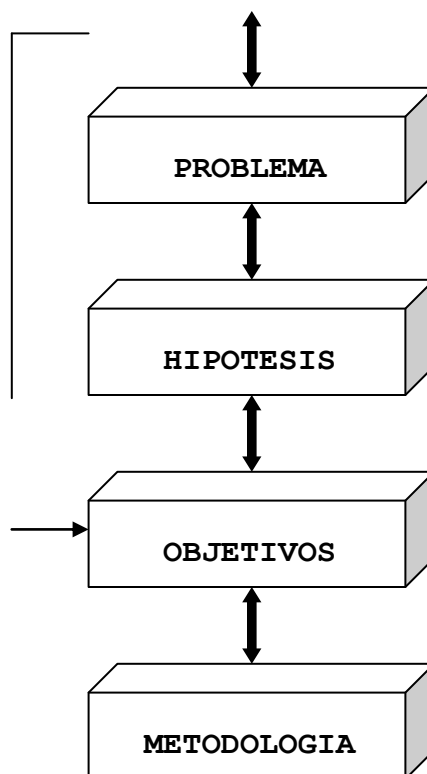
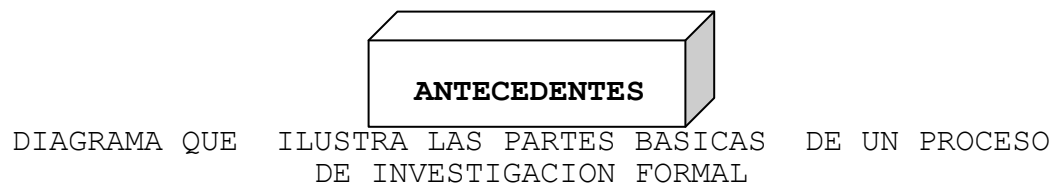
Una vez se han identificado las variables que es necesario medir a través de un proceso de investigación, es necesario entrar a planear éste proceso.

Las partes básicas de un proceso de investigación formal, son en su orden: Antecedentes, problema, hipótesis, objetivo(s) general(es), objetivo(s) específico(s) y metodología (ver gráfico # 2).

Las partes anteriores tienen una interrelación que consiste en lo siguiente:

- Los antecedentes son una narración de hechos que indican la procedencia del problema y dan pistas para hipótesis.
- El problema se define con base en los antecedentes.
- Las hipótesis son supuestos acerca del problema y las ideas se extraen de los antecedentes.
- Los objetivos generales indican la utilidad de investigar el problema.
- Los objetivos específicos contienen las variables que es necesario medir, para probar o refutar las hipótesis.

GRAFICO # 2



Fuente: El autor.

- La metodología contiene la forma como se va a proceder para medir las variables contenidas en los objetivos específicos.

Dentro de la metodología se identifican tres partes básicas: Las características del modelo de investigación, los alcances de la investigación y las fuentes de información a consultar.

Cuando se decide usar el muestreo, en el proceso de investigación, en la parte de fuentes de información deben aparecer aspectos tales como estos:

- Universo.
- Tamaño de muestra.
- Método de elección de la muestra.
- Elemento muestral.
- Trabajo de campo.
- Supervisión.
- Tabulación.

Todos los aspectos anteriores son de dominio del muestreo y cada uno tiene sus técnicas respectivas.

Como se puede observar, el muestreo esta incrustado dentro de la parte de la metodología de un proceso de investigación.

#### D. EJERCICIO DE AUTOEVALUACION.

Tome el caso usado para el ejercicio de autoevaluación de la unidad 1 y realice las siguientes actividades:

- Indique la decisión pendiente de tomar.
- Describa la información que se necesita conocer.
- Enumere las variables que se desprenden de la información que se necesita.
- Plantee el problema y las hipótesis (si las hay).
- Redacte el o los objetivos generales y específicos.
- Redacte la metodología que se ajuste al caso y

específicamente lo relativo a fuentes de información.

### UNIDAD 3

#### MUESTREO ALEATORIO SIMPLE

##### A. OBJETIVOS ESPECIFICOS.

Al finalizar esta unidad, el estudiante estará en capacidad de:

1. Definir muestreo aleatorio simple.
2. Calcular un tamaño de muestra.
3. Elegir un tamaño de muestra.
4. Diseñar el formato para registrar la información que se va a compilar.
5. Elaborar los cuadros de tabulación y estadísticos que salen de una muestra elegida.
6. Hacer inferencias desde el punto de vista de la estimación y la prueba de hipótesis.

##### B. DEFINICION.

El muestreo aleatorio simple es una técnica donde todos los elementos de un universo tienen la misma posibilidad o probabilidad de ser elegidos y hay homogeneidad entre los elementos.

El término aleatorio o al azar, se refiere a que todos

los elementos tienen la misma posibilidad o probabilidad.

El término simple, se refiere a que hay homogeneidad entre los elementos.

### C. CALCULO DEL TAMAÑO DE LA MUESTRA.

Todas las variables cuantitativas y cualitativas dicótomas (que tienen dos alternativas) intervienen o participan en el cálculo del tamaño de muestra.

Las variables cualitativas no dicótomas, no tienen fórmulas o ecuaciones para el cálculo del tamaño de la muestra.

Se deben calcular tantos tamaños de muestra como variables cuantitativas y cualitativas dicótomas haya.

El tamaño de muestra definitivo será el mayor de todos los calculados.

Las variables cualitativas no dicótomas, al no tener fórmulas o ecuaciones, estarán REPRESENTADAS por las otras dos clases de variables.

1. Fórmulas o ecuaciones para el cálculo del tamaño de muestra de las variables cuantitativas.

a. Cuando la población es infinita:

$$n = \frac{(Z^2 \cdot s^2)}{d} \cdot \left(1 + \frac{2}{n}\right)$$

b. Cuando la población es finita:

1). Usando parámetros:

$$n_o = \frac{(Z^2 \cdot S^2)}{d} \cdot n = n_o / \left(1 + \frac{n_o}{N}\right)$$

2). Usando estadígrafos.

$$n_o = \frac{(Z^2 \cdot s^2)}{d} \cdot \left(1 + \frac{2}{n}\right) \cdot n = n_o / \left(1 + \frac{n_o}{N}\right)$$

2. Formulas o ecuaciones para el calculo del tamaño de muestra de las variables cualitativas dicótomas.

a. Cuando la población es infinita:

$$n = \frac{(Z^2 \cdot p \cdot q)}{d} \cdot \left(1 + \frac{2}{n}\right)$$

b. Cuando la población es finita:

1). Usando parámetros.

$$n_o = \frac{Z^2 \cdot P \cdot Q}{d} \quad n = \frac{n_o}{1 + (n_o / N)}$$

2). Usando estadígrafos.

$$n_o = \frac{Z^2 \cdot p \cdot q}{d} \cdot (1 + \frac{2}{n}) \quad n = \frac{n_o}{1 + (n_o / N)}$$

3. Significado de las letras y signos usados en las formulas o ecuaciones.

$Z$  = Márgen de confiabilidad. Es el porcentaje del área de la distribución que se esta teniendo en cuenta. Este porcentaje lo da el investigador y se tiene que traducir a desviaciones estandar, por intermedio de la tabla de la distribución de la normal.

$s$  = Varianza estadígrafo. Se halla con base en una premuestra o se toma de una investigación anterior.

$S$  = Varianza parámetro. Se halla del universo total.

$d$  = Error de muestreo. Lo da el investigador y el intervalo común para escogerlo esta entre 0% y 10%. Para las variables cualitativas dicótomas se eleva el porcentaje al cuadrado. Ej: 10% de error sería: (0.10) al cuadrado. Para las variables cuantitativas se multiplica el porcentaje por la media de la variable y el resultado se eleva al cuadrado. Ej: 10% de error sería: (media \* 0.10) al cuadrado.

$n$  = Premuestra o muestra piloto o muestra preliminar su tamaño lo fija el investigador.

$n_o$  = Muestra no ajustada. Hay que ajustarla usando el dato del universo.

$n$  = Muestra definitiva de cada variable.

$N$  = Universo o población.

$p$  y  $P$  = Porcentaje de elementos que tienen determinada característica (si). El primero es estadígrafo y el segundo es parámetro.

$q$  y  $Q$  = Porcentaje de elementos que no tienen determinada característica (no). El primero es estadígrafo y el segundo es parámetro.

4. Pasos para el calculo del tamaño de una muestra.  
En forma secuencial, los pasos para calcular el tamaño de una muestra son:

a. Se listan las variables que van a participar en el calculo del tamaño de la muestra. En el caso de la CMA, éstas son:

- Gasto semanal por comida.
- Ingreso anual de la familia.
- Número de personas de la familia.
- Edad del jefe de la unidad familiar.
- Si hay niños menores de 6 años en la familia.
- Si hay personas entre 6 y 18 años en la familia.

Las 4 primeras son cuantitativas y las 2 últimas son cualitativas dicótomas. La educación del jefe de la unidad familiar es una variable cualitativa no dicótoma y no entra en el calculo del tamaño de la muestra.

b. Se elige una premuestra al azar y se anota al frente de cada elemento el valor de cada variable. Usando los números contenidos en el anexo # 2 (tabla de números aleatorios) y con el siguiente criterio: página 411, columna 2, fila 51, en forma vertical y los tres primeros números iguales o menores de 160, se escogen 10 elementos:

1	2	3	4	5	6	7	8
009	7.500	2	1.250.000	2	51	no	si
124	7.800	2	2.100.000	2	26	no	no
080	11.000	3	1.500.000	1	19	no	no
052	6.000	3	1.250.000	1	36	si	no
012	10.000	2	4.000.000	1	35	no	no
064	5.500	2	250.000	1	26	no	no
143	18.000	2	1.400.000	2	59	no	si
156	20.000	5	3.000.000	3	61	si	si
017	21.000	2	1.500.000	3	19	no	no
041	15.000	1	3.000.000	2	34	no	no

El criterio depende del investigador, los encabezados de las columnas tienen el mismo significado del anexo # 1 y los códigos de los elementos no se deben repetir.

c. Se calculan los estadígrafos para cada variable, que se requieren en la respectiva fórmula para calcular la muestra:

- Gasto semanal por comida. Media = \$ 12.180. Varianza = \$ 34.535.111,11.

- Número de personas de la familia. Media = 2,4 personas. Varianza = 1,16 personas.

- Ingreso anual de la familia. Media = \$ 1.925.000. Varianza = \$ 1.222.361.111.000.

- Edad del jefe de la unidad familiar. Media = 36,6 años. Varianza = 239,82 años.

- Si hay niños menores de 6 años en la familia. Si hay (p) = 20%. No hay (q) = 80%.

- Si hay personas entre 6 y 18 años en la familia. Si hay (p) = 30%. No hay (q) = 70%.

d. Se determina el margen de error de muestreo y el margen de confiabilidad a usar en todos los cálculos. Como el margen de error de muestreo lo determina el investigador, para este caso se determinó en un 20%.

El margen de confiabilidad se determinó en 90%, pero como hay que convertirlo en desviaciones estándar con el anexo # 3, el valor resultante es 1.65.

Para hallar el 1.65 se hace lo siguiente: Se divide 90% entre 2 y luego entre 100, eso da 0.4500. Se busca ese valor en el cuerpo del anexo # 3 y se observa el valor del encabezado de la fila (1.6) y el valor del encabezado de la columna (0.05). Se suman los dos valores anteriores y da 1.65.

e. Se halla el tamaño de muestra para cada variable. En este caso hay que hallar 6 tamaños de muestra:

- Gasto semanal por comida.

$$no = \frac{(Z^2 * s^2)}{d^2 * (1 + 2/n)}$$

$$no = \frac{(1.65^2 * 34.535.111.11)}{(12.180^2 * 0.20)}$$

$$* (1 + 2/10). \quad no = 19$$

$$n = no / (1 + no/N). \quad n = 19 / (1 + 19/160) = 17$$



- Número de personas de la familia.

$$no = \frac{(Z^2 * s^2) / d^2 * (1 + 2/n)}{1}$$

$$no = \frac{(1.65^2 * 1.16^2) / (2.4^2 * 0.20^2) * (1 + 2/10)}{1} \cdot no = 16$$

$$n = no / (1 + no / N) \cdot n = 16 / (1 + 16 / 160) = 14$$

- Ingreso anual de la familia.

$$no = \frac{(Z^2 * s^2) / d^2 * (1 + 2/n)}{1}$$

$$no = \frac{(1.65^2 * 1.222.361.111.000^2) / (1.925.000^2 * 0.20^2) * (1 + 2/10)}{1} \cdot no = 22$$

$$n = no / (1 + no/N) \cdot n = 22 / (1 + 22/160) = 18$$

- Edad del jefe de la unidad familiar.

$$no = \frac{(Z^2 * s^2) / d^2 * (1 + 2/n)}{1}$$

$$no = \frac{(1.65^2 * 239.82^2) / (36.6^2 * 0.20^2) * (1 + 2/10)}{1} \cdot no = 12$$

$$n = no / (1 + no/N) \cdot n = 12 / (1 + 12/160) = 11$$

- Hay niños menores de 6 años en la familia.

$$no = \frac{(Z^2 * p * q) / d^2 * (1 + 2/n)}{1}$$

$$no = \frac{(1.65^2 * 0.20 * 0.80) / 0.20^2 * (1 + 2/10)}{1} \cdot no = 11$$

$$n = no / (1 + no/N) \cdot n = 11 / (1 + 11/160) = 10$$

- Hay personas entre 6 y 18 años en la familia.

$$no = \frac{(Z^2 * p * q) / d^2 * (1 + 2/n)}{1}$$

$$no = \frac{(1.65^2 * 0.30 * 0.70) / 0.20^2 * (1 + 2/10)}{1} \cdot no = 14$$

$$n = no / (1 + no/N) \cdot n = 14 / (1 + 14/160) = 13$$

f. Se determina el tamaño de la muestra definitivo. El tamaño de muestra a usar en la investigación será de 18 elementos ( n = 18). Esta cifra es suficiente para la variable que la generó y más que suficiente para

las variables que dieron tamaños de muestra menores.

#### D. ELECCION DEL TAMAÑO DE LA MUESTRA.

Para elegir los 18 elementos de la muestra entre los 160 elementos del universo, se debe usar el anexo # 2 y definir un criterio. El criterio para el caso fué el siguiente: Página 412, columna 1, fila 101, en forma horizontal y los tres primeros números de cada columna. Los códigos de los 18 elementos y los valores de cada una de las variables son:

1	2	3	4	5	6	7	8
132	19.000	2	3.000.000	2	30	no	no
007	7.500	1	1.750.000	2	76	no	no
021	16.500	3	6.500.000	5	23	si	no
072	30.000	2	5.000.000	4	22	no	no
003	9.000	1	1.000.000	1	20	no	no
038	3.500	1	1.250.000	1	19	no	no
001	6.000	1	1.250.000	1	56	no	no
112	10.000	1	3.000.000	2	64	no	no
110	9.000	1	2.100.000	2	59	no	no
064	5.500	2	250.000	1	26	no	no
079	10.000	3	4.500.000	2	58	si	no
019	15.000	2	3.000.000	5	41	no	no
142	9.000	4	2.500.000	2	40	no	no
120	28.000	2	6.500.000	2	29	no	no
025	26.500	3	3.250.000	2	51	no	si
019	15.000	2	3.000.000	5	41	no	no
074	3.750	2	1.200.000	1	51	no	no
124	7.800	2	2.100.000	2	26	no	no

No se debe repetir ningun elemento, pues sería como entrevistar dos veces o más a la misma persona y sobre lo mismo.

#### E. DISEÑO DEL FORMATO PARA REGISTRAR LA INFORMACION.

1. Partes básicas. Un formato para registrar información tiene tres grandes partes:

a. Información de identificación.

Esta parte contiene datos sobre el entrevistado, el entrevistador, el supervisor y el estado de la encuesta.

b. Información de clasificación.

Esta parte contiene datos sobre la o las variables que

se usan para estratificar el universo. Solo es necesaria cuando se esta usando el muestreo estratificado. Contra esta o estas variables se crusan, al tabular, las variables contenidas en la parte de información básica.

c. Información básica.

Esta parte contiene las variables que se van a medir y que son de interes para la investigación.

Las variables estan implícitas en las preguntas que contenga el formato. Cada pregunta debe contener una variable.

En el caso de la CMA, son 7 las variables que interesa medir y por lo tanto se necesitarán 7 preguntas.

Las partes pueden ir en distinto orden y la parte de información de identificación puede ir o no, según la investigación. En algunos casos la parte de identificación inhibe la colaboración de los entrevistados.

2. Pasos para la estructura del formato. En su orden, los pasos pueden ser:

- Elaborar una lista de las variables que se van a medir.
  - Ordenar las variables de lo general a lo particular.
  - Determinar la información de identificación necesaria para la investigación.
  - Determinar las variables que irían en la parte de información de clasificación, si se esta usando el muestreo estratificado.
  - Confeccionar el formato con las tres partes básicas y en el orden deseado.
- Para el caso de la CMA, el formato puede ser el siguiente:

INFORMACION DE IDENTIFICACION

Número de la encuesta \_ \_ \_

Fecha en que se lleno la encuesta \_\_\_\_\_

Nombre del entrevistado \_\_\_\_\_



- |       |      |
|-------|------|
| 1. Si | fin. |
| 2. No | fin. |

---

Este pudo haber sido el diseño de la encuesta que se usó para compilar la información, que esta contenida en la matriz del anexo # 1 (160 filas y 8 columnas).

#### F. ELABORACION DE LOS CUADROS DE TABULACION Y LOS CUADROS ESTADISTICOS.

Como mínimo debe salir un cuadro de tabulación y un cuadro estadístico por cada variable. Como en el caso de la CMA son 7 variables, saldrán 7 cuadros de tabulación y 7 cuadros estadísticos. Los cuadros de tabulación son el borrador de los cuadros estadísticos.

1. Cuadros de tabulación. Cuando la tabulación se hace en forma manual, cada cuadro de tabulación permite resumir lo que se encontró de cada variable, en todos los elementos de la muestra.

En el caso de la CMA, el cuadro de tabulación que saldría de la variable: "gasto semanal por comida de cada socio", sería el siguiente:

ALTERNATIVAS.	TOTAL.
0        5.000	///
5.000   10.000	/////
10.000   15.000	//
15.000   20.000	///
20.000   25.000	
25.000   30.000	//
30.000   35.000	/

NOTA: Los intervalos se deben interpretar así: \$ 0 -

5.000, quiere decir de \$ 0 a 5.000 inclusive. \$ 5.000 - 10.000, quiere decir de más de \$ 5.000 a \$ 10.000 inclusive. Lo mismo para los demas. El número de personas ideal para tabular manualmente son dos: uno lee y el otro anota las respuestas usando "palos".

2. Cuadros estadísticos. Las partes de un cuadro estadístico, de arriba hacia abajo, son las siguientes: número, título, cuerpo, número de entrevistados, notas(opcional) y fuente. El cuadro estadístico que sale del cuadro de tabulación anterior, es el siguiente:

#### NUMERO 1

#### GASTO SEMANAL POR COMIDA DE LOS SOCIOS DE LA CMA

ALTERNATIVAS.		TOTAL.	
\$		#	%
0	5.000	3	16.6
5.000	10.000	6	33.4
10.000	15.000	2	11.1
15.000	20.000	4	22.3
20.000	25.000	-	-
25.000	30.000	2	11.1
30.000	35.000	1	5.5
Total respuestas		18	100.0
Número de entrevistados		(18)	

Fuente: Este estudio.

#### G. INFERENCIA.

Hay dos clases de inferencia: la estimación y la prueba de hipótesis.

1. Estimación. Hay dos clases: la estimación puntual y la estimación de intervalo.

La estimación puntual hace el parámetro igual al estadígrafo. La estimación de intervalo ubica el parámetro dentro de un límite inferior y un límite superior, calculados con base en el estadígrafo.

a. Formulas o ecuaciones para la estimación de intervalo.

1). Cuando la variable es cuantitativa.

a). Cuando la población es infinita.

- Si  $n$  es mayor que 30.

$$\hat{X} = \bar{x} - Z * s / \text{raiz}(n)$$

- Si  $n$  es menor o igual a 30.

$$\hat{X} = \bar{x} - t * s / \text{raiz}(n)$$

b). Cuando la población es finita.

- Si  $n$  es mayor que 30.

$$\hat{X} = \bar{x} - Z * s / \text{raiz}(n) * \text{raiz}(1-f)$$

- Si  $n$  es menor o igual a 30.

$$\hat{X} = \bar{x} - t * s / \text{raiz}(n) * \text{raiz}(1-f)$$

2). Cuando la variable es cualitativa dicótoma.

a). Cuando la población es infinita.

- Si  $n$  es mayor que 30.

$$\hat{p} = p - Z * \text{raiz}[(p * q) / n]$$

- Si  $n$  es menor o igual a 30.

$$\hat{p} = p - t * \text{raiz}[(p * q)/(n-1)]$$

b). Cuando la población es finita.

- Si n es mayor que 30.

$$\hat{p} = p - Z * \text{raiz}[(p * q)/n] * \text{raiz}(1-f)$$

- Si n es menor o igual a 30.

$$\hat{p} = p - t * \text{raiz}[(p * q)/(n-1)] * \text{raiz}(1-f)$$

b. Significado de las letras usadas en las formulas o ecuaciones.

$\hat{X}$   
-  
X = Posible parámetro de la media.

-  
x = Media aritmetica(estadígrafo).

Z = Distribución Z. Significa el porcentaje de las posibles muestras, en que el parámetro cae en el intervalo. Lo determina el investigador y se busca en la tabla anexo # 3.

s = Desviación estandar (estadígrafo).

t = Distribución t. Significa lo mismo que Z. Lo determina el investigador y se busca en la tabla anexo # 4.

f = Fracción de muestreo.  $f = n/N$ . Se usa cuando la división da mayor o igual a 5%.

$\hat{p}$   
p = Posible parámetro de p.

c. Pasos para una estimación de intervalo. Siguiendo un orden, los pasos a seguir son los siguientes:

- Se calculan los estadígrafos de la variable. En el caso del cuadro # 1 estos serían: media = 13.056.6 y s = 9.056.8.

- Se escoje la formula o ecuación adecuada. En el caso del cuadro # 1 ésta sería:

$\hat{p}$



$$X = \bar{x} - t * s / \sqrt{n} * \sqrt{1-f}.$$

- Se determina el valor de t y se busca en la tabla. Si se indica que t sea 90%, hay que trabajar con el complemento o sea 10%. En la tabla anexa # 4, se debe buscar el valor de t, dividiendo el 10% entre 2 (por lo de las 2 colas) y hallando los grados de libertad (n-1). En la tabla habría que buscar el valor correspondiente a 0.05 y 17, esto da 1.74 para sustituir en la fórmula. Si el tamaño de muestra

hubiera sido mayor de 30, se debe buscar en la tabla anexo # 3. Habría que dividir 90% entre 2 (por lo de las 2 colas) y luego entre 100%. En la tabla de la Z habría que buscar el valor correspondiente a 0.45, lo cual da 1.65 aproximadamente.

- Se aplica la fórmula o ecuación respectiva, para hallar el posible parámetro.

$$X = 13.055.6 - 1.74 * 9.056.8 / \sqrt{18} * \sqrt{1-0.113}.$$

$$X = 13.055.6 - 3.498.2$$

$$X = 9.557.4 - 16.553.8$$

- Se interpreta el intervalo y se aconseja acerca de la decisión a tomar. La interpretación sería: en el 90% de las posibles muestras, el gasto semanal por comida de los socios de la cooperativa está entre \$ 9.557.4 y \$ 16.553.8. Si el aumento de la cantidad y variedad de productos, estuviera supeditado a que el gasto fuera superior a \$ 10.000 semanales. Entonces, se aconsejaría que se ampliara la mezcla de productos, pues hay más probabilidad de que el gasto sea superior a \$ 10.000 semanales.

2. Prueba de hipótesis. Consiste en determinar si el parámetro será igual o no, mayor o igual, o menor o igual a determinado valor, teniendo como base el estadígrafo.

La prueba de hipótesis dicotomiza las situaciones, mientras que la estimación de intervalos da varios valores para el parámetro. La prueba de hipótesis

se basa en el planteamiento de dos hipótesis: Una nula( $H_0$ ) y la otra alterna( $H_1$ ).

El contraste o prueba de hipótesis se hace alrededor de la hipótesis nula. Siempre hay que comparar un valor teórico o de tabla, contra un valor calculado con base en el estadígrafo y el posible parámetro.

Siempre se debe concluir si se acepta o se rechaza la hipótesis nula.

a. Cuando la variable es cuantitativa.

1). Formas de las hipótesis. Hay tres alternativas de hipótesis nula y alterna para escoger.

-  $H_0: U_0 = ?$

$H_1: U_0$  **no igual** ?

-  $H_0: U_0 > o$  igual a ?

$H_1: U_0 < ?$

-  $H_0: U_0 < o$  igual a ?

$H_1: U_0 > ?$

La primera alternativa se llama de dos colas, por que hay dos zonas de rechazo de la hipótesis nula. La segunda y tercera alternativas se llaman de una cola, porque solo hay una zona de rechazo de la hipótesis nula.

2). Procedimiento para hallar el valor calculado y el valor teórico, cuando  $n$  es mayor que 30. Cuando  $n$  es mayor que 30 se usa la distribución Z y la formula para hallar el valor calculado es la siguiente:

$$Z_c = \frac{\bar{x} - U_0}{s/\mathbf{raiz}(n)}$$

El Z de tabla( $Z_t$ ) se halla de dos formas, dependiendo del número de colas. Cuando la forma de las hipótesis es de una cola, el  $Z_t$  se halla así:

A 0.50 se le resta el valor dado al error tipo I o alfa y ese valor se busca en la tabla de la distribución Z(anexo # 3). Cuando la forma de las hipótesis es de dos colas, el  $Z_t$  se halla así: A 1 se le resta el valor dado al error tipo I o alfa y ese valor se busca en la tabla de la distribución Z.

3). Procedimiento para hallar el valor calculado y el valor teórico, cuando  $n$  es menor o igual a 30.

Cuando  $n$  es menor o igual a 30 se usa la distribución T y la formula para hallar el valor calculado es la siguiente:

$$T_c = (\bar{x} - U_0) / s / \mathbf{raiz}(n).$$

El T de tabla ( $T_t$ ) se halla de dos formas, dependiendo del número de colas. Cuando la forma de las hipótesis es de una cola, el  $T_t$  se halla así: Se fija un error tipo I o alfa y los grados de libertad ( $n - 1$ ), y con esos dos valores se halla el valor de  $T_t$  en la tabla de la distribución T (anexo # 4). Cuando la forma de las hipótesis es de dos colas, el  $T_t$  se halla así: Se fija un error tipo I o alfa y se divide por dos, se determina los grados de libertad ( $n - 1$ ) y con esos dos valores se halla el  $T_t$  en la tabla de la distribución T (anexo # 4).

b. Cuando la variable es cualitativa dicótoma.

1). Formas de las hipótesis. Hay tres alternativas de hipótesis nula y alterna para escoger.

-  $H_0: P_0 = ?$

$H_1: P_0$  **no igual** ?

-  $H_0: P_0 > \text{o igual a } ?$

$H_1: P_0 < ?$

-  $H_0: P_0 < \text{o igual a } ?$

$H_1: P_0 > ?$

La primera alternativa es de dos colas y la segunda y tercera son de una cola.

2). Procedimiento para hallar el valor calculado y el valor teórico. Solo se trabaja con la distribución Z. Pero, se debe cumplir con las siguientes condiciones:  $n * P_0 > \text{o igual a } 5$  y  $n * (1 - P_0) > \text{o igual a } 5$ . Si no se cumplen estas dos condiciones, es necesario aumentar el tamaño de la muestra hasta cumplirlas. La formula para hallar el  $Z_c$  es la siguiente:

$$Z_c = p * P_0 / \mathbf{raiz}\{[P_0 * (1 - P_0)]/n\}$$

El  $Z_t$  se calcula en la misma forma que para las variables cuantitativas.

c. significado de las letras usadas en las formulas o ecuaciones.

$H_0$  = Hipótesis nula.

$H_1$  = Hipótesis alterna.

$\mu_0$  (miu subcero) = Parámetro de la media.

$p_0$  = Parámetro de  $p$ . También se usa la letra  $P$ .

$Z_c$  =  $Z$  calculado.

$Z_t$  =  $Z$  de tabla. El signo es el mismo de  $Z_c$ .

$T_c$  =  $T$  calculado. El signo es el mismo de  $T_c$ .

$T_t$  =  $T$  de tabla.

Error tipo I o  $\alpha$  = Probabilidad de rechazar  $H_0$  siendo verdadera. Su valor lo determina el investigador, teniendo en cuenta el costo de cometer éste error o el error B(beta) o tipo II.

El error B es la probabilidad de aceptar  $H_0$  siendo falsa. Si es más costoso cometer el error tipo I, la probabilidad (porcentaje) debe ser baja. En éste caso, se incitaría que se presentará el error tipo II.

Si es más costoso cometer el error tipo II, la probabilidad del error tipo I debe ser alta. En éste caso, se incitaría que se presentará el error tipo I.

G.L. = Grados de libertad. Son iguales a  $n - 1$ .

d. Zonas de aceptación de la hipótesis nula. En un contraste o prueba de hipótesis de una cola, la zona de aceptación de  $H_0$  estará a la derecha del  $Z_t$  o  $T_t$ , si  $H_0$  indica  $>$  o igual que  $?$ . En un contraste o prueba de hipótesis de una cola, la zona de aceptación de  $H_0$  estará a la izquierda de  $Z_t$  o  $T_t$ , si  $H_0$  indica  $<$  o igual que  $?$ . En un contraste o prueba de hipótesis de dos colas, la zona de aceptación de  $H_0$  estará entre la derecha de menos  $Z_t$  o  $T_t$  y entre la izquierda de más  $Z_t$  o  $T_t$ . El  $Z_t$  o  $T_t$  tienen o asumen el signo que tenga el  $Z_c$  o  $T_c$ . Para determinar la aceptación o rechazo de  $H_0$  es mejor graficar los valores de  $Z_c$  y  $Z_t$  (o  $T_c$  y  $T_t$ ) en la curva que representa la respectiva distribución.

e. Pasos para una prueba o contraste de hipótesis.

En su orden, los pasos son los siguientes:

- Se indica una decisión a tomar, contemplando un posible parámetro. En el caso, una de las decisiones a tomar es: "si el gasto semanal por comida promedio es por lo menos de \$ 10.000, se aumentará la cantidad y variedad de productos de la cooperativa".

- Se plantean las hipótesis nula y alterna, de la forma que mejor se ajuste a la decisión. En éste caso la forma es:

$H_0: U_0 \geq 10.000$  se aumentará...

$H_1: U_0 < 10.000$  no se aumentará ...

- Se toma una muestra y se halla el estadígrafo de la respectiva variable. En éste caso, el estadígrafo del gasto semanal por comida promedio actual es de \$ 13.055.6.

- Se halla el valor del  $T_c$  o  $Z_c$ . En éste caso, sería el  $T_c$  porque el tamaño de muestra es 18.

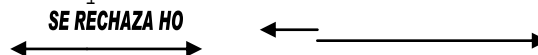
$$T_c = (\bar{x} - U_0) / s / \sqrt{n}$$

$$T_c = (13.055.6 - 10.000) / 9.056.8 / \sqrt{18} \quad T_c = + 1.43$$

- Se halla el valor de  $T_t$  o  $Z_t$ . En éste caso, sería el  $T_t$ . Como es más costoso aceptar  $H_0$  siendo falsa (error tipo II), se debe asignar un alfa alto. puede ser 10%. Los grados de libertad (G.L.) son:  $18 - 1 = 17$ .

Hay que buscar en la tabla de la distribución T, estos dos valores: 17 y 10%. Hay que aclarar que la prueba o el contraste es de una cola.

El valor de  $T_t$  es + 1.33. No hay que olvidar que el signo es el mismo que tiene el  $T_c$ .



- Se grafica  $T_c$  y  $T_t$ , y se indica la zona de aceptación y rechazo de  $H_0$ .

Se ACEPTA  $H_0$

| |  
| |

| |

0 1.33

La zona de aceptación y rechazo de  $H_0$  lo indica el  $T_t$ , y la decisión de si acepta o rechaza  $H_0$  lo indica el  $T_c$ . En éste caso, se acepta  $H_0$ . Es decir, que el estadígrafo de \$ 13.055.6 de gasto semanal por comida promedio, da para pensar que el parámetro será igual o mayor a \$ 10.000.

- Se concluye acerca de la decisión a tomar. Según el planteamiento de las hipótesis, la cantidad y variedad de productos se aumentará si el parámetro del gasto semanal por comida promedio es igual o superior a \$ 10.000. Como el estadígrafo da para pensar que el parámetro es igual o superior a \$ 10.000, la decisión a tomar sería aumentar la cantidad y variedad de productos de la cooperativa.

#### H. EJERCICIO DE AUTOEVALUACION.

Tome el caso usado en las unidades 1 y 2, y realice las siguientes actividades (según el muestreo aleatorio simple):

- Calcule el tamaño de muestra.
- Elija la muestra calculada.
- Diseñe el formato a usar para registrar la información que se va a compilar.
- Aplique el formato diseñado a la muestra elejida.
- Elabore los cuadros de tabulación y estadísticos que salgan del formato.
- Tome una variable y haga la inferencia desde el punto de vista de la estimación de intervalos y desde el punto de vista de la prueba de hipótesis.

## UNIDAD 4

## MUESTREO ALEATORIO ESTRATIFICADO

## A. OBJETIVOS ESPECIFICOS.

Al finalizar esta unidad, el estudiante estará en capacidad de:

1. Definir muestreo aleatorio estratificado.
2. Calcular un tamaño de muestra.
3. Elegir un tamaño de muestra.
4. Diseñar el formato para registrar la información.
5. Elaborar los cuadros de tabulación y estadísticos que salen de una muestra elegida.
6. Hacer inferencias desde el punto de vista de la estimación y la prueba de hipótesis.

## B. DEFINICION.

El muestreo aleatorio estratificado, es una técnica de muestreo donde todos los elementos, de un universo, tienen la misma posibilidad de ser elegidos y hay heterogeneidad entre los elementos con relación a las variables que se van a medir.

El término aleatorio o al azar, se refiere a que todos

los elementos de cada grupo tienen la misma posibilidad o probabilidad.

El término estratificado, se refiere a que el universo hay que dividirlo en grupos, porque hay heterogeneidad entre los elementos con relación a las variables a medir.

Los grupos son homogéneos entre si, pero heterogéneos de un grupo a otro.

### C. CALCULO DEL TAMAÑO DE LA MUESTRA.

Todo lo planteado en el muestreo aleatorio simple, con relación a las variables, es valido para el muestreo aleatorio estratificado.

#### 1. Estratificación del universo.

Es necesario dividir el universo en grupos, usando por lo menos una variable. Hay que codificar por separado los elementos de cada grupo de 1 hasta x.

En el caso de la CMA, vamos a usar la variable: "existen niños por debajo de los 6 años en la unidad familiar", para estratificar el universo en dos grupos: los que si y los que no. Hay 100 elementos que no tienen niños menores de 6 años.

Los códigos de esos elementos y su equivalencia con los códigos de muestreo aleatorio simple son:

CODIGO	CODIGO	CODIGO	CODIGO
ESTRATO NO	A.SIMPLE	ESTRATO NO	A.SIMPLE
001	001	051	078
002	002	052	079
003	003	053	080
004	004	054	081
005	005	055	083
006	006	056	084
007	007	057	085
008	008	058	087
009	009	059	088
010	010	060	089
011	012	061	105
012	015	062	106



013	016	063	107
014	017	064	108
015	019	065	109
016	022	066	110
017	023	067	111
018	024	068	112
019	025	069	113
020	026	070	114
021	030	071	115
022	038	072	116
023	039	073	117
024	040	074	120
025	041	075	122
026	042	076	124
027	043	077	125
028	044	078	126
029	045	079	127
030	046	080	129
031	047	081	130
032	048	082	131
033	049	083	132
034	050	084	133
035	051	085	134
036	055	086	135
037	056	087	136
038	057	088	137
039	060	089	141
040	061	090	142
041	062	091	143
042	063	092	144
043	064	093	146
044	065	094	147
045	066	095	148
046	071	096	151
047	072	097	154
048	073	098	155
049	074	099	157
050	076	100	159

Hay 60 elementos que si tienen niños menores de 6 años.  
Los códigos de esos elementos y su equivalencia con los  
códigos de muestreo aleatorio simple son:

CODIGO	CODIGO	CODIGO	CODIGO
ESTRATO SI	A.SIMPLE	ESTRATO SI	A.SIMPLE
01	11	31	091
02	13	32	092
03	14	33	093
04	18	34	094

05	20	35	095
06	21	36	096
07	27	37	097
08	28	38	098
09	29	39	099
10	31	40	100
11	32	41	101
12	33	42	102
13	34	43	103
14	35	44	104
15	36	45	118
16	37	46	119
17	052	47	121
18	053	48	123
19	054	49	128
20	058	50	138
21	059	51	139
22	067	52	140
23	068	53	145
24	069	54	149
25	070	55	150
26	075	56	052
27	077	57	153
28	082	58	156
29	086	59	158
30	090	60	160

2. Formulas o ecuaciones para el calculo del tamaño de muestreo de las variables cuantitativas.

a. Cuando la población es infinita.

$$n = 1/v * \sum_{i=1}^2 (Wi * si)$$

b. Cuando la población es finita.

1). Usando parámetros.

$$no = 1/v * \sum_{i=1}^2 (Wi * Si)$$

$$n = no/1+(no/N)$$

2). Usando estadígrafos.

$$no = 1/v * \sum_{i=1}^2 (Wi * si)$$

$$n = no/1+(no/N)$$

3. formulas o ecuaciones para el calculo del tamaño de muestra de las variables cualitativas dicótomas.

a. Cuando la población es infinita.

$$n = 1/v * \text{sumatoria}(W_i * p_i * q_i)$$

b. cuando la población es finita.

1). Usando parámetros.

$$n_o = 1/v * \text{sumatoria}(W_i * P_i * Q_i)$$

$$n = n_o / (1 + (n_o / N))$$

2). Usando estadígrafos.

$$n_o = 1/v * \text{sumatoria}(w_i * p_i * q_i)$$

$$n = n_o / (1 + (n_o / N))$$

4. Significado de las letras usadas en las formulas o ecuaciones.

$v$  = División del error sobre el margen de confiabilidad  
 $= d / Z$ . En las variables cuantitativas  $d = (\bar{x}_{st} * \%)$ .

$W_i$  = Porcentaje o proporción de elementos de cada estrato, con relación al total del universo.

$s_i^2$  = varianza estadígrafo de cada estrato.

$S_i^2$  = Varianza parámetro de cada estrato.

$\bar{x}_i$  = Media aritmetica de cada estrato.

$\bar{x}_{st}$  = Promedio de las medias de los estratos =  $\text{sumatoria}(\bar{x}_i * W_i)$ .

$p_i$  y  $P_i$  = Proporción de elementos que tiene determinada característica, en cada estrato. La primera es estadígrafo y la segunda es parámetro.

$q_i$  y  $Q_i$  = Proporción de elementos que no tienen determinada característica, en cada estrato. La primera es estadígrafo y la segunda es parámetro.

5. Pasos para el calculo del tamaño de la muestra. En forma secuencial, los pasos para calcular el tamaño de muestra son:

a. Se listan las variables que van a participar en el calculo del tamaño de la muestra. En el caso de la CMA estas son:

- Gasto semanal por comida.

- Ingreso anual de la familia.
- Número de personas de la familia.
- Edad del jefe de la unidad familiar.
- Si hay personas entre 6 y 18 años en la familia.

El nivel de educación, no participa porque es cualitativa no dicótoma. Si hay niños menores de 6 años, no participa porque se convirtió en variable de clasificación o estratificación.

b. Se elige una muestra al azar, en cada uno de los estratos y se anota al frente de cada elemento el valor de cada variable.

1). Para el estrato de los que si tienen niños menores de 6 años.

De ahora en adelante será el estrato 1. El criterio para la elección fue: página 410, columna 1, fila 1, en sentido horizontal y los dos primeros números de cada serie.

Los 5 elementos de cada muestra elegidos y los valores de las variables son:

1	2	3	4	5	6	8
10	3.750	3	1.400.000	1	20	no
15	26.250	6	1.250.000	1	37	si
01	11.250	2	2.500.000	2	47	no
02	48.500	3	2.750.000	4	58	no
14	17.250	5	6.000.000	2	38	si

2). Para el estrato de los que no tienen niños menores de 6 años.

De ahora en adelante será el estrato 2. El criterio para la elección fue: página 411, fila 51, columna 1, en sentido vertical y los tres primeros números de cada serie.

Los 5 elementos de la muestra elegidos y los valores de las variables son:

1	2	3	4	5	6	8
039	11.000	3	1.250.000	2	40	si
090	9.000	4	2.500.000	2	40	no

032	6.000	2	250.000	1	26	no
083	19.000	2	3.000.000	2	30	no
099	18.000	5	3.000.000	2	53	si

c. Se calculan los estadígrafos para cada variable (en cada estrato), que se requieren en las respectivas formulas para calcular la muestra.

1). Para el estrato de los que si tienen niños menores de 6 años:

- Gasto semanal por comida. Media = \$ 21.400. Varianza = \$ 297.424.999.9.

- Número de personas de la familia. Media = 3.8 personas. Varianza = 2.7 personas.

- Ingreso anual de la familia. Media = \$ 2.780.000. Varianza = \$ 367.324.999.800.

- Edad del jefe de la unidad familiar. Media = 40 años. Varianza = 196.5 personas.

- Si hay personas entre 6 y 18 años en la familia. Si hay(p) = 40%. No hay(q) = 60%.

2). Para el estrato de los que no tienen niños menores de 6 años:

- Gasto semanal por comida. Media = \$ 12.600. Varianza = \$ 32.300.000.

- Número de personas de la familia. Media = 3.2 personas. Varianza = 1.7 personas.

- Ingreso anual de la familia. Media = \$ 2.000.000. Varianza = \$ 146.874.999.900.

- Edad del jefe de la unidad familiar. Media = 37.8 años. Varianza = 110.2 años.

- Si hay personas entre 6 y 18 años en la familia. Si hay(p) = 40%. No hay(q) = 60%.

d. Se determina el margen de error de muestreo y el margen de confiabilidad a usar en todos los calculos.

Acá se aplica todo lo planteado en muestreo aleatorio simple y de la misma forma. Se va usar un error del 20% y un margen de confiabilidad del 90%. El margen de confiabilidad equivale a 1.65 desviaciones estandar.

e. Se halla el tamaño de muestra para cada variable. En éste caso hay que hallar 5 tamaños de muestra:

1). Gasto semanal por comida.

$$no = 1/v * \sumatoria(Wi * si)^2$$

$$no = 1/[(15.900 * 0.20)/1.65^2] * (0.375 * 297.424.999.9 + 0.625 * 32.300.000)$$

$$no = 1/3.714.380.2 * 131.721.875 = 35$$

$$n = no/(1+no/N) = 35/(1+35/160) = 29$$

2). Número de personas de la familia.

$$no = 1/v * \sumatoria(Wi * si)^2$$

$$no = 1/[(3.425 * 0.20)/1.65^2] * (0.375 * 2.7 + 0.625 * 1.7)$$

$$no = 1/0.172 * 2,075 = 12$$

$$n = no/(1+no/N) = 12/(1+12/160) = 11$$

3). Ingreso anual de la familia.

$$no = 1/v * \sumatoria(Wi * si)^2$$

$$no = 1/[(2.292.500 * 0.20)/1.65^2] * (0.375 * 367.324.999.800 + 0.625 * 146.874.999.900)$$

$$no = 1/7.721.662.075 * 22.954.374.990 = 3$$

$$n = no/(1+no/N) = 3/(1+3/160) = 3$$

4). Edad del jefe de la unidad familiar.

$$no = 1/v * \sumatoria(Wi * si)^2$$

$$no = 1/[(38.63 * 0.20)/1.65^2] * (0.375 * 196.5 + 0.625 * 110.2)$$

$$no = 1/22 * 143 = 7$$

$$n = no/(1+no/N) = 7/(1+7/160) = 7$$

5). Si hay personas entre 6 y 18 años en la familia.

$$n_o = 1/v * \text{sumatoria}(W_i * p_i * q_i)$$

$$n_o = 1 / \left( \frac{0.20}{1.65} \right)^2 * \left( 0.375^2 * 0.40^2 * 0.60^2 + 0.625^2 * 0.40^2 * 0.60^2 \right)$$

$$n_o = 1/0.015 * 0.24 = 16$$

$$n = n_o / (1 + n_o/N) = 16 / (1 + 16/160) = 15$$

f. Se determina el tamaño de la muestra definitiva. El tamaño de muestra a usar en la investigación será de 29 elementos ( $n = 29$ ). Esta cifra es suficiente para la variable que lo generó y más que suficiente para las variables que dieron menores tamaños de muestra.

#### D. ELECCION DEL TAMAÑO DE MUESTRA.

Antes de entrar a elegir el tamaño de muestra, hay que decidir sobre la repartición de los 29 elementos en los dos estratos.

Existen dos alternativas para repartir los 29 elementos en los dos estratos: de una forma proporcional y de una forma no proporcional.

No proporcional, sería repartir los 29 elementos de cualquier forma: 15 y 14, 10 y 19, etc..

Proporcional, sería repartir los 29 elementos de acuerdo al peso que tiene cada estrato dentro del universo: estrato 1  $0.375 * 29 = 11$  y estrato 2  $0.625 * 29 = 18$ .

En éste caso vamos a aplicar el método proporcional. Vamos a elegir 11 elementos en el estrato 1 y 18 elementos en el estrato 2.

1. Elección del tamaño de muestra del estrato 1. Para elegir los 11 elementos dentro de los 60 que componen el estrato 1, hay que usar la tabla de números aleatorios y definir un criterio.

El criterio para el caso fué el siguiente: página 412, fila 101, columna 1, en sentido vertical y los dos primeros números de cada serie de estos.

1	2	3	4	5	6	8
13	14.250	5	3.500.000	3	33	si
21	7.500	3	1.300.000	3	39	no
60	11.000	6	6.000.000	2	48	si
35	30.000	9	8.000.000	1	58	si
38	15.000	6	1.900.000	1	39	si
10	3.750	3	1.400.000	1	20	no
36	19.000	5	4.000.000	1	51	si
24	15.000	4	3.500.000	1	44	si
45	10.000	3	5.000.000	3	58	no
19	10.000	4	12.500.000	5	45	si
42	50.000	7	7.000.000	1	53	si

No se debe repetir ningún elemento, pues sería como entrevistar dos veces o más a la misma persona y sobre lo mismo.

2. Elección del tamaño de muestra del estrato. Para elegir los 18 elementos dentro de los 100 que componen el estrato 2, hay que usar la tabla de números aleatorios y definir un criterio.

El criterio para el caso fue el siguiente: página 413, columna 1, fila 151, en sentido horizontal y los 3 últimos números de cada serie de estos.

1	2	3	4	5	6	8
096	18.000	4	4.000.000	2	38	si
055	18.000	3	3.800.000	2	47	no
011	10.000	2	4.000.000	1	35	no
067	7.500	1	1.300.000	2	21	no
056	15.000	4	4.500.000	3	48	si
019	26.500	3	3.250.000	2	51	si
060	18.000	6	3.000.000	3	32	si
094	22.000	3	3.000.000	3	48	si
075	26.000	2	6.000.000	2	50	no
032	6.000	2	250.000	1	26	no
013	15.000	2	2.000.000	2	40	si
012	19.500	3	1.500.000	1	38	no
072	8.000	2	1.000.000	2	39	no
077	21.000	2	1.500.000	3	29	no
064	18.000	1	2.000.000	4	42	no
086	30.000	2	3.500.000	3	44	si
081	15.000	3	3.000.000	2	48	si
059	38.000	6	3.500.000	2	49	si

No se debe repetir ningún elemento, pues sería como entrevistar dos veces a la misma persona y sobre lo mismo.



#### E. DISEÑO DEL FORMATO PARA REGISTRAR LA INFORMACION.

Existen dos diferencias con el formato usado para registrar la información en muestreo aleatorio simple:

- La pregunta número 6 no va.
- Con la variable de la pregunta 6, se estructura la parte de "información de clasificación".

La parte de información de clasificación va al terminar la información de identificación y antes de la información básica.

Esta parte iría más o menos así:

#### INFORMACION DE CLASIFICACION

Hay niños menores de 6 años en la familia?.

Si 1

No 2

#### F. ELABORACION DE LOS CUADROS DE TABULACION Y DE LOS CUADROS ESTADISTICOS.

Como mínimo debe salir un cuadro de tabulación y un cuadro estadístico por cada variable.

En éste caso serían 6 cuadros de tabulación y 6 cuadros estadísticos. Una de las variables (hay niños menores de 6 años en la familia) pasa a ser variable de clasificación y sirve de variable de cruce de las otras 6.

No se debe olvidar que los cuadros de tabulación son el borrador de los cuadros estadísticos.

1. Cuadros de tabulación. Como ejemplo, se presenta el cuadro de tabulación que saldría de la variable: "hay personas entre 6 y 18 años en la familia". El cuadro es el siguiente:

ALTERNATIVAS	ESTRATO 1 (SI)	ESTRATO 2 (NO)
Si	////////	////////

No

///

//////////

En el muestreo aleatorio estratificado, las alternativas de cada una de las 6 variables básicas, estarán cruzadas contra las alternativas de la variable de clasificación.

2. Cuadros estadísticos. Como ejemplo, se presenta el cuadro estadístico que saldría del cuadro de tabulación anterior. El cuadro es el siguiente:

## Número 2

Socios que tienen o no personas entre 6 y 18 años en la familia. Por total y estrato.

ALTERNATIVAS	TOTAL		ESTRATO 1		ESTRATO 2	
	#	%	#	%	#	%
Si	17	58.6	8	72.7	9	50
No	12	41.4	3	27.3	9	50
Total respuestas	29	100.0	11	100.0	18	100
Número de entrevistados	(29)		(11)		(18)	

Fuente: Este estudio.

## G. INFERENCIA.

Existen dos clases de inferencia: Estimación y prueba de hipótesis.

1. Estimación de intervalo. Lo planteado en muestreo aleatorio simple, sirve para ésta parte.

a. Formulas o ecuaciones para la estimación de intervalo:

1). Cuando la variable es cuantitativa:

$$\hat{X}_{st} = \bar{X}_{st} - t_{\alpha} \cdot \sqrt{V_{xst}}$$

2). Cuando la variable es cualitativa dicótoma.

$$\hat{Pst} = \hat{Pst} - t \text{ o } Z\text{raiz}(\hat{Vpst})$$

b. Significado de las letras usadas en las formulas o ecuaciones:

$\hat{Pst}$  = Proporción muestral estratificada = **sumatoria**( $W_i * P_i$ )

$\hat{Xst}$  = Media muestral estratificada = **sumatoria**( $W_i * \bar{X}_i$ )

$\hat{Vxst}$  = Estimador de la varianza =  $\frac{1}{N} \sum_{i=1}^k [N_i(N_i - n_i) * \frac{s_i^2}{n_i}]$  o  $[(1-f)/n] * \sum_{i=1}^k (W_i * s_i^2)$

$\hat{Vpst}$  = Estimador de la varianza =  $[(1-f)/n] * \sum_{i=1}^k (W_i * p_i * q_i)$

t = Distribución t. Se usa cuando n es menor o igual a 30.

Z = Distribución Z. Se usa cuando n es mayor que 30.

c. Pasos para una estimación de intervalo. Llevando un orden, los pasos a seguir son los siguientes:

- Se calculan los estadígrafos de la variable. En éste caso la del cuadro # 2. Estos son:

$$p_1 = 72.7 \%$$

$$q_1 = 27.3 \%$$

$$p_2 = 50 \%$$

$$q_2 = 50 \%$$

$$\hat{Pst} = 58.6 \%$$

- Se escoje la formula o ecuación adecuada. En el caso del cuadro # 2, ésta sería:

$$\hat{Pst} = \hat{Pst} - t\text{raiz}(\hat{Vpst})$$

- Se determina el valor de t y se busca en la tabla.

Acá se aplica lo indicado en la parte de muestreo aleatorio simple. Lo unico que cambia es que los grados

de libertad(G.L) son iguales a  $n_1 + n_2 - 2$ .

En éste caso serían:  $29 - 2 = 27$ .

Si se determina que  $t$  sea igual al 10%, en la tabla(anexo # 4) se debe buscar el valor que corresponda a 0.05 y 27. Este valor es 1.703.

- Se aplica la formula o ecuación respectiva, para hallar el posible parámetro.

$$\hat{Pst} = 0.586 - 1.703 * \text{raiz}(0.0166)$$

$$\hat{Pst} = 0.586 - 0.219$$

$$\hat{Pst} = 0.367 \text{ -- } 0.805$$

- Se interpreta el intervalo y se aconseja acerca de la decisión a tomar. La interpretación sería: en el 90% de las posibles muestras el porcentaje de socios que tienen personas entre 6 y 18 años en la familia, esta entre el 36.7% y el 80.5%.

Si la introducción de textos escolares estuviera supeditada a que el porcentaje de socios que tengan personas entre 6 y 18 años en la familia, fuera de por lo menos el 40%, entonces se aconsejaría que se introdujera esa línea de productos, pues hay más probabilidad de que el porcentaje sea superior al 40% que inferior a ésta cifra.

2. Prueba de hipótesis. Acá se aplica lo escrito, en ésta parte, en muestreo aleatorio simple.

Lo planteado acá sobre prueba de hipótesis, se puede aplicar cuando se trabaja con dos estratos y cuando se comparan dos investigaciones de muestreo aleatorio simple.

a. Cuando la variable es cuantitativa.

1). Formas de las hipótesis. Hay tres alternativas de hipótesis nula y alterna para escoger:

$$- H_0: U_1 - U_2 = 0$$

$$H_1: U_1 - U_2 \text{ no igual } 0$$

>

$$- H_0: U_1 - U_2 = 0$$

$$H_1: U_1 - U_2 < 0$$

$$- H_0: U_1 - U_2 = 0$$

$$H_1: U_1 - U_2 > 0$$

Lo demás, es lo mismo que se escribió en muestreo aleatorio simple.

2). Procedimiento para hallar el valor calculado y el valor teórico, cuando  $n$  es mayor que 30. Es el mismo que se describió en muestreo aleatorio simple. Lo único diferente son las formulas que se usan para hallar  $Z_c$  y  $T_c$ . Estas estan contenidas en el anexo No 5.

3). Procedimiento para hallar el valor calculado y el valor teórico, cuando  $n$  es menor o igual a 30. Es el mismo que se describió en muestreo aleatorio simple. pero hay dos aspectos diferentes:

- Las formulas que se usan para hallar  $Z_c$  y  $T_c$  estan contenidas en el anexo No 5.

- La forma para calcular los grados de libertad (G.L.). Estos se calculan asi:  $n_1 + n_2 - 2$ .

b. Cuando la variable es cualitativa dicótoma.

1). Formas de las hipótesis. Hay tres alternativas de hipótesis nula y alterna para escoger:

$$- H_0: P_1 - P_2 = 0$$

$$H_1: P_1 - P_2 \text{ no igual } 0$$

$$- H_0: P_1 - P_2 = 0$$

$$H_1: P_1 - P_2 < 0$$

$$- H_0: P_1 - P_2 = 0$$

$$H_1: P_1 - P_2 > 0$$

Lo demás es lo mismo que se escribió en muestreo aleatorio simple.

2). Procedimiento para hallar el valor calculado y el valor teórico. Para el valor calculado, solo se trabaja con la distribución Z. Las condiciones se encuentran en el anexo No 5, dependiendo de la formula a usar. Si no se cumplen las condiciones, es necesario aumentar el tamaño de muestra hasta cumplirlas.

El Z de tabla ( $Z_t$ ) se calcula de la misma forma que para las variables cuantitativas.

c. Significado de las letras usadas en las formulas o ecuaciones.

$H_0$  = Hipótesis nula.

$H_1$  = Hipótesis alterna.

$U_1$  = Parámetro de  $\bar{X}$  del primer estrato o del primer estudio.

$U_2$  = Parámetro de  $\bar{X}$  del segundo estrato o del segundo estudio.

$P_1$  = Parámetro de  $p$  del primer estrato o del primer estudio.

$P_2$  = Parámetro de  $p$  del segundo estrato o del segundo estudio.

$Z_c$  = Z calculado.

$Z_t$  = Z de tabla. El signo es el mismo del Z calculado.

$T_c$  = T calculado.

$T_t$  = T de tabla. El signo es el mismo del T calculado.

Alfa = Probabilidad de rechazar  $H_0$ , siendo verdadera. También se conoce como error tipo I. Todo lo demás, es lo mismo que se escribió en muestreo aleatorio simple.

G.L. = Grados de libertad. Son iguales a :  $n_1 + n_2 - 2$ .

d. Zonas de aceptación y de rechazo de la hipótesis nula. Aquí sirve todo lo escrito en muestreo aleatorio simple.

Solo habría que aclarar que cuando se indica que  $U_1 > P_1 - U_2$  o  $P_2 - 0$ , se está diciendo que lo de el primer

estrato o estudio es mayor o igual que lo del segundo estrato o estudio.

Cuando se indica que  $U1$  o  $P1 - U2$  o  $P2 < 0$ , se esta diciendo que lo del primer estrato o estudio es menor que lo del segundo estrato o estudio.

En otros terminos, hay que tener en cuenta que es una resta algebraica cuyo resultado puede ser con signo mas (+) o menos (-).

c. Pasos para una prueba o contraste de hipótesis. En su orden los pasos son los siguientes:

- Se indica una decisión a tomar, contemplando un posible parámetro. En el caso, una de las decisiones a tomar es: " Si el porcentaje de socios que tienen personas entre 6 y 18 años en la familia, en el primer estrato, es igual o superior al del segundo estrato, se abrirá el servicio de jardin infantil en los puntos de venta".

- Se plantean las hipótesis nula y alterna, de la forma que mejor se ajuste a la decisión. En este caso la forma es:

$H_0: P1 - P2 - 0$  Se abrirá el jardin ...

$H_1: P1 - P2 < 0$  No se abrirá el jardin ...

- Se toma una muestra y se hallan los estadígrafos de la respectiva variable. En este caso, el estadígrafo para el estrato 1 es 72.7 % y estadígrafo para el estrato 2 es 50 %.

Se halla el valor de Z calculado. La formula es la siguiente ( ver anexo No 5):

$$Z_c = (p_1 - p_2) / \sqrt{p * (1-p) * (1/n_1 + 1/n_2)}$$

$$P = (x_1 + x_2) / (n_1 + n_2)$$

$$p = (8 + 9) / (11 + 18) = 58.6 \%$$

Las condiciones son:

$$n_i * p > 5 \quad y \quad n_i * (1 - p) > 5$$

$$11 * 0.586 = 6 \quad y \quad 11 * (1 - 0.586) = 5$$

$$18 * 0.586 = 11 \quad y \quad 18 * (1 - 0.586) = 8$$

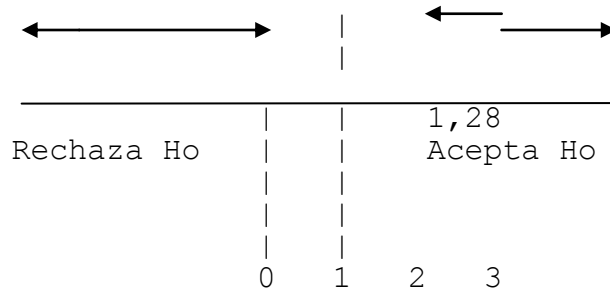
Aproximando a enteros se cumplen las condiciones.

$$Z_c = (0.727 - 0.50) / \sqrt{0.586 * (1 - 0.586) * (1/11 + 1/18)}$$

$$Z_c = 1.2$$

- Se halla el valor de Z de tabla ( $Z_t$ ). En este caso, como es más costoso aceptar  $H_0$  siendo falsa (error tipo II), se debe asignar un alfa alto. Puede ser 10%. Se trabaja con el complemento: 90%. Como el contraste es de una cola se resta 50%, esto da 40%. 40% se divide por 100 y se busca en la tabla. El valor es: 1.28 (signo más).

Se grafica  $Z_c$  y  $Z_t$ , y se indica la zona de aceptación y  $H_0$ .



En este caso se rechaza  $H_0$ . Es decir, que los estadígrafos no dan para pensar que en el estrato 1 haya un porcentaje, de socios que tienen personas entre 6 y 18 años en la familia, mayor o igual al estrato 2.

- Se concluye acerca de la decisión anterior. Como la hipótesis nula fué rechazada, entonces no se deberá crear el jardín infantil en cada punto de venta.

#### H. EJERCICIO DE AUTOEVALUACION.

Tome el caso usado en las unidades anteriores y realice las siguientes actividades(según el muestreo aleatorio estratificado):

- Use una variable dicótoma para estratificar el universo y estructure los dos estratos.
- Calcule el tamaño de muestra.
- Elija la muestra calculada.



- Diseñe el formato a usar para registrar la información que se va a compilar.
- Aplique el formato diseñado a la muestra elegida.
- Elabore los cuadros de tabulación y estadísticos que salgan del formato.
- Tome una variable y haga la inferencia desde el punto de vista de la estimación de intervalo y la prueba de hipótesis.

## UNIDAD 5

### OTROS METODOS ALEATORIOS O AL AZAR PARA ELEGIR UNA MUESTRA

#### A. OBJETIVOS ESPECIFICOS.

Al finalizar ésta unidad, el estudiante estará en capacidad de:

1. Definir muestreo sistemático, de conglomerados, polietapico y por fases multiples.
2. Elegir una muestra por el método sistemático, de conglomerados, polietapico y por fases multiples.

#### B. GENERALIDADES.

El calculo de un tamaño de muestra en forma aleatoria o al azar tiene dos alternativas: por el método simple o por el método estratificado. El proceso para cada método esta indicado en las unidades 3 y 4.

Para la elección de una muestra en forma aleatoria o al azar, hay varios métodos. En las unidades 3 y 4 solo se uso el de las tablas de números aleatorios. En ésta unidad se describirán los pasos a seguir, para la

elección de una muestra, en los métodos sistemático, de conglomerados, polietapico y por fases multiples. Para cada uno se hará referencia cuando sea muestreo simple y cuando sea muestreo estratificado.

Por último, hay que tener en cuenta que los métodos para el calculo y la elección de una muestra, se pueden mezclar en un mismo estudio o investigación, dando origen a los llamados métodos mixtos.

### C. MUESTREO SISTEMATICO.

1. Concepto. Es un método para elegir un tamaño de muestra, tomando los elementos muestrales a intervalos regulares.

Este método se considera aleatorio o al azar, cuando el marco muestral esta ordenado al azar.

2. Procedimiento cuando el muestreo es simple. Los pasos a seguir son los siguientes:

a. Se calcula el tamaño de la muestra. En la unidad 3 dió 18 elementos.

b. Se calcula el intervalo de elección o salto sistemático.  $i = N/n$ .

Para el caso de la unidad 3, éste sería:  $i = 160/18 = 8.88$ . Se aproxima y daría 9.

c. Se elige un número al azar entre 1 y 9. Supongamos que sea 2. Este será el primer elemento a elegir.

d. Se eligen los códigos de los elementos que van a hacer parte de la muestra, usando como constante el intervalo de elección(9) así:

CODIGOS.

002  
 002 + 009 = 011  
 011 + 009 = 020  
 020 + 009 = 029  
 029 + 009 = 038  
 038 + 009 = 047  
 047 + 009 = 056  
 056 + 009 = 065  
 065 + 009 = 074

CODIGOS.

074 + 009 = 083  
 083 + 009 = 092  
 092 + 009 = 101  
 101 + 009 = 110  
 110 + 009 = 119  
 119 + 009 = 128  
 128 + 009 = 137  
 137 + 009 = 146  
 146 + 009 = 155

e. Se busca para cada código de tarjeta el valor de cada variable y se anota al frente, como en la unidad 3.

f. Lo demás se hace lo mismo que en la unidad 3.

3. Procedimiento cuando el muestreo es estratificado. Los pasos a seguir son los siguientes:

a. Se calcula el tamaño de muestra y se divide entre los estratos. En la unidad 4 dio 11 para el estrato 1 y 18 para el estrato 2.

b. Se calcula el intervalo de elección o salto sistemático para cada estrato.  $i = N_i / n_i$ .

- Estrato 1.

$i = N_1 / n_1$ .  $i = 60 / 11 = 5.45$ . Se aproxima a 5.

- Estrato 2.

$i = N_2 / n_2$ .  $i = 100 / 18 = 5.55$ . Se aproxima a 6.

c. Se elige un número al azar entre 1 y 5 para el estrato 1 y entre 1 y 6 para el estrato 2.

Supongamos que es 3 para el estrato 1 y 4 para el estrato 2.

d. Se eligen los códigos de los elementos que van a hacer parte de la muestra en cada estrato, usando como constantes los intervalos de elección (5 y 6 respectivamente), así:

- Estrato 1.

CODIGOS.

03

03 + 05 = 08

08 + 05 = 13

13 + 05 = 18

18 + 05 = 23

23 + 05 = 28

CODIGOS.

28 + 05 = 33

33 + 05 = 38

38 + 05 = 43

43 + 05 = 48

48 + 05 = 53

33 + 05 = 38

- Estrato 2.

CODIGOS.

004

CODIGOS.

052 + 006 = 058

004 + 006 = 010	058 + 006 = 064	
010 + 006 = 016	064 + 006 = 070	
016 + 006 = 022	070 + 006 = 076	
022 + 006 = 028	076 + 006 = 082	
028 + 006 = 034	082 + 006 = 088	
034 + 006 = 040		088 + 006 = 094
040 + 006 = 046	094 + 006 = 100	
046 + 006 = 052	100 + 006 = 106	

Como no hay 106, se da la vuelta y el 106 será el 006.

e. Se busca para cada código de tarjeta el valor de cada variable y se anota al frente, como en la unidad 4.

f. Lo demás se hace lo mismo que en la unidad 4.

#### D. MUESTREO POR CONGLOMERADOS.

1. Concepto. Es un método para elegir un tamaño de muestra, del conjunto de todas las posibles muestras (de tamaño  $n$ ) que pueden salir de un universo determinado.

Se denomina conglomerado a cada una de las muestras (de tamaño  $n$ ) que salen de un universo. Cualquiera de las muestras (de tamaño  $n$ ) sirven.

2. Procedimiento cuando el muestreo es simple. Los pasos a seguir son los siguientes:

a. Se calcula el tamaño de muestra. En la unidad 3 dió 18 elementos.

b. Se calcula el número de conglomerados o intervalo de elección o salto sistemático.  $i = N/n$ .

Para el caso de la unidad 3, éste sería:  
 $i = 160/18 = 8.88$ . Se aproxima y daría 9.

El 9 indica que salen (aproximadamente) ese número de muestras de 18 elementos cada una.

c. Se eligen los códigos de los elementos que van a formar parte de cada conglomerado. La constante 9 se usa en la misma forma que en el sistemático.

- Primer conglomerado.

001  
 001 + 009 = 010

010 + 009 = 019  
019 + 009 = 028  
028 + 009 = 037  
Etc.

- Segundo conglomerado.

002  
002 + 009 = 011  
011 + 009 = 020  
020 + 009 = 029  
029 + 009 = 038  
Etc.

- Tercer conglomerado.

003  
003 + 009 = 012  
012 + 009 = 021  
021 + 009 = 030  
030 + 009 = 039  
Etc.

- Cuarto conglomerado.

004  
004 + 009 = 013  
013 + 009 = 022  
022 + 009 = 031  
031 + 009 = 040  
Etc.

- Quinto conglomerado.

005  
005 + 009 = 014  
014 + 009 = 023  
023 + 009 = 032  
032 + 009 = 041  
Etc.

- Sexto conglomerado.

006  
006 + 009 = 015  
015 + 009 = 024  
024 + 009 = 033  
033 + 009 = 042  
Etc.

- Séptimo conglomerado.

007  
 007 + 009 = 016  
 016 + 009 = 025  
 025 + 009 = 034 + 009 = 034  
 034 + 009 = 043  
 Etc.

- Octavo conglomerado.

008  
 008 + 009 = 017  
 017 + 009 = 026  
 026 + 009 = 035  
 035 + 009 = 044  
 Etc.

- Noveno conglomerado.

009  
 009 + 009 = 018  
 018 + 009 = 027  
 027 + 009 = 036  
 036 + 009 = 045  
 Etc.

d. Se elige uno de los conglomerados, al azar, como muestra.

e. Se busca para cada código de tarjeta el valor de cada variable y se anota al frente, como en la unidad 3.

f. Lo demás se hace lo mismo que en la unidad 3.

3. Procedimiento cuando el muestreo es estratificado. Los pasos a seguir son los siguientes:

a. Se calcula el tamaño de muestra y se divide entre los estratos. En la unidad 4 dió 11 para el estrato 1 y 18 para el estrato 2.

b. Se calcula el número de conglomerados o intervalo de elección o salto sistemático para cada estrato.  $i = N_i/n_i$ .

- Estrato 1.

$i = N_1/n_1 = 60/11 = 5.45$ . Se aproxima a 5.

- Estrato 2.

$i = N2/n2 = 100/18 = 5.55$ . Se aproxima a 6.

c. Se eligen los códigos de los elementos que van a formar parte de cada conglomerado, en cada estrato. Las constantes son 5 y 6, y se usan de la misma forma que en el sistemático.

1). Estrato 1.

- Primer conglomerado.

01				
01 + 05 = 06				
06	+	05	=	11
11 + 05 = 16				
16 + 05 = 21				
Etc.				

- Segundo conglomerado.

02  
 02 + 05 = 07  
 07 + 05 = 12  
 12 + 05 = 17  
 17 + 05 = 22  
 Etc.

- Tercer conglomerado.

03  
 03 + 05 = 08  
 08 + 05 = 13  
 13 + 05 = 18  
 18 + 05 = 23  
 Etc.

- Cuarto conglomerado.

04  
 04 + 05 = 09  
 09 + 05 = 14  
 14 + 05 = 19  
 19 + 05 = 24  
 Etc.

- Quinto conglomerado.

05  
 05 + 05 = 10  
 10 + 05 = 15  
 15 + 05 = 20

20 + 05 = 25  
Etc.

2). Estrato 2.

- Primer conglomerado.

001  
001 + 006 = 007  
007 + 006 = 013  
013 + 006 = 019  
019 + 006 = 025  
Etc.

- Segundo conglomerado.

002  
002 + 006 = 008  
008 + 006 = 014  
014 + 006 = 020  
020 + 006 = 026  
Etc.

- Tercer conglomerado.

003  
003 + 006 = 009  
009 + 006 = 015  
015 + 006 = 021  
021 + 006 = 027  
Etc.

- Cuarto conglomerado.

004  
004 + 006 = 010  
010 + 006 = 016  
016 + 006 = 022  
022 + 006 = 028  
Etc.

- Quinto conglomerado.

005  
005 + 006 = 011  
011 + 006 = 017  
017 + 006 = 023



023 + 006 = 029  
Etc.

- Sexto conglomerado.

006  
006 + 006 = 012  
012 + 006 = 018  
018 + 006 = 024  
024 + 006 = 030  
Etc.

d. Se elige, al azar, un conglomerado en cada estrato. El conglomerado elegido es la muestra para cada estrato.

e. Se busca para cada código de tarjeta el valor de cada variable y se anota al frente, como en la unidad 4.

f. Lo demás se hace lo mismo que en la unidad 4.

#### E. MUESTREO POR ETAPAS.

1. Concepto. Es un método para elegir un tamaño de muestra, en el cual se llega al elemento muestral después de más de una etapa.

Lo que se elige en cada etapa se llama unidad muestral y lo que se elige en la última etapa es unidad muestral y elemento muestral al mismo tiempo.

La elección de casas de una ciudad se puede hacer en cinco etapas:

- Primera etapa. Se eligen comunas.
- Segunda etapa. Se eligen barrios dentro de las comunas elegidas.
- Tercera etapa. Se eligen manzanas dentro de los barrios.
- Cuarta etapa. Se eligen viviendas dentro de las manzanas.
- Quinta etapa. Se eligen hogares dentro de las viviendas y se entrevista al ama de casa.

Este método permite la concentración de la elección de la muestra y evita andar, en forma dispersa, buscando los elementos a entrevistar.

2. Procedimiento cuando el muestreo es simple. Los pasos a seguir son los siguientes:

a. Se calcula el tamaño de muestra. En la unidad 3 dió 18 elementos.

b. Se determina el número de etapas y el número de unidades muestrales a elegir en cada etapa.

Como el marco muestral esta compuesto de varias planas, las etapas pueden ser dos: la primera sería elegir planas y la segunda elegir los códigos de las tarjetas.

Como la muestra es de 18 elementos y cada plana tiene más de 18 elementos, se necesitaría elegir una plana en la primera etapa y 18 elementos dentro de la plana en la segunda etapa.

c. Se elige, al azar, la plana y los 18 códigos de las tarjetas dentro de la plana.

d. Se busca para cada código de tarjeta el valor de cada variable y se anota al frente, como en la unidad 3.

e. Lo demas se hace lo mismo que en la unidad 3.

3. Procedimiento cuando el muestreo es estratificado. Los pasos a seguir son los siguientes:

a. Se calcula el tamaño de muestra y se divide entre los estratos. En la unidad 4 dió 11 para el estrato 1 y 18 para el estrato 2.

b. Este paso y los pasos c, d y e, son los mismos que para el muestreo simple. Pero, hay que aplicarlos separadamente para cada estrato.

#### F. MUESTREO POR FASES.

1. Concepto. Es un método para elegir un tamaño de muestra, donde en cada fase se entrevista a la unidad o elemento muestral elegido, se cambian las variables a medir y se varia el tamaño de muestra (reduciendose a medida que se cambia de fase).

Este método es apropiado para hacer un estudio de profundidad de una muestra.

En el caso solo se puede hacer muestreo de una sola fase, pues no se especifican otras variables a medir en

una segunda fase.

En cada fase se ejecutan todos los pasos del muestreo simple o estratificado.

La primera muestra se toma dentro del universo, la segunda muestra se toma dentro de la primera, la tercera muestra se toma dentro de la tercera y así sucesivamente. Las variables deben ir profundizando el tema que se está investigando.

Si en el muestreo simple la primera fase tiene un tamaño de muestra de 18 elementos, la información se toma de las tarjetas y hay que medir 7 variables. En una segunda fase se elige un tamaño de muestra de entre los 18 elementos, la información se debe extraer de las tarjetas y se deben medir otras variables diferentes a las 7 primeras.

#### G. METODOS MIXTOS.

En una investigación se pueden combinar los métodos vistos hasta éste momento. Por ejemplo, en un muestreo por fases los elementos se pueden elegir en forma sistemática en cada fase o en una fase en forma sistemática y en otra a través de números aleatorios.

Más aun, pueden combinarse métodos aleatorios y métodos no aleatorios. Por ejemplo, el tamaño de muestra se puede calcular en forma aleatoria y la elección hacerse en forma no aleatoria y viceversa.

#### H. EJERCICIO DE AUTOEVALUACION.

Tome el caso usado en las unidades anteriores y elija las muestras de muestreo aleatorio simple y aleatorio estratificado, usando los métodos:

- Sistemático.
- Conglomerados.
- Por fases.
- Polietapico.

## UNIDAD 6

## METODOS DE MUESTREO NO ALEATORIOS O NO AL AZAR

## A. OBJETIVOS ESPECIFICOS.

Al finalizar ésta unidad, el estudiante estará en capacidad de:

1. Definir muestreo no aleatorio o no al azar, dirigido, por cuotas y por conveniencia.
2. Calcular y elegir una muestra por el método dirigido, por cuotas y por conveniencia.

## B. CONCEPTO DE MUESTREO NO ALEATORIO O NO AL AZAR.

Se llama así a los métodos para calcular y elegir una muestra, donde todos los elementos no tienen la misma posibilidad o probabilidad de participar.

## C. MUESTREO DIRIGIDO.

1. Concepto. Es un método de muestreo, donde se calcula y se elige la muestra de acuerdo al criterio del investigador.

2. Procedimiento. Los pasos a seguir son:

a. El investigador determina el tamaño de muestra, de acuerdo a su experiencia.

b. Se elige el tamaño de muestra, de acuerdo al criterio del investigador.

c. Los otros pasos son los mismos que en el muestreo aleatorio simple.

D. MUESTREO POR CUOTAS.

1. Concepto. Es un método de muestreo, donde se calcula y elige la muestra (por grupos o estratos) de acuerdo al criterio del investigador.

2. Procedimiento. Los pasos a seguir son:

a. El investigador determina el tamaño de la muestra.

b. El investigador reparte la muestra entre los grupos o estratos, en que ha dividido el universo.

c. Se elige el tamaño de muestra, en cada estrato, de acuerdo al criterio del investigador.

d. Los otros pasos son los mismos que en el muestreo aleatorio estratificado.

E. MUESTREO POR CONVENIENCIA.

1. Concepto. Es un método de muestreo, donde se calcula y se elige la muestra buscando la facilidad en todo el proceso.

Este método tiende a ser sesgado, precisamente por buscar la facilidad en lo que se hace.

2. Procedimiento. Los pasos a seguir son:

a. El investigador determina el tamaño de muestra que más se le facilite.

b. Si es por cuotas, se reparte el tamaño de muestra entre los grupos o estratos. Si es dirigido no se reparte la muestra.

c. Se elige el tamaño de muestra, buscando os elementos

más fáciles de contactar.

d. Los demás pasos se hacen lo mismo que en el muestreo aleatorio simple o aleatorio estratificado, según el caso.

#### F. EJERCICIO DE AUTOEVALUACION.

Tome el caso usado para las unidades anteriores y realice las siguientes actividades:

1. Indique cuál puede ser el tamaño de muestra y cómo elegirlo, desde el punto de vista del muestreo dirigido. Sustente la respuesta.

2. Indique cuál puede ser el tamaño de muestra y cómo elegirlo, desde el punto de vista del muestreo por cuotas. Sustente la respuesta.

3. Indique cuál puede ser el tamaño de muestra y cómo elegirlo, desde el punto de vista del muestreo por conveniencia. Sustente la respuesta.

## BIBLIOGRAFIA

- A. Aaker, David A. y George S. Day. Investigación de mercados. 3 ed. México, Ed. Mc. Graw Hill, 1988. 715 p.
- B. Acosta, J. José A. Manual para la planeación y desarrollo de una investigación formal de mercadeo. Tunja, trabajo de acenso a profesor titular, 1988. 166 p.
- C. Chao, Lincoln L. Estadística para las ciencias administrativas. 2 ed. México, Ed. Mc. Graw Hill, 1980. 472 p.
- D. Kinnear, C. Thomas y James R. Taylor. Investigación de mercados. Bogotá, Ed. Mc. Graw Hill, 1979. 740 p.
- E. Martinez, B. Ciro. Muestreo: Algunos métodos y sus aplicaciones prácticas. Bogotá, Ed. Ecoe, 1984. 299 p.
- F. Pope, Jeffrey. Investigación de mercados: Guía maestra para el investigador profesional. Bogotá, Ed. Norma, 1984. 348 p.
- G. Richards, Larry E. y Jerry J. La Cava. Estadística en los negocios: Por qué y cuando. Bogotá, Ed. Mc. Graw Hill, 1980. 454 p.
- H. Tagliacarne, Guglielmo. Teoría y práctica de las

investigaciones de mercadeo. 2 ed. Barcelona, Ed.  
Ariel, 1962. 821 p.

ANEXO # 1

CASO PARA ILUSTRAR LA APLICACION DE LA TEORIA



#### COOPERATIVA MUISCA DE ALIMENTOS

La señorita Vélez es la gerente general de la Cooperativa Muisca de alimentos (CMA).

Recientemente, se dió cuenta que había perdido contacto con las costumbres de compra de los miembros de la cooperativa.

La CMA parecía mucho más grande ahora, en comparación con algunos años atrás.

La gerente pensaba si podría hacer uso de algún tipo de información que estuviera a su alcance, con el fin de ampliar y actualizar sus conocimientos en relación con los hábitos de compra de los miembros de la cooperativa.

La gerente esperaba poder utilizar estos conocimientos, para planificar mejor la mezcla y cantidad de productos que la CMA ofrecía.

La señorita Vélez y un pequeño grupo de voluntarios fundaron la CMA en 1984.

La cooperativa había aumentado los 10 miembros originales de 1984, a 160 miembros de Diciembre de 1989.

La empresa está localizada en una vieja casona al norte de la ciudad de Tunja.

La CMA obtuvo socios de varios barrios de la ciudad de Tunja y de algunas poblaciones cercanas.

El objetivo de la CMA es suministrar productos alimenticios de alta calidad a precios bajos, por debajo de los que ofrecían los supermercados y tiendas locales.

Para poder lograr el objetivo anterior, la CMA utilizaba las cajas de embarque como estanterías, solicitaba la colaboración de los proveedores para que marcaran sus propios precios en los artículos, vendía solamente las mejores marcas del mercado y generalmente no ofrecía los "lujos" de los supermercados tradicionales.

Las personas que podían gozar de los servicios que ofrecía la CMA tenían que ser socios de las mismas.

La cuota anual para los socios era de \$ 10.000. Todos los beneficios que obtenía la CMA en un año se devolvían a los socios, en forma de créditos para sus compras.

La señorita Vélez estaba convencida de que todos los socios compraban sus alimentos en la CMA.

Durante los primeros años que siguieron a la creación de la CMA, la señorita Vélez se enorgullecía de tener el privilegio de conocer a todos los socios. Había colaborado gran parte del tiempo con el almacén y consideraba que sabía lo que la gente compraba y la cantidad de dinero que estaba gastando.

A medida que creció el número de socios, sus deberes administrativos la mantenían más tiempo en la oficina. Por lo tanto ya no podía decir que conocía a todos los socios, ni que tenía idea de sus patrones de gastos.

Deseaba conocer mejor estos aspectos de su negocio y pensó que, tal vez, alguna información que se había recopilado a este respecto, la podría ayudar a resolver sus dudas.

En Junio de 1989, se utilizó un cuestionario con el fin de recopilar información relacionada con los socios de la cooperativa. Durante ese mes, todos los miembros, sin excepción, visitaron la cooperativa por lo menos una vez. Por lo tanto, logró recopilar información

sobre todos y cada uno de los socios.

La información consistía en características demográficas de los socios, además de incluir sus gastos semanales por comida.

La información se encuentra en las tarjetas que los socios llenarón, en el momento de la entrevista. La señorita Vélez archivó las tarjetas en su oficina.

Con el fin de conocer mejor a los socios, la señorita Vélez quería saber, primero que todo, el gasto promedio semanal por comida. Debido a la premura del tiempo, ella quería lograrlo sin tener que analizar cada una de las 160 tarjetas. también quería que el promedio resultante fuera exacto.

En la siguiente matriz se encuentra, resumida, toda la información que contienen las 160 tarjetas. La columna 1 contiene el código que se le asigno a cada tarjeta.

La columna 2 contiene el gasto semanal por actual de cada socio.

La columna 3 contiene el número de personas que conforman la familia de cada uno de los socios.

La columna 4 contiene el ingreso anual de cada unidad familiar.

La columna 5 contiene el código correspondiente al nivel de educación alcanzado por cada jefe de la unidad familiar.

La columna 6 contiene la edad de cada jefe de la unidad familiar.

La columna 7 contiene la información sobre si existen o no niños por debajo de 6 años, en la unidad familiar.

La columna 8 contiene la información sobre si existen o no personas entre 6 y 18 años, en la unidad familiar.

1	2	3	4	5	6	7	8
001	6.000	1	1.250.000	1	56	no	no
002	8.250	1	1.400.000	1	70	no	no
003	9.000	1	1.000.000	1	20	no	no
004	8.500	1	2.250.000	1	60	no	no
005	23.250	1	4.000.000	1	40	no	no

006	22.500	1	3.500.000	1	51	no	no	
007	7.500	1	1.750.000	2	76	no	no	
008	30.000	2	1.400.000	3	20	no	no	
009	7.500	2	1.250.000	2	51	no	si	
010	9.000	3	2.000.000	1	32	no	no	
011	11.250	2	2.500.000	2	47	si	no	
012	10.000	2	4.000.000	1	35	no	no	
013	48.500	3	2.750.000	4	58	si	no	
014	28.500	3	3.000.000	5	27	si	no	
015	19.500	3	1.500.000	1	38	no	no	
016	15.000	2	2.000.000	2	40	no	si	
017	21.000	2	1.500.000	3	19	no	no	
018	15.000	4	1.500.000	1	50	si	si	
019	15.000	2	3.000.000	5	41	no	no	
020	16.000	3	3.250.000	2	34	si	no	
021	16.500	3	6.500.000	5	23	si	no	
022	15.000	2	7.500.000	1	34	no	no	
023	18.750	2	11.000.000	2	42	no	no	
024	55.000	3	3.300.000	1	43	no	si	
025	26.500	3	3.250.000	2	51	no	si	
026	17.000	3	14.000.000	5	30	no	si	
027	13.500	4	2.000.000	2	46	si	si	
028	19.500	5	6.000.000	1	56	si	si	
029	33.750	7	5.500.000	2	48	si	si	
030	11.250	2	950.000	1	40	no	no	
031	3.750	3	1.400.000	1	20	si	no	
032	9.500	3	1.600.000	3	24	si	no	
033	11.250	4	2.500.000	1	30	si	si	
034	14.250	5	3.500.000	3	33	si	si	
035	17.250	5	6.000.000	2	38	si	si	
036	26.250	6	1.250.000	1	37	si	si	
037	29.000	8	6.000.000	5	42	si	si	
038	3.500	1	1.250.000	1	19	no	no	
039	13.000	1	250.000	1	68	no	no	
040	7.500	1	1.250.000	2	20	no	no	
041	15.000	1	3.000.000	2	34	no	no	
042	12.750	1	3.000.000	2	66	no	no	
043	3.750	1	750.000	1	67	no	no	
044	6.000	1	750.000	1	72	no	no	
045	6.750	2	2.000.000	1	28	no	no	
046	9.000	2	1.250.000	1	20	no	no	
047	15.000	2	1.250.000	2	21	no	no	
048	6.000	2	250.000	1	26	no	no	
049	7.500	2	750.000	2	19	no	no	
050	3.750	2	250.000	3	19	no	no	
051	5.250	2	1.300.000	1	33	no	no	
052	6.000	3	1.250.000	1	36	si	no	
053	26.500	3	1.000.000	2	30	si	no	
054	10.000	4	12.500.000	5	45	si	si	
055	8.000	2	250.000	2	54	no	no	
056	7.500	2	250.000	1	54	no	no	
057	7.500	2	250.000	1	56	no	no	

058	13.000	4	3.100.000	1	26	si	no	
059	7.500	3	1.300.000	3	39	si	no	
060	11.000	3	1.250.000	2	40	no	si	
061	11.250	3	1.400.000	3	48	no	si	
062	12.000	3	1.300.000	3	60	no	si	
063	10.000	2	1.750.000	1	23	no	no	
064	5.500	2	250.000	1	26	no	no	
065	8.250	3	1.500.000	4	64	no	si	
066	7.500	2	250.000	1	70	no	no	
067	7.500	3	1.500.000	3	31	si	no	
068	9.000	2	3.200.000	4	32	si	no	
069	15.000	4	3.500.000	1	44	si	si	
070	15.000	5	4.000.000	2	49	si	si	
071	19.000	2	2.000.000	2	35	no	no	
072	30.000	2	5.000.000	4	22	no	no	
073	12.000	2	1.500.000	1	28	no	no	
074	3.750	2	1.200.000	1	51	no	no	
075	43.500	3	7.000.000	5	69	si	si	
076	15.000	3	3.500.000	4	43	no	si	
077	10.200	3	2.000.000	1	33	si	no	
078	60.000	3	7.000.000	5	39	no	si	
079	10.000	3	4.500.000	2	58	si	no	
080	11.000	3	1.500.000	1	19	no	no	
081	9.000	3	5.000.000	3	33	no	no	
082	20.000	4	3.000.000	2	41	si	si	
083	18.000	3	3.000.000	2	47	no	no	
084	15.000	4	4.500.000	3	48	no	si	
085	19.000	4	1.500.000	1	55	no	si	
086	30.000	4	4.000.000	5	36	si	si	
087	20.000	5	3.500.000	2	57	no	si	
088	38.000	6	3.500.000	2	49	no	si	
089	18.000	6	3.000.000	3	32	no	si	
090	38.000	4	2.800.000	1	32	si	no	
091	25.000	4	3.800.000	3	44	si	si	
092	40.000	7	5.500.000	3	40	si	si	
093	25.000	5	1.800.000	2	29	si	si	
094	20.000	5	3.000.000	1	47	si	si	
095	30.000	9	8.000.000	1	58	si	si	
096	19.000	5	4.000.000	1	51	si	si	
097	40.000	6	7.000.000	3	43	si	si	
098	15.000	6	1.900.000	1	39	si	si	
099	22.000	6	7.000.000	2	39	si	si	
100	36.000	6	2.800.000	1	33	si	si	
101	10.000	7	1.500.000	1	49	si	si	
102	50.000	7	7.000.000	1	53	si	si	
103	15.000	9	1.500.000	2	43	si	si	
104	18.000	9	2.500.000	3	60	si	si	
105	30.000	1	4.500.000	2	45	no	no	
106	21.000	1	5.500.000	5	45	no	no	
107	25.000	1	7.500.000	5	53	no	no	
108	18.000	1	2.000.000	4	42	no	no	
109	17.000	1	2.500.000	4	33	no	no	

110	9.000	1	2.100.000	2	59	no	no		
111	7.500	1	1.300.000	2	21	no	no		
112	10.000	1	3.000.000	2	64	no	no		
113	4.000	1	250.000	2	69	no	no		
114	5.000	1	750.000	2	39	no	no		
115	3.000	1	3.000.000	2	26	no	no		
116	8.000	2	1.000.000	2	39	no	no		
117	33.000	2	9.000.000	5	62	no	si		
118	10.000	3	5.000.000	3	58	si	no		
119	25.000	4	5.500.000	3	38		si		si
120	28.000	2	6.500.000	2	29	no	no		
121	25.000	5	6.500.000	5	60	si	si		
122	26.000	2	6.000.000	2	50	no	no		
123	23.000	7	10.000.000	5	40	si	si		
124	7.800	2	2.100.000	2	26	no	no		
125	21.000	2	1.500.000	3	29		no		no
126	25.000	2	1.500.000	3	29	no	no		
127	7.800	2	1.300.000	1	31	no	no		
128	10.000	4	4.200.000	4	46	si	si		
129	12.000	2	1.800.000	2	30	no	no		
130	15.000	3	3.000.000	2	48	no	si		
131	15.000	2	2.500.000	2	36	no	no		
132	19.000	2	3.000.000	2	30	no	no		
133	9.000	2	3.000.000	2	70	no	no		
134	9.000	2	1.800.000	2	39	no	no		
135	30.000	2	3.500.000	3	44	no	si		
136	25.000	2	3.000.000	2	56	no	no		
137	26.000	3	12.500.000	2	59	no	si		
138	8.000	3	2.200.000	4	29	si	no		
139	20.000	3	1.800.000	3	46	si	no		
140	25.000	3	8.500.000	5	32	si	no		
141	18.000	4	4.500.000	2	73	no	no		
142	9.000	4	2.500.000	2	40	no	no		
143	18.000	2	1.400.000	2	59	no	si		
144	20.000	2	1.800.000	3	49	no	si		
145	16.000	3	1.300.000	2	47	si	no		
146	20.000	3	1.900.000	1	52	no	si		
147	22.000	3	3.000.000	3	48	no	si		
148	22.000	3	3.000.000	4	56	no	si		
149	15.000	3	5.000.000	3	24	si	no		
150	34.000	4	4.500.000	5	51	si	si		
151	18.000	4	4.000.000	2	38	no	si		
152	22.000	4	3.000.000	2	47	si	si		
153	26.000	4	4.800.000	3	58	si	si		
154	38.000	5	6.000.000	5	50	no	si		
155	15.000	5	3.200.000	1	49	no	si		
156	20.000	5	3.000.000	3	61	si	si		
157	18.000	5	3.000.000	2	53	no	si		
158	25.000	5	3.500.000	4	42	si	si		
159	22.000	5	4.500.000	5	39	no	si		
160	11.000	6	6.000.000	2	48	si	si		

La columna 5 contiene los códigos que se le asignó a cada nivel de educación. Las alternativas del nivel de educación y el código asignado a cada una se presentan a continuación:

CODIGO	ALTERNATIVAS
1	Por debajo de segundo de bachillerato.
2	Entre tercero y quinto de bachillerato.
3	Bachiller.
4	Algunos años de universidad.
5	Universitario.

ANEXO 3  
DISTRIBUCION Z



ANEXO 4  
DISTRIBUCION T

## ANEXO 5

## TABLA 8 - 6

TOMADA DEL LIBRO: RICHARDS, LARRY E. Y JERRY J.LACAVA.  
ESTADISTICA EN LOS NEGOCIOS: POR QUE Y CUANDO ?.  
BOGOTA, ED. McGRAW - HILL, 1980. PAGINAS 174 y 175.

