Capitulo tres

Procedimientos de muestreo

1. PLAN DE MUESTREO

Todo estudio de mercado requiere de recolección de información. Ahora bien, dependiendo del objetivo u objetivos que se persiguen, puede realizarse una investigación exhaustiva o una investigación parcial.

En la investigación exhaustiva se deben observar todas las unidades que constituyen la población o universo. La enumeración total de una población en un tiempo dado recibe el nombre de CENSO. El censo requiere de una organización compleja y la ocupación de un gran número de personas en cada una de las diferentes etapas de investigación.

Desventajas de realizar un censo

- Se requiere demasiado tiempo para realizarlo.
- Los costos son elevados en las etapas de planeamiento, sistematización y análisis y publicación de la información.
- Tiene una alta probabilidad de cometerse el error sistemático.
- En algunos casos no se podrá obtener información especializada, por ejemplo, no se podrá dar pruebas de degustación a toda una población.
- Los estudiantes carecen de tiempo y recursos para realizar una investigación exhaustiva.

Las anteriores consideraciones obligan a buscar un método práctico para obtener información, en especial para poblaciones grandes. Este método consiste en realizar un estudio parcial o tomar una muestra de la población total.

Ventajas de realizar una muestra

- Reduce los costos y el tiempo
- Es aplicable en poblaciones infinitas y muy grandes

Al hablar de muestreo nos referimos al conjunto de técnicas estadísticas que estudian la forma de seleccionar una muestra lo suficientemente representativa de una población cuya información permita inferir las propiedades o características de toda la población cometiendo un error medible o contable. A partir de la muestra, seleccionada mediante un determinado método de muestreo, se estiman las características poblacionales (media, total, proporción, etc.), estas estimaciones se realizan a través de funciones matemáticas de la muestra denominadas estadísticos, que se convierten en variables aleatorias al considerar la variabilidad de las muestras. Los errores se cuantifican mediante varianzas, desviaciones típicas o errores cuadráticos medios de los estimadores, que miden la precisión de éstos. La metodología que permite inferir resultados, predicciones y generalizaciones sobre la

población estadística, basándose en la información contenida en las muestras representativas previamente elegidas por métodos de muestreo formales, se denomina inferencia estadística.

A continuación se dan algunos conceptos generales de muestreo, se parte del supuesto de que los estudiantes tienen alguna formación en estadística básica para abordar el tema.

3.1. ELEMENTO.

Es la unidad acerca de la cual se solicita información. Éste suministra la base del análisis que se llevará a cabo. Los elementos más comunes del muestro en investigación de mercados son los individuos. En otros casos, los elementos podrían ser productos, almacenes, empresas, familias, etc. Los elementos de cualquier muestra específica dependerán de los objetivos del estudio.

3.2. POBLACIÓN.

Una población o universo, como también se llama, es el conjunto de todos los elementos definidos antes de selección de la muestra. Una población adecuadamente designada debe definirse en términos de: 1) elementos, 2) unidades de muestreo, 3) alcance y 4) tiempo: Por ejemplo, una encuesta de consumidores debe especificar la población pertinente de la siguiente manera:

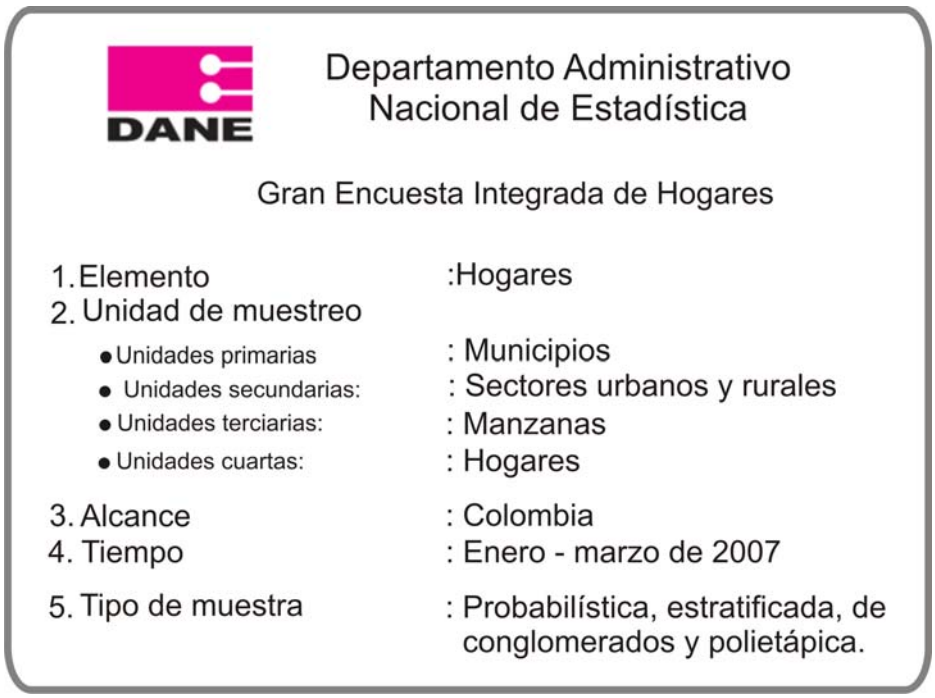
1. Elemento	: Mujeres entre 18 – 50 años
2. Unidades de muestreo	: Mujeres entre 18 – 50 años
3. Alcance	: Popayán
4. Tiempo	: 1 de mayo al 15 de junio de 2000

3.3. UNIDAD DE MUESTREO.

Es el elemento o los elementos disponibles para su selección en alguna etapa del proceso de muestreo. El ejemplo anterior es un tipo de muestreo simple, es decir, de una sola etapa, en este caso las unidades y los elementos de muestro son los mismos.

En procedimientos de muestreo más complejos pueden utilizarse diferentes niveles de unidades de muestreo, entonces las unidades de muestreo y los elementos difieren en todo, menos en la última etapa. Veamos el siguiente ejemplo:

La encuesta continúa de hogares realizada por el DANE, es Polietápica porque para la selección de los hogares objeto de investigación se seleccionan secuencialmente las unidades de muestreo.



3.4. MARCO MUESTRAL.

Es una lista de todas las unidades de muestreo disponibles para su selección en una etapa del proceso de muestreo. Uno de los procedimientos más creativos en un proyecto de investigación de mercados puede relacionarse con la especificación de un marco muestral.

Un marco puede ser una lista de alumnos, una lista de votantes inscritos, un directorio telefónico, una lista de empleados o incluso un mapa.

3.5. POBLACIÓN DE ESTUDIO.

La población de estudio es el conjunto de elementos del cual se saca la muestra. Existe varios tipos de muestreo, dependiendo si la población es finita o infinita, materia sobre la que existe amplia literatura estadística. En éste texto se considera solamente el muestreo en poblaciones finitas.

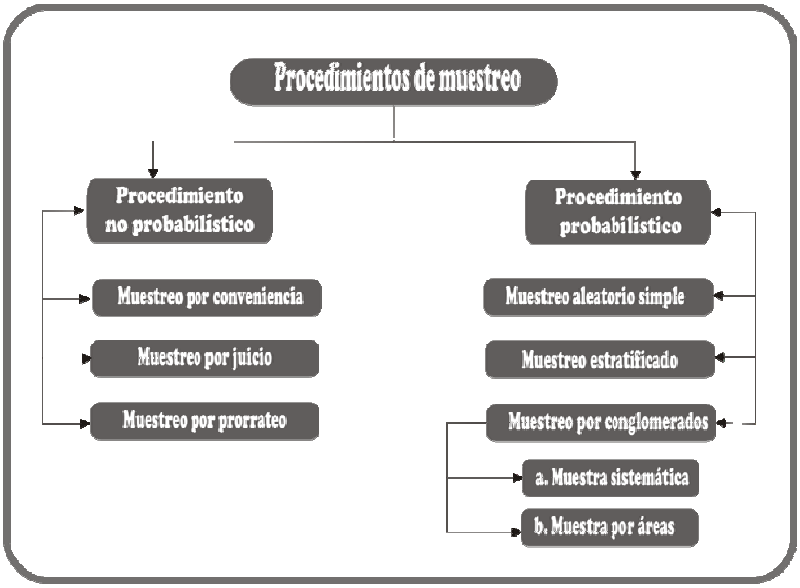
3.6. PROCEDIMIENTOS DE MUESTREO

Como se mencionó existen muchos procedimientos mediante los cuales los investigares pueden seleccionar sus muestras, pero se debe establecer la diferencia entre una muestra probabilística y una muestra no probabilística.

3.6.1. Procedimiento Probabilístico. Cada elemento de la población tiene una probabilidad conocida de ser seleccionado para la muestra. Una muestra probabilística permite calcular el grado probable hasta el cual el valor de la muestra puede diferir del valor de la población de interés. Esta diferencia recibe el nombre de **error muestral.**

3.6.2. Procedimiento no Probabilístico. La selección de un elemento de la población que va a formar parte de la muestra se basa hasta cierto punto en el criterio del investigador o entrevistador de campo. No existe una posibilidad conocida para seleccionar cualquier elemento particular de la población.

Figura 3.1. Procedimientos de Muestreo



3.6.3. Parámetro. Un parámetro es una descripción de una medida de la población bajo estudio. Ejemplos: Edad promedio de los estudiantes, Ingreso promedio, etc.

3.6.4. Estadístico. Un estadístico es una descripción resumida de una medida en la muestra seleccionada. Así, la edad promedio de los estudiantes sería un estadístico si se mide a través de una muestra**.**

3.7. NOMENCLATURA UTILIZADA.

La nomenclatura que se utilizará será la siguiente:

Tabla 3.1. Nomenclatura

Parámetros / Estadísticos	Símbolo de Población	Símbolo Muestral
Media o promedio	μ	$\overline{X}$
Varianza	σ^2	S^2
Probabilidad de ocurrencia del fenómeno de estudio.	π	p
Probabilidad de no ocurrencia del fenómeno de estudios	$(1-\pi)$	(1-p) o q

3.8. DISTRIBUCIONES MUESTRALES.

Consideramos los sucesos elementales asociados a un fenómeno o experimento aleatorio dado S_1, S_2, …, S_m, entendiendo por suceso elemental los más simples posibles, es decir, aquellos que no pueden ser descompuestos en otros sucesos. El conjunto $\{S_1, S_2, \ldots, S_m\}$ se denomina espacio muestral asociado al fenómeno o experimento. Si consideramos como fenómeno o experimento la extracción aleatoria de muestras dentro de una población por un procedimiento o método de muestreo dado, podemos considerar como sucesos elementales las muestras obtenidas, constituyendo el conjunto de las mismas el espacio muestral.

Habitualmente en los métodos de muestreo se consideran iguales muestras con los mimos elementos, aunque estén colocados en orden diferente (el orden de colocación no interviene), es decir, cuando la selección se hace con reposición. Una muestra de tamaño *n* extraída de una población $U=\{U_1, U_2, \ldots, U_N\}$ de tamaño N mediante un método de muestreo dado, suele denotarse como $s=\{u_1, u_2, \ldots, u_n\}$. De esta forma, El conjunto de las N^n muestras posibles de tamaño s que se pueden formar con los elementos de la población U es el espacio muestral S.

Cuando el método de muestreo se realiza sin reposición el conjunto de las $\frac{N!}{(N-n)!n!}$ muestras posibles de tamaño n que se pueden formar con los elementos de la población U es el espacio muestral S.

Evidentemente, para establecer la probabilidad de todas las muestras posibles derivadas de un procedimiento de muestreo dado, será necesario conocer ese conjunto de muestras, es decir, será necesario delimitar tanto el método de muestreo como el especio muestral

derivado del mismo, por tanto el método aleatorio empleado para seleccionar la muestra define en el espacio muestral S una función de probabilidad P tal que:

$$P(S_i) \geq 0 \forall i$$
$$\sum_s P(S_i) = 1 \qquad (1)$$

Ahora consideremos una población de N elementos, con media μ y desviación típica σ, si se obtiene M número de muestras posibles, de tamaño n, simbolizamos a cada media muestral por $\bar{x}_1; \bar{x}_2; \bar{x}_3 ..... \bar{x}_M$ y cada desviación típica muestral por: s_1; s_2; s_3 ……s_m

Teorema. Dada una población, si extraemos todas las muestras posibles de un mismo tamaño, entonces las medias de la distribución de todas las medias muestrales posibles será igual a la media poblacional.

$$\mu_x = \frac{\Sigma \overline{X}_j}{M} = \frac{\overline{X}_1 + \overline{X}_2 + \overline{X}_3 + ... + \overline{X}_M}{M} = \mu \qquad (2)$$

La varianza de todas las medias muestrales se simboliza por: $\sigma^2_{\bar{x}}$
El error estándar será simbolizado por: $\sigma_{\bar{x}}$

$$\sigma_x = \sqrt{\frac{\sum \left(\overline{X}_j - \mu\right)^2}{M}} = \sqrt{\frac{\left(\overline{X}_1 - \mu\right)^2 + \left(\overline{X}_2 - \mu\right)^2 + ... + \left(\overline{X}_M - \mu\right)^2}{M}} \qquad (3)$$

Siendo $\sigma_x = \frac{\sigma}{\sqrt{n}}$ para muestras grandes o sea n>30 y se denomina: error estándar de la media.

La media de todas las medias muestrales debe ser exactamente igual a la media poblacional (μ), debido a que la distribución de muestreo resulta de todas las muestras posibles que se pueden extraer de una población; por tal razón incluye a todos sus elementos.

Expliquemos lo anterior mediante un ejemplo. Supongamos una población de 5 elementos (N=5), los valores que toman las variables son:

X_1=2 X_2=4 X_3=6 X_4=8 X_5=10

Con los anteriores valores se calcula la media poblacional, la varianza y la desviación típica poblacional así:

$$\mu = \frac{\sum X_i}{N} = \frac{2+4+6+8+10}{5} = \frac{30}{5} = 6 \qquad (4)$$

$$\sigma^2 = \frac{(2-6)^2 + (4-6)^2 + (6-6)^2 + (8-6)^2 + (10-6)^2}{5} = \frac{40}{5} = 8 \quad (5)$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{8} = 2{,}83 \quad (6)$$

Ahora determinamos el número de muestras posibles (M) de esta población, si el tamaño de la muestra que fijamos es de 2 y la selección se hace sin reposición se tiene:

$$M = c_n^N = \frac{N!}{(N-n)!n!} = \frac{5!}{(5-2)!2!} = 10 \quad (7)$$

Las combinaciones que se pueden obtener son las siguientes:

X_1X_2	X_2X_3	X_3X_4	X_4X_5
X_1X_3	X_2X_4	X_3X_5	
X_1X_4	X_2X_5		
X_1X_5			

Calculamos la media aritmética para cada una de las muestras:

$$\bar{x}_1 = \frac{x_1 + x_2}{n} = \frac{2+4}{2} = 3 \qquad \bar{x}_6 = \frac{x_2 + x_4}{n} = \frac{4+8}{2} = 6$$

$$\bar{x}_2 = \frac{x_1 + x_3}{n} = \frac{2+6}{2} = 4 \qquad \bar{x}_7 = \frac{x_2 + x_5}{n} = \frac{4+10}{2} = 7$$

$$\bar{x}_3 = \frac{x_1 + x_4}{n} = \frac{2+8}{2} = 5 \qquad \bar{x}_8 = \frac{x_3 + x_4}{n} = \frac{6+8}{2} = 7$$

$$\bar{x}_4 = \frac{x_1 + x_5}{n} = \frac{2+10}{2} = 6 \qquad \bar{x}_9 = \frac{x_3 + x_5}{n} = \frac{6+10}{2} = 8$$

$$\bar{x}_5 = \frac{x_2 + x_3}{n} = \frac{4+6}{2} = 5 \qquad \bar{x}_{10} = \frac{x_4 + x_5}{n} = \frac{8+10}{2} = 9$$

La media de las medias muestrales será igual a:

$$\mu_{\bar{x}} = \frac{\sum \bar{x}_i}{M} = \frac{3+4+5+6+5+6+7+7+8+9}{10} = 6 \text{, es decir, } \mu_{\bar{x}} = \frac{\sum \bar{x}_i}{M} = \mu$$

La desviación típica de todas las medias muestrales, se calcula a continuación

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum(\bar{x}_i - \mu)^2}{M}}$$

$$\sigma_{\bar{x}} = \sqrt{\frac{(3-6)^2 + (4-6)^2 + (5-6)^2 + (6-6)^2 + (5-6)^2 + (6-6)^2 + (7-6)^2 + (7-6)^2 + (8-6)^2 + (9-6)^2}{10}} = \sqrt{\frac{30}{10}} = \sqrt{3} = 1.73$$

Lo anterior debe ser igual a $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, esta expresión se denomina error estándar de la media, lo cual se cumple para muestras grades, algunos autores consideran que una muestra es grande cuando n>30, por esta razón los resultados difieren:

Con los datos poblacionales tenemos:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.8384}{1.4142} = 2 \neq 1{,}73$$

3.9. TEOREMA DEL LÍMITE CENTRAL.

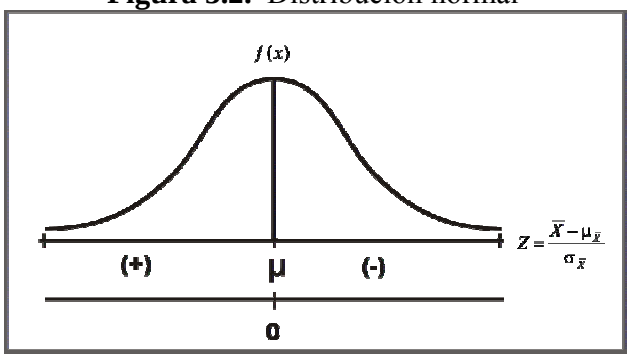
Se cumple cuando independientemente de la población origen la distribución de las medias aleatorias se aproxima a una distribución normal a medida que el tamaño de la muestra crece. Se podrá decir también que si las muestras provienen de una población que no es normal. Si el tamaño de la muestra es pequeño, la distribución obtenida de sus medias muestrales tendrá un comportamiento similar al de la población de donde se extrajeron. Por el contrario, si el tamaño muestral es grande, el comportamiento de estas medias muestrales será igual al de una distribución normal independientemente de la población de donde fueron extraídas.

En su forma más simple el teorema indica que si **n** variables aleatorias independientes tienen varianzas finitas, su suma, cuando se le expresa en medida estándar, tienden a estar normalmente distribuidas cuando **n** tiende a infinito.

De acuerdo al teorema, la varianza estadística para distribuciones de medias muestrales será:

$$Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Figura 3.2. Distribución normal

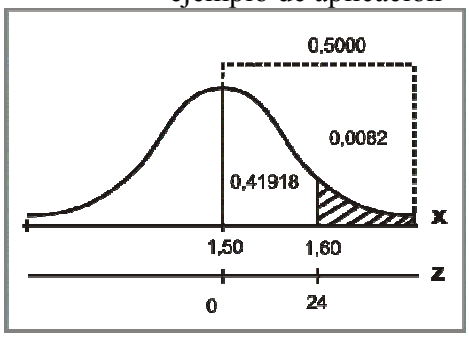


Es decir, se aproxima a una distribución normal

Ejemplo 1[1].

La altura media de 400 alumnos de un plantel de secundaria es de 1,50 m. y su desviación típica es de 0,25 m. Determinar la probabilidad de que en una muestra de 36 alumnos, la media sea suprior a 1,60 m.

Figura 3.3. Distribución normal ejemplo de aplicación



$$P(\overline{X} \geq 1{,}60) = ?$$

$$Z = \frac{1{,}60 - 1{,}50}{\frac{0{,}25}{\sqrt{36}}} = \frac{0{,}10}{\frac{0{,}25}{6}} = \frac{0{,}60}{0{,}25} = 2{,}40$$

$$Z = 2{,}40 \rightarrow A(0{,}4918)$$

$$P = 0{,}5000 - 0{,}4918 = 0{,}0082 = 0{,}82\%$$

Ejercicio para realizar en clase

- Se tiene para la venta un lote de 1.000 pollos, con un peso promedio de 3,50 Kg. y una desviación estándar de 0,18 Kg., ¿Cuál es la probabilidad de que en una muestra aleatoria, 100 pollos de esta población, pesen entre 3,53 y 3,54 Kg.?

- Un fabricante de cierto champú para el cabello distribuye el tamaño profesional de su producto en 100 salones de belleza. Se ha determinado que el consumo promedio de su producto es de 2.800 cojines mensuales, con desviación estándar de 280 cojines. Si se toma una muestra probabilística de 36 salones, ¿cuál es la probabilidad de que el consumo promedio en un mes sea inferior a 2.700?

3.10. FACTOR DE CORRECCIÓN PARA POBLACIONES FINITAS.

En aquellos casos de poblaciones finitas, es decir, cuando se da información sobre el tamaño poblacional y cuando el tamaño de la muestra es mayor del 5% de la población, se puede aplicar el factor de corrección, representado de diferentes maneras; cualquiera de estas formas podrá ser aplicada:

[1] Martínez Bernandino. Ciro, Estadística y muestreo, 11ª ed., Ecoe Ediciones, Bogotá, D.C. 2002, p 321 y 322

$$\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{N-n}{N}} = \sqrt{\frac{N}{N} - \frac{n}{N}} = \sqrt{1-\frac{n}{N}} = \sqrt{1-f} \; ...donde.... f = \frac{n}{N} \qquad (8)$$

En distribuciones de medias muestrales, la estandarización de Z, incluyendo el factor de corrección será:

$$Z = \frac{\overline{X} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)\sqrt{\frac{N-n}{N}}} \qquad (9)$$

3.11. DISTRIBUCIÓN MUESTRAL DE UNA PROPORCIÓN.

En el análisis de una característica cualitativa o atributo se emplea la proporción de la ocurrencia (P) o éxito y no ocurrencias (Q) del fenómeno de estudio, siguiendo una distribución binomial.

Se define la proporción de éxitos o ocurrencia del fenómeno de estudio como:

$$P = \frac{Numero \;\; de \;\; casos \;\; favorables \;\; o \;\; exitos}{tatal \;\; de \;\; casos \;\; posibles} = \frac{\sum a_j}{n} \qquad (10)$$

La nomenclatura que se utilizará es la siguiente:

$A = \sum A_j = NP$ Total de elementos que presentan la característica en la población

$P = \frac{A}{N} = \frac{\sum A_j}{N}$ Proporción de elementos que presentan la ocurrencia del fenómeno de estudio (éxito).

$Q = \frac{N-A}{N} = 1 - P$ Proporción de elementos que no presentan la ocurrencia del fenómeno estudio.

Entonces: $P + Q = 1$ (11)

Varianza de la proporción en la población $\sigma^2{}_p = PQ$ (12)

Desviación estándar $\sigma_P = \sqrt{PQ}$ (13)

$\sigma_{\overline{P}} = \frac{\sigma_P}{\sqrt{n}} = \sqrt{\frac{PQ}{n}}$ = Error estándar de la proporción (14)

Ejemplo 2[2]:

[2] Ibíd., p. 332.

Se tiene que el 4% de las piezas producidas por cierta máquina son defectuosas, ¿cuál es la probabilidad de que en un grupo de 200 piezas, el 3% o más sean defectuosas?

Solución:

$$\mu_P = P = 0{,}04 \rightarrow \overline{P} = P = 0.03$$

$$\sigma_{\overline{P}} = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{(0{,}04)(0{,}96)}{200}} = 0{,}014$$

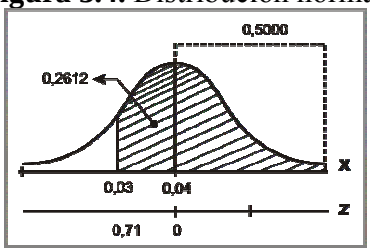
Se desea determinar la P(p ≥ 0,03)=?

$$Z = \frac{p - \mu_p}{\sqrt{\frac{PQ}{n}}} = \frac{0{,}03 - 0{,}04}{\sqrt{\frac{(0{,}04)(0{,}96)}{200}}} = -0{,}71$$

$$Z = -0{,}71 \rightarrow A(0{,}2612)$$

$$P = 0{,}2612 + 0{,}5000 = 0{,}7612 = 76{,}12\%$$

Figura 3.4. Distribución normal



Ejercicios para resolver en clase

- Se desea estudiar una muestra de 49 personas para saber la proporción de las mayores de 40 años; sabiendo que la proporción en la población es 0,4, ¿cuál es la probabilidad de que la proporción en la muestra sea menor de 0,5?

- Cuarenta y seis por ciento de los sindicatos del país están en contra de comerciar con la China Continental; ¿cuál es la probabilidad de que una encuesta a 100 sindicatos muestre que más del 52% tenga la misma posición?

3.12. TAMAÑO DE LA MUESTRA.

Para determinar **n** el tamaño de la muestra es necesario identificar los siguientes componentes o elementos técnicos:

3.12.1. La Varianza (σ^2_X)**.** En el caso de variables discretas (σ^2_X)=PQ.

3.12.2. El Nivel de Confianza. Tiene relación directa con el tamaño de la muestra, por lo tanto se dirá que a mayor nivel de confianza más grande debe ser el tamaño de la muestra. El nivel es fijado por el investigador, de acuerdo a su experiencia.

3.12.3. Precisión de la Estimación. Corresponde al margen de error que el investigador fija de acuerdo al conocimiento que tenga acerca del parámetro que piensa estimar. Se le conoce como error de muestreo (E), siendo:

$$E = Z\frac{\sigma}{\sqrt{n}} \qquad E = Z\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N}} \tag{15}$$

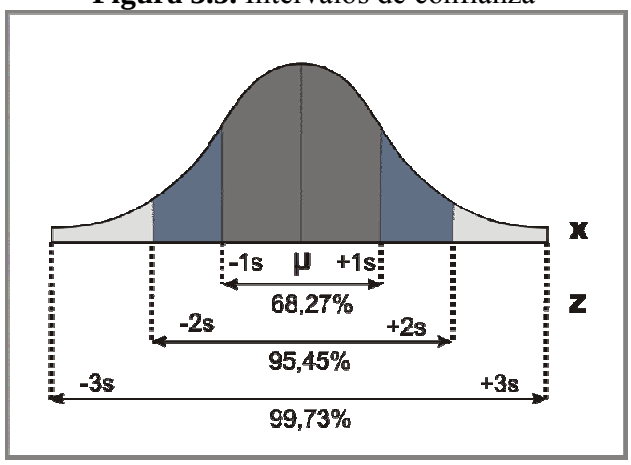
3.13. ESTIMACIÓN DE INTERVALOS DE CONFIANZA PARA PARÁMETROS DE POBLACIÓN.

Sean μ_s y σ_s la media y la desviación (error típico) de la distribución de muestreo de un estadístico S. Entonces si la distribución de muestreo de S es aproximadamente normal (que como hemos visto es cierto para muchos estadísticos si el tamaño de la muestra n ≥ 30), se espera hallar un estadístico real S que esté en los intervalos.

$$\left[\mu_s - \sigma_s ; \mu_s + \sigma_s\right] \;;\; \left[\mu_s - 2\sigma_s ; \mu_s + 2\sigma_s\right] \;;\; \left[\mu_s - 3\sigma_s ; \mu_s + 3\sigma_s\right]$$

Aproximadamente del 68.7%, 95,45% y 99,73% respectivamente.

Figura 3.5. Intervalos de confianza



3.14. CÁLCULO DE LOS COEFICIENTES DE CONFIANZA.

$S \pm 1.96\sigma_s$ son los límites de confianza 95% para S. El porcentaje de confianza suele llamarse nivel de confianza. Los números 1.96, 2.58, etc. en los límites de confianza se llaman ***coeficientes de confianza o valores críticos*** y se denota por Z. De los niveles de confianza podemos deducir los coeficientes de confianza y viceversa.

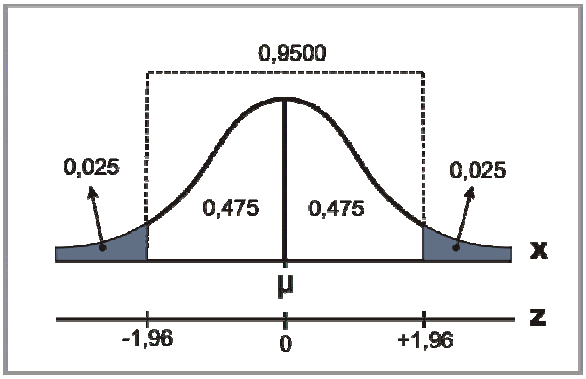
Para el calcular el Coeficiente de confianza para un nivel de confianza del 95%, haga el siguiente procedimiento:

- 1-0,95=0,05

- 0,05/2=0,025 (Ésta operación se realiza por la taba de distribución de frecuencia utilizada).

- 0,500 - 0,025=0,475 Se busca en la tabla de distribución normal (Apéndice B) el área 0,475 el valor de Z $\rightarrow$ le corresponde = 1.96

Figura 3.6. Distribución normal



En la Tabla 2 se dan algunos cálculos de los coeficientes de confianza más utilizados, si el estudiante requiere de algún coeficiente que no este especificado en la tabla, lo puede calcular haciendo uso de la tabla de distribución normal (Apéndice B) y de los procedimientos dados en la sección anterior.

Tabla 3.2. Niveles de confianza.

Nivel de confianza	99.73%	99%	98%	96%	95.45%	95%	90%	80%	68.27%	50%
Z	3.00	2.58	2.33	2.05	2.00	1.96	1.645	1.28	1.00	0.6745

3.13. TEORÍA ESTADÍSTICA DE LAS DECISIONES

A menudo no vemos obligados a tomar decisiones relativas a una población sobre la base de la información provenientes de muestras. Tales decisiones se llaman decisiones estadísticas. Por ejemplo, deseamos decidir basados en datos muestrales que producir el producto A es mejor que producir el producto B, o si la aceptación del producto C es lo suficientemente alta para invertir en determinado proyecto.

Al tomar una decisión, es útil hacer conjeturas sobre la hipótesis implicada. Las hipótesis estadísticas es un enunciado acerca de las distribuciones de probabilidad de las poblaciones.

3.13.1. Hipótesis nula. En muchos casos formulamos hipótesis estadística con el único propósito de rechazarla o invalidarla. Así, si queremos decidir si una producto es aceptado, formulamos la hipótesis de que este tiene aceptación (o sea p=0,5, donde p es la probabilidad de que definitivamente o probablemente lo compren). Este enunciado o conjetura se denomina hipótesis nula y se denotará por H_0.

3.13.2. Hipótesis alternativa. Toda hipótesis que difiera de una dada (hipótesis nula H_0) se llamará hipótesis alternativa, si una hipótesis es p=0,5, la hipótesis alternativa podría ser

p=0,7, es decir un p≠0,5 o p>0,5. Una hipótesis alternativa a la hipótesis nula se denotará por H_1.

3.13.3. Contrastes de hipótesis y significación, o reglas de decisión. Supongamos que vemos que los resultados hallados en una muestra aleatoria difieren notablemente de los esperados bajo la hipótesis (o sea, esperado sobre la base del azar, por teoría de muestreo), entonces diremos que las diferencias observadas son significativas y tendríamos que rechazar la hipótesis (o al menos no aceptarla ante la evidencia obtenida).

Los procedimientos que nos capacitan para determinar si las muestras observadas difieren significativamente de los resultados esperados, y por tanto nos ayudan a decidir si aceptamos o rechazamos las hipótesis planteadas, se llama contrastes (o tests) de hipótesis o de significación o reglas de decisión.

3.13.4. Errores de tipo I y de tipo II. Si rechazamos una hipótesis cuando debe ser aceptada, diremos que se ha cometido un error de Tipo I. De otra parte, si aceptamos una hipótesis que debe ser rechazada, diremos que se ha cometido un error de tipo II. Es decir, que ambos casos se ha producido un juicio erróneo.

Tabla 3.3. Tipos de error

	Aceptada	Rechazada
Hipótesis verdadera	Juicio acertado	Error de tipo I
Hipótesis falsa	Error de tipo II	Juicio acerado

Las reglas de decisión (o contraste de hipótesis) deben diseñarse de modo que minimicen los errores de la decisión. Esto no es una cuestión fácil, porque para cualquier tamaño de la muestra, un intento de disminuir un tipo de error suele ir acompañado de un crecimiento del otro tipo. La única forma de disminuir ambos es aumentar el tamaño de la muestra.

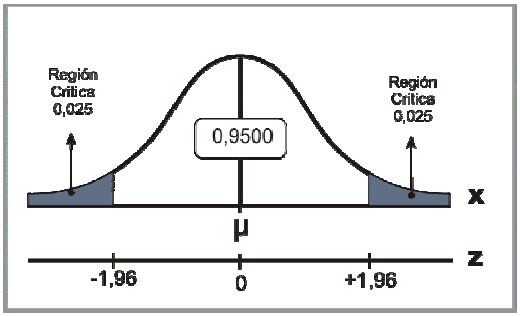
3.13.5. Nivel de significación. Al contrastar una hipótesis, la máxima probabilidad con la que estamos corriendo el riesgo de cometer un error de tipo I se llama nivel de significación del contraste. Esta probabilidad, denotada a menudo por α, se suele especificar antes de tomar la muestra.

En la práctica, es frecuente un nivel de significación de 0,05 ó 0.01, si bien se usan otros valores Si, por ejemplo, se escoge el nivel de significación 0,05 (o 5%) al diseñar una regla de decisión, entonces diremos que entre 100 muestras hay 5 oportunidades de rechazar la hipótesis, es decir, tenemos un 95% de confianza de que hemos tomado una decisión acertada. Dicho de otra forma, quiere decir, que la hipótesis tiene una probabilidad de 0.05 de ser falsa.

3.13.6. Contraste mediante la distribución normal. Para ilustrar los conceptos dados, supongamos que bajo cierta hipótesis la distribución de muestreo de un estadístico S es una distribución normal con media μ y desviación típica σ, entonces la distribución de la variable tipificada z, estandarizada está dada por:

$$Z = \frac{S - \mu}{\sigma} \quad dónde \quad (0,1) \quad \mu = 0 \quad y \quad \sigma^2 = 1$$

Figura 3.7. Contraste de hipótesis



Como se ve en la Figura 3.7, se puede tener 95% de confianza de que si la hipótesis es verdadera, entonces el valor de z para un estadístico muestral S estará entre -1,96 y 1.96 (porque el área bajo la curva normal entre esos valores es 0,95). Si al escoger una sola muestra al azar hallamos que el valor de z de sus estadístico está fuera de ese rango, debemos concluir que tal suceso puede ocurrir con una probabilidad de 0,05 (el área sombreada en la figura 3.7) si la hipótesis dada fuera cierta, entonces se puede afirmar que este z difiere de forma significativa de lo que se espera bajo la hipótesis y tendríamos que rechazar la hipótesis.

El área total sombreada 0,05 es el nivel de significación del contraste. Representa la probabilidad de equivocarnos al rechazar la hipótesis (o sea la probabilidad de cometer un error de tipo I).

El conjunto de z fuera del rango -1.96 a 1.96 se llama región crítica de la hipótesis región de rechazo de la hipótesis, o región de significación. El conjunto de z en el rango -1.96 a 1.96 se conoce como región de aceptación de la hipótesis o región de no significación.

Basado en lo anterior, se puede formular la siguiente regla de decisión o contrastes de hipótesis o significación:

Rechazar la hipótesis al nivel de significación 0,05 si el valor de z para el estadístico S está fuera del rango -1.96 a 1.96. Esto equivale a decir que el estadístico muestral observado es significativo al nivel 0,05.

3.13.7. Curva de operación características. Potencia de un contraste. Es posible evitar el riesgo de cometer un error de Tipo II, recurriendo a las curvas de operación características, o curvas OC, que son gráficos que muestran las probabilidades de error de Tipo II bajo

diversas hipótesis, es decir, nos indica la potencia de un test a la hora de prevenir decisiones erróneas. Son útiles en el diseño de muestreo porque sugieren entre otras cosas el tamaño de muestra a calcular.
Mediante un ejemplo vamos a explicar la elaboración de estos gráficos.

Ejemplo 3[3]

Para contrastar la hipótesis de que una moneda es buena, adoptamos la siguiente regla de decisión:

H_0: Aceptarla si el número de caras en una sola muestra de 100 tiradas está entre 40 y 60 inclusive.
H_1: Rechazar en caso contrario.

Como Np=100(½) y Nq(½), la aproximación normal a la distribución binomial es correcta a la hora de evaluar la suma. La media y la desviación típica de número de caras en 100 tiradas son

$$\mu = Np = 100\left[\frac{1}{2}\right] = 50$$

$$\sigma = \sqrt{Npq} = \sqrt{(100)(½)(½)} = 5$$

En una escala continua, decir entre 40 y 60 inclusive es como decir entre 39,5 a 60.5 caras, luego:

$$39{,}5 \;\; en \;\; unidades \;\; estándar = Z = \frac{S-\mu}{\sigma} = \frac{39{,}5-50}{5} = -2{,}10$$

$$60.5 \;\; en \;\; unidades \;\; estándar = Z = \frac{S-\mu}{\sigma} = \frac{60{,}5-50}{5} = 2{,}10$$

La probabilidad es igual al área bajo la curva normal entre Z=-2,1 y Z=2.1

2(área entre z= 0 y z=2) (Ver Apéndice B. Tabla de distribución normal)
= 2(0,4821)=0.9642

b) La probabilidad de no obtener entre 40 y 60 caras inclusive si la moneda es buena, es 1-0.9642 = 0,0358. Luego la probabilidad de rechazar la hipótesis cuando sea correcta es 0.0358.

En la tabla 6.1 muestra los valores correspondientes a valores dados de p, es decir la probabilidad de aceptar la hipótesis y rechazarla, como se puede observar el punto ideal es cuando p=0,5.

Tabla 3.4. Probabilidades de cometer un error de tipo I y II

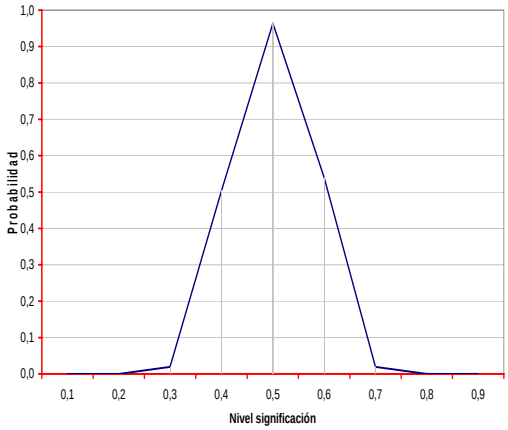
[3] Murray R. Spiegel. Estadística de Schaum. Segunda edición. Edit McGrawHill. pagina 235.

p	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
Prob.	0,0000	0,0000	0,0192	0,5040	0,9642	0,5400	0,0192	0,0000	0,0000
1-Prob	1,0000	1,0000	0,9808	0,4960	0,0358	0,4600	0,9808	1,0000	1,0000

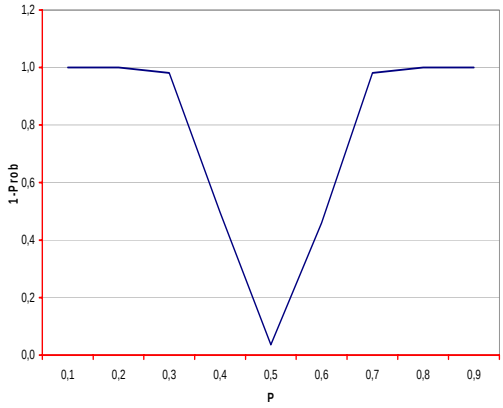
Graficando los valores de la taba 3.4 se obtienen las figuras 3.8 (a) y 3.8 (b).

Figura 3.8.

(a) Curva de operación característica Curva OC



(b) Curva de Potencia de la regla de decisión



La grafica 3.8. (a) se llama la curva de operación característica, o curva OC, de la regla de decisión (o contraste de hipótesis), en general lo que se observa es que cuanto más agudo es el pico de la curva OC, mejor es la regla de decisión a la hora de rechazar hipótesis incorrectas.

La Figura 3.8 (b) se llama la curva de potencia de la regla de decisión se obtiene sin más que invertir la curva OC, luego ambas Figuras son equivalentes, esta curva indica la potencia de un test (o contraste) para rechazar hipótesis falsas.

3.14. TAMAÑO OPTIMO EN POBLACIONES FINITAS.

Partiendo de la ecuación (9)

$$Z = \frac{\overline{X} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)\sqrt{\frac{N-n}{N}}}$$

Se considera la diferencia entre $\overline{X} - \mu$, como el error máximo permitido en la muestra y se reemplaza por E, hecho este cambio se despeja la ecuación así:

$$Z = \frac{E}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N}}} \quad ; \quad E = \frac{Z\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N}} = E^2 = \left(\frac{Z^2\sigma^2}{n}\right)\left(\frac{N-n}{N}\right)$$

$$E^2 = \frac{Z^2\sigma^2 N - Z^2\sigma^2 n}{Nn}$$

$$E^2 Nn = Z^2\sigma^2 N - Z^2\sigma^2 n$$

$$E^2 Nn + Z^2\sigma^2 n = Z^2\sigma^2 N$$

$$n(NE^2 + Z^2\sigma^2) = Z^2\sigma^2 N$$

$$n = \frac{Z^2 N\sigma^2}{NE^2 + Z^2\sigma^2}$$

La formula que se utilizará para el cálculo de la muestra es:

$$n = \frac{Z^2 N\sigma^2}{NE^2 + Z^2\sigma^2}$$

Realizando algunas manipulaciones algebraicas tenemos las siguientes formulas que son muy utilizadas también:

$$n = \frac{\sigma^2}{\left(\frac{E}{Z}\right)^2 + \frac{\sigma^2}{N}} \tag{16}$$

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \quad donde \quad n_0 = \frac{Z^2\sigma^2}{E^2} = \left(\frac{Z\sigma}{E}\right)^2$$

3.14.1. Formulas utilizadas con variables discretas (atributos).

$$n = \frac{Z^2 NPQ}{(N-1)E^2 + Z^2 PQ} \quad Ó \quad n = \frac{PQ}{\left(\frac{E}{Z}\right)^2 + \frac{PQ}{N}}$$

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \quad Siéndo \quad n_0 = \frac{Z^2 PQ}{E^2} \tag{17}$$

3.15. MUESTREO ALEATORIO SIMPLE (M.A.S.).

El muestreo aleatorio simple sin reposición es un procedimiento de selección de muestras con probabilidades iguales, que consiste en obtener la muestra de forma aleatoria sin

reposición a la población de las unidades previamente seleccionadas, teniendo presente que el orden de colocación de los elementos en las muestras no intervienen (es decir, que muestras con los mismos elementos colocados en orden distinto se consideran iguales). De esta forma, las muestras con elementos repetidos son imposibles. Como el procedimiento de selección es con probabilidades iguales, todas las muestras son equiprobables, y además se cumple que todas las unidades de la población tienen la misma probabilidad de pertenecer a la muestra $\pi_i = n/N$.

3.15.1. Cálculo del tamaño de la muestra. Para diseñar la muestra es indispensable contar con el marco muestral, es decir, la lista, mapa o otra especificación de las unidades que resulta de la información previamente disponible, respecto a la población sobre la cual se basan los esquemas particulares de muestreo. Para el ejemplo que se realizará el marco muestral lo constituye 50 tiendas (Tabla 3.5), cuya característica es que los establecimientos tienen activos mayores o iguales a $1'000.000.

Cuando no se conoce la varianza poblacional de una o algunas de las características que tienen que ver con el objetivo de la investigación, se procede a tomar una muestra piloto, para estimar la muestra piloto no hay normas ni reglas específicas, esta es una decisión del investigador, en algunos casos depende del tamaño de la población, tiempo y costos. Para el ejercicio tomaremos 10 tiendas del marco muestral (Tabla 3.5), haciendo uso de una tabla de números aleatorios o de la generación de los números aleatorios de una calculadora. Sin tomar números superiores a 50 (total de la población) o repetidos se construyó la Tabla 3.6.

Tabla 3.5. Tiendas de la Ciudad de Popayán, discriminadas por el valor en activos, ingresos y propiedad del local.

# de ord.	Nombre o Razon Social	Direccion	Activos en millones de pesos	Ingresos en millones de pesos	Local propio
1	ANACONA AMELIA	-CARRERA 7E # 17BIS-17	1,000	13,888	no
2	ANACONA DE ANACONA LUCINDA	-CALLE 19 # 30-24	1,170	12,075	no
3	ANACONA PIAMBA LIOVIGILDO	-CALLE 63N NO. 7A-09	1,414	3,000	si
4	ANDRADE DASA WILLIAM JULIAN	-CALLE 68 # 10-92	1,150	0,450	no
5	ARIAS DE GAMBOA GRACIELA	-CARRERA 1AE # 9A-41	1,050	9,600	si
6	ARANGO MARULANDA JENNY LILIANA	-CARRERA 6 # 43N-50	7,350	11,550	si
7	BELTRAN PENA TITO GERARDINO	-CARRERA 11 # 12A-10	8,795	2,450	si
8	CABRERA SALAS HUGO HERNANDO	-CARRERA 6 # 12-51	10,500	521,794	no
9	CABRERA SALAS ONASIS ERNESTO	-CALLE 5 No 18-50 POPAYAN	7,200	117,197	si
10	AUSECHA ROJAS MARIA CONSUELO	-TRANSV. 19 # 10-121	1,000	7,518	si
11	BARBOSA LUZ DARY	-MANZANA 3 NO. 42A-11	1,200	1,200	no
12	BASTIDAS AVILA IRMA SIRLEY	-CALLE 29 BLOQUE H CASA No 5	1,500	4,900	si
13	BENAVIDES RODRIGUEZ BLANCA BELLANIRE	-CALLE 5B # 18-31	1,080	13,311	si
14	BENITEZ GUERRERO ENOELIO	-CARRERA 41 # 2-13	1,200	0,400	no
15	BERRIO BUITRAGO FABIO ELIAS	-CALLE 12 # 28A-04	1,120	14,300	no
16	BETANCOURT MOLANO GILMA	-MANZ. 25 #25-16 TOMAS CIPRIAN	1,270	11,270	no
17	BOLANOS BELALCAZAR LUIS EDUARDO	-CALLE 5 # 43-24	1,473	1,573	si
18	BOLANOS DE MARTINEZ ROSAURA	-GALERIA SUR PUESTO # 12	1,100	11,910	no
19	ACOSTA DE CERTUCHE PAULINA DUPERLY	-CALLE 7 # 19-114	1,900	11,900	no
20	ACOSTA DE PARAMO MATILDE	-CARRERA 2A No.7A-40	2,185	28,652	no
21	ACOSTA VILLEGAS ARNULFO	-CALLE 4 # 36-11	1,690	5,800	si
22	ACUNA REY OSCAR DARLEY	CRA 12 66N-72 BELLOHORIZONTE	2,300	44,308	si
23	AGREDO SOLIS RAUL	-CARRERA 41 # 4-11	1,547	2,160	si
24	ALVEAR DE SOLARTE MARIA MAGDALENA	-CALLE 12 13-03	2,200	6,938	no
25	AREVALO FIGUEROA JUAN BAUTISTA	-CARRERA 3 # 8-07	1,650	19,187	no
26	ARTUNDUAGA MOSQUERA EDGAR JESUS	-CALLE 7 # 21-62	2,140	0,945	no
27	ASTUDILLO JUSPIAN ADAN	-CALLE 9 # 17-25	2,800	15,500	si
28	AVILES SILVA OMAR	-CARRERA 9 # 7N-02	2,343	7,819	si
29	BARRERA DE CERON ISMENIA - SUCESORES -	-CALLE 12 # 4-93	1,710	1,030	si
30	BOLANOS MORALES ROLANDO	-CARRERA 9 No 7-99	1,650	13,500	si
31	ACOSTA DIAZ NUMAR ARNEY	CALLE 17 # 6E-19 B/ LOS SAUCES.	4,600	10,500	si
32	AGREDO DE VASQUEZ MARTHA CLELIA	-CRA 3 # 9-84	3,300	16,720	si
33	ALEGRIA DE CALVACHE EMMA CLEMENCIA	-CARRERA 2 # 3-93	5,550	14,590	si
34	ASTORQUIZA ORDONEZ FABIO	-CALLE 5B # 18- 17	6,500	13,005	si
35	BEJARANO DINA ISABEL	-CALLE 20N # 8-47 POPAYAN	5,000	80,000	si
36	BENITEZ GUERRERO GELVI VIRGILIO	-CARRERA 41 # 2-13	3,600	24,166	si
37	CABRERA LARA JUAN CARLOS	PASAJE CENTRO COMERCIAL LOCAL 46	3,750	21,676	si
38	CALVACHE ROJAS LUIS ALBERTO	-CALLE 9 # 17-55	4,000	10,800	si
39	CEBALLOS CANO SANDRA LORENA	-CALLE 56N # 10-110	4,115	9,295	si
40	CERTUCHE BARRERA LILIA NIRA	-CALLE 70D # 7-15	6,920	31,650	si
41	ARTUNDUAGA ROJAS LUIS ALBERTO	-CALLE 5 # 27A-12	1,175	2,450	no
42	ASTAIZA DE PIEDRAHITA GEORGINA	-CALLE 8 # 5-28	1,433	14,900	si
43	ASTAIZA DE VIANA LUCILA	-CALLE 4 # 25-73	1,150	11,450	no
44	ASTUDILLO DE ANACONA IRMA	-CALLE 16A # 4-51	1,500	1,500	si
45	AGREDO CIFUENTES ANA LUCIA	-CARRERA 4 # 7-35	29,469	747,544	si
46	ALVAREZ LOPEZ LTDA	-CALLE 11N # 9-44	14,384	75,990	no
47	ANACONA ANACONA LUIS ANGEL	-CALLE 19 No 31-14	10,000	30,641	si
48	ANDRADE DIAZ MARIA ELENA	-CARRERA 9 # 63N-18	10,028	35,381	si
49	CASTANEDA CASTANEDA MARIA NIDIA	-CARRERA 7 # 12 -106	18,000	58,535	no
50	CASTANO CAMPO CLAUDIA MARCELA	-CARRERA 6A # 9N-92 B/BOLIVAR	23,178	111,957	no

Fuente: Cámara de Comercio del Cauca

Tabla 3.6. Muestra Piloto, muestreo aleatorio simple (M.A.S.)

# Aleatorio	048	033	024	003	018	035	039	041	044	049
	10,028	5,550	2,200	1,414	1,100	5,000	4,115	1,175	1,500	18,000

Activos										
Local	si	si	no	si	no	si	si	no	si	no

Se procede a calcular la varianza para la variable continua (Activos) de la siguiente forma:

Tabla 3.7. Calculo de la varianza, para variables continuas

No. De orden	Números aleatorios	Local propio	Activos en miles de pesos	$(X - \overline{X})^2$
1	48	si	10,028	25,198
2	33	si	5,550	0,294
3	24	no	2,200	7,886
4	3	si	1,414	12,918
5	18	no	1,100	15,274
6	35	si	5,000	0,000
7	39	si	4,115	0,798
8	41	no	1,175	14,693
9	44	si	1,500	12,307
10	49	no	18,000	168,787
Totales			50,082	258,156
			Media	Varianza
Media y Varianza			5,008	28,684

$$\overline{X} = \frac{\sum X_i}{n} = \frac{50.082}{10} = 5.008$$

Se estima un error del 10% del promedio de los activos así:

$$E = 0{,}10(\overline{X}) = 0{,}10 \times 5.008 = 0.5008$$

$$\textit{Varianza mustral} \quad S^2 = \frac{(X - \overline{X})^2}{n-1} = \frac{258.156}{9} = 28.684$$

Una vez calculados estos estadísticos se procede a calcular el tamaño de la muestra. Se trabajará con un nivel de confianza del 90%, es decir un valor de z=1,645.

$$n = \frac{Z^2 \times N \times S^2}{N \times E^2 + Z^2 \times S^2} = \frac{(1{,}645)^2 \times 50 \times 28..684}{50 \times (0{,}5008)^2 + (1.645)^2 \times 28.684} = 43{,}045 \approx 43$$

$$.n = \frac{S^2}{\left(\frac{E}{Z}\right)^2 + \frac{S^2}{N}} = \frac{28.684}{\left(\frac{0.5008}{1.645}\right)^2 + \frac{28.684}{50}} = 43{,}045 \approx 43$$

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \quad donde \quad n_0 = \frac{Z^2 S^2}{E^2} = \frac{\frac{Z^2 S^2}{E^2}}{1 + \frac{\frac{Z^2 S^2}{E^2}}{N}} = \frac{\frac{(1,645)^2 * 28.684}{(0.5008)^2}}{1 + \frac{\frac{(1,645)^2 * 28.684}{(0.5008)^2}}{50}} = 43,045 \approx 43$$

Se observa más claramente en esta última formula que cuando N→ ∞ (la fracción de muestreo n/N tiende a cero) el tamaño muestral $n \to \frac{Z^2 S^2}{E^2} = n_0$ (n es inversamente proporcional al cuadrado del error de muestreo).

La expresión del tamaño muestral n puede ponerse en función de N y del valor n_0 como se aprecia en la ecuación 18.

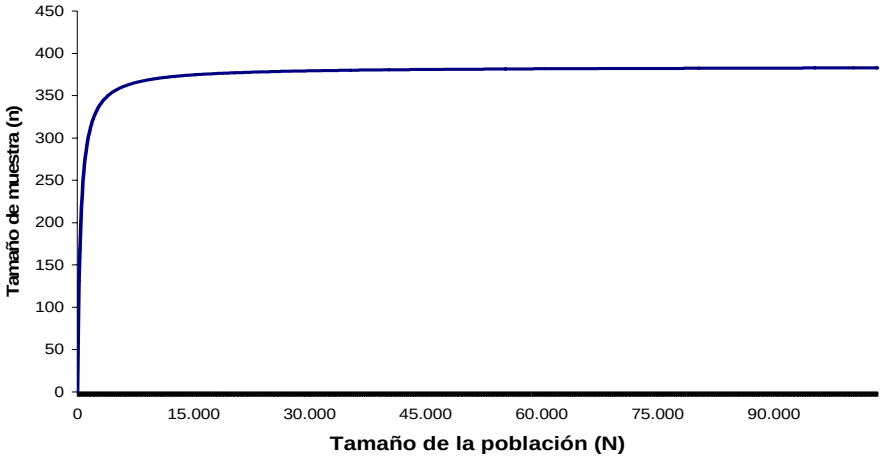
$$n = \frac{n_0}{1 + \frac{n_0}{N}} = \frac{n_0 N}{n_0 + N} = f(N) \qquad (18)$$

Si representamos gráficamente la curva de la ecuación $n = f(N)$ observamos que pasa por el origen de las coordenadas, ya que $f(0) = 0$, que tiene una asíntota paralela al eje de las X de la ecuación $n = n_0$, dado que el $\lim_{N \to \infty} f(N) = n_0$, que es siempre creciente puesto que la primera derivada:

$$f'(N) = \frac{n_0^2}{(n_0 + N)^2} \qquad (19)$$

Es siempre positiva, que no tiene máximos ni mínimos dado que la ecuación definida por $f'(N) = 0$ no tiene solución en N. Por tanto, la representación gráfica de $n = f(N)$ es la siguiente:

Figura 3.9. Representación gráfica de la curva de la ecuación $n = f(N)$



Como la curva $n = f(N)$ es creciente, al aumentar el tamaño poblacional N también aumenta el tamaño muestral n necesario para un error de muestreo dado. Pero como n ha de ser un número entero y la curva $n = n_0$ es una asíntota horizontal, desde un cierto N en adelante los aumentos de N no producen aumentos en n. (Apéndice C).

FICHA TÉCNICA

Una vez calculado el tamaño de la muestra, se requiere hacer la ficha técnica donde se relaciona en forma resumida el plan de muestreo utilizado así:

Elemento : Tiendas de la ciudad de Popayán
Unidad de muestreo : Tiendas de la ciudad de Popayán
Alcance : Ciudad de Popayán
Tiempo : Septiembre de 2003
Nivel de confianza : 90%
Z : 1,645
Varianza muestral S^2 : 28.684
Error : 5% sobre la media (0.5008)
N : 50
Tamaño de la muestra: 43

3.15.2. Tamaño de la muestra con la variable discreta (cualitativa) de proporción de tiendas con local propio

$$p = \frac{\sum a_i}{n} = \frac{6}{10} = 0{,}6$$

$$pq = 0{,}6 \times 0{,}4 = 0{,}24$$

$$n = \frac{Z^2 \times N \times p \times q}{N \times E^2 + Z^2 \times p \times q} = \frac{(1{,}645)^2 \times 50 \times 0{,}6 \times 0{,}40}{50 \times (0{,}10)^2 + (1.645)^2 \times 0{,}6 \times 0{,}40} = 28{,}25 \approx 28$$

FICHA TÉCNICA

Elemento : Tiendas de la ciudad de Popayán
Unidad de muestreo : Tiendas de la ciudad de Popayán
Alcance : Ciudad de Popayán
Tiempo : Septiembre de 2003
Nivel de confianza : 90%
Z : 1,645
P : 0.6
Q : 0.4
Error : 10% (0.10)
N : 50
Tamaño de la muestra: 28

3.16. TAMAÑO OPTIMO DE MUESTRA CON VARIABLES DISCRETAS.

Teniendo en cuenta el punto 3.13 las curvas de operación características, se procede a calcular el tamaño de la muestra con diferentes valores de p así:

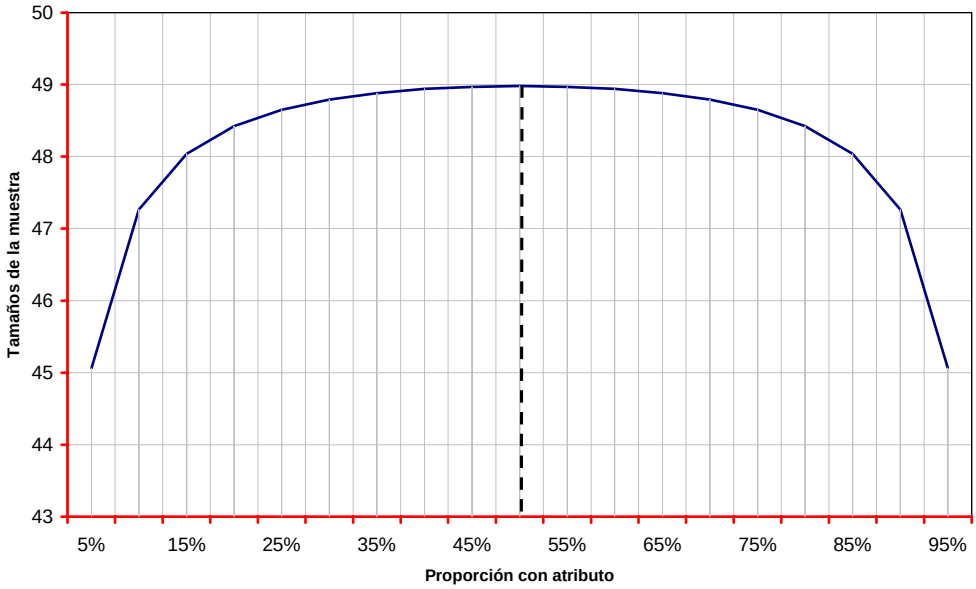
Estadísticos de la ficha técnica:

N = 50
E= 2%
Z= 1,96

Tabla 3.8. Tamaño de la muestra, con diferentes valores de p.

p	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%
q	95%	90%	85%	80%	75%	70%	65%	60%	55%	50%	45%	40%	35%	30%	25%	20%	15%	10%	5%
n=	45	47	48	48	49	49	49	49	48,97	48,98	48,97	49	49	49	49	48	48	47	45

Figura 3.10. Tamaño de la muestra con diferentes proporciones de (p)



Fuente: Tabla No. 3.8. Tamaño de la muestra, con diferentes valores de p.

Como se aprecia en la figura 3.10 el tamaño de la muestra es optimo cuando se toma las proporciones equivalentes a p=0,5 y q=0,5[4]

3.17. HOMOGENEIDAD DE LA INFORMACIÓN.

Cuando hay cierto grado de homogeneidad en la característica investigada, el tamaño tiende a ser pequeño. El grado de homogeneidad se da cuando el coeficiente de variación es menor del 30%; en estos casos es recomendable la aplicación del muestreo aleatorio simple (m.a.s.), a continuación se calcula el coeficiente de variación.

$$S = \sqrt{S^2} = \sqrt{28.684} = 5{,}36 \qquad (20)$$

[4] Martínez Bernardina, expresa al respecto que "...se tiene la costumbre de tomar P=0,5 con lo cual se obtiene el máximo valor posible de n". Estadística y muestreo. 2002. p. 349.

$$Coefciente \quad de \quad Variación \quad CV = \frac{S}{\overline{X}} \times 100 = \frac{5{,}36}{5{,}008} \times 100 \cong 106{,}94\% \qquad (21)$$

En este caso el coeficiente es mayor a 30% por tanto no es recomendable aplicar el muestreo aleatorio simple.

3.18. RELACIÓN DEL TAMAÑO DE LA MUESTRA CON EL NIVEL DE CONFIANZA Y EL ERROR.

En la tabla 3.9 se calculó el tamaño de la muestra dado los niveles de confianza del 96%, 95% y 80%, para errores del 4%, 5% y el 20% sobre el valor de la media ($\overline{X} = 5$) y $S^2 = 28{,}684$

Tabla 3.9. Calculo del tamaño de la muestra

Un mayor nivel de confianza menor tamaño de la muestra

Un mayor error menor tamaño de la muestra

N=50		NIVEL DE CONFIANZA		
		96%	95%	80%
		Z=2,05	Z=1,96	Z=1,28
ERROR	E=4%*5=0,2	49	49	48
	E=5%*5=0,25	48	48	47
	E=20%*5=1,0	35	33	23

Se concluye que un mayor nivel de confianza por ejemplo 96% y un error bajo en éste caso del 4% el tamaño de la muestra es de 49, que es casi el tamaño de la población; caso contrario ocurre con un nivel de confianza bajo como 80% y un error alto como del 20%, el tamaño de la muestra es de 23, es decir que la muestra se reduce considerablemente, lo cual puede traer como consecuencia un estudio o investigación poco confiable.

3.19. MUESTREO ESTRATIFICADO.

Este es un método que permite una selección más eficiente que el obtenido mediante el Muestreo Aleatorio Simple (M.A.S), en especial cuando la característica que se investiga es de gran variabilidad, lo cual implica un tamaño muestral relativamente más grande.

Una muestra estratificada se selecciona de la siguiente manera:

Se divide la población en grupos o estratos mutuamente excluyentes y colectivamente exhaustivos. Los estratos son mutuamente excluyentes, sólo si los miembros de un estrato no pueden ser miembros de cualquier otro. Los estratos son colectivamente exhaustivos si se utilizan todas las categorías posibles de una variable para definirlos. Es decir, las categorías "Masculino" y "Femenino" definen el área completa de la variable "sexo". Ninguna otra categoría es posible y, por tanto, los estratos son colectivamente exhaustivos.

Dependiendo de la manera como se distribuyen los elementos dentro de los estratos muestrales, hay tres métodos a saber:

1. Asignación o fijación igual
2. Asignación o fijación proporcional
3. Asignación o fijación óptima

Se trabajará con la asignación proporcional, por ser este método uno de los más utilizados en las investigaciones de mercado.

3.19.1. Muestreo estratificado - asignación proporcional. Este método permite determinar el tamaño óptimo de la muestra, así como los estimados puntuales y límites de confianza para el promedio, proporción, razón y proporciones en conglomerados. Con este método los tamaños muestrales en cada estrato se distribuyen en la misma proporción que las unidades en la población de cada uno de ellos; en otras palabras, el peso relativo dado por el número de unidades en cada estrato, en relación al total de elementos de la población, debe ser igual al obtenido en la muestra.

Antes de determinar el tamaño de la muestra, se elaboró la estratificación para la población de 50 tiendas. Se distribuyo la población de tiendas en dos estratos así:

Tabla 3.10. **Estrato I:** Tiendas con Activos menores e iguales a $2.000.000

# de ord.	Nombre o Razon Social	Direccion	Activos en millones de pesos	Ingresos en millones de pesos	Local propio
1	ANACONA AMELIA	-CARRERA 7E # 17BIS-17	1,000	13,888	no
2	AUSECHA ROJAS MARIA CONSUELO	-TRANSV. 19 # 10-121	1,000	7,518	no
3	ARIAS DE GAMBOA GRACIELA	-CARRERA 1AE # 9A-41	1,050	9,600	si
4	BENAVIDES RODRIGUEZ BLANCA BELLANIRE	-CALLE 5B # 18-31	1,080	13,311	si
5	BOLANOS DE MARTINEZ ROSAURA	-GALERIA SUR PUESTO # 12	1,100	11,910	si
6	BERRIO BUITRAGO FABIO ELIAS	-CALLE 12 # 28A-04	1,120	14,300	si
7	ANDRADE DASA WILLIAM JULIAN	-CALLE 68 # 10-92	1,150	0,450	si
8	ASTAIZA DE VIANA LUCILA	-CALLE 4 # 25-73	1,150	11,450	si
9	ANACONA DE ANACONA LUCINDA	-CALLE 19 # 30-24	1,170	12,075	si
10	ARTUNDUAGA ROJAS LUIS ALBERTO	-CALLE 5 # 27A-12	1,175	2,450	no
11	BARBOSA LUZ DARY	-MANZANA 3 NO. 42A-11	1,200	1,200	no
12	BENITEZ GUERRERO ENOELIO	-CARRERA 41 # 2-13	1,200	0,400	no
13	BETANCOURT MOLANO GILMA	-MANZ. 25 #25-16 TOMAS CIPRIAN	1,270	11,270	si
14	ANACONA PIAMBA LIOVIGILDO	-CALLE 63N NO. 7A-09	1,414	3,000	no
15	ASTAIZA DE PIEDRAHITA GEORGINA	-CALLE 8 # 5-28	1,433	14,900	si
16	BOLANOS BELALCAZAR LUIS EDUARDO	-CALLE 5 # 43-24	1,473	1,573	no
17	ASTUDILLO DE ANACONA IRMA	-CALLE 16A # 4-51	1,500	1,500	si
18	BASTIDAS AVILA IRMA SIRLEY	-CALLE 29 BLOQUE H CASA No 5	1,500	4,900	si
19	AGREDO SOLIS RAUL	-CARRERA 41 # 4-11	1,547	2,160	no
20	AREVALO FIGUEROA JUAN BAUTISTA	-CARRERA 3 # 8-07	1,650	19,187	si
21	BOLANOS MORALES ROLANDO	-CARRERA 9 No 7-99	1,650	13,500	no
22	ACOSTA VILLEGAS ARNULFO	-CALLE 4 # 36-11	1,690	5,800	no
23	BARRERA DE CERON ISMENIA - SUCESORES -	-CALLE 12 # 4-93	1,710	1,030	si
24	ACOSTA DE CERTUCHE PAULINA DUPERLY	-CALLE 7 # 19-114	1,900	11,900	no

Fuente: Cámara de Comercio del Cauca

Tabla 3.11. Estrato II. Tiendas con activos de más de $2.000.0001

# de ord.	Nombre o Razon Social	Direccion	Activos en millones de pesos	Ingresos en millones de pesos	Local propio
1	ARTUNDUAGA MOSQUERA EDGAR JESUS	-CALLE 7 # 21-62	2,140	0,945	si
2	ACOSTA DE PARAMO MATILDE	-CARRERA 2A No.7A-40	2,185	28,652	si
3	ALVEAR DE SOLARTE MARIA MAGDALENA	-CALLE 12 13-03	2,200	6,938	si
4	ACUNA REY OSCAR DARLEY	CRA 12 66N-72 BELLOHORIZONTE	2,300	44,308	si
5	AVILES SILVA OMAR	-CARRERA 9 # 7N-02	2,343	7,819	si
6	ASTUDILLO JUSPIAN ADAN	-CALLE 9 # 17-25	2,800	15,500	no
7	AGREDO DE VASQUEZ MARTHA CLELIA	-CRA 3 # 9-84	3,300	16,720	si
8	BENITEZ GUERRERO GELVI VIRGILIO	-CARRERA 41 # 2-13	3,600	24,166	si
9	CABRERA LARA JUAN CARLOS	PASAJE CENTRO COMERCIAL LOCAL 46	3,750	21,676	si
10	CALVACHE ROJAS LUIS ALBERTO	-CALLE 9 # 17-55	4,000	10,800	si
11	CEBALLOS CANO SANDRA LORENA	-CALLE 56N # 10-110	4,115	9,295	si
12	ACOSTA DIAZ NUMAR ARNEY	CALLE 17 # 6E-19 B/ LOS SAUCES.	4,600	10,500	no
13	BEJARANO DINA ISABEL	-CALLE 20N # 8-47 POPAYAN	5,000	80,000	si
14	ALEGRIA DE CALVACHE EMMA CLEMENCIA	-CARRERA 2 # 3-93	5,550	14,590	no
15	ASTORQUIZA ORDONEZ FABIO	-CALLE 5B # 18- 17	6,500	13,005	no
16	CERTUCHE BARRERA LILIA NIRA	-CALLE 70D # 7-15	6,920	31,650	si
17	CABRERA SALAS ONASIS ERNESTO	-CALLE 5 No 18-50 POPAYAN	7,200	117,197	no
18	ARANGO MARULANDA JENNY LILIANA	-CARRERA 6 # 43N-50	7,350	11,550	si
19	BELTRAN PENA TITO GERARDINO	-CARRERA 11 # 12A-10	8,795	2,450	si
20	ANACONA ANACONA LUIS ANGEL	-CALLE 19 No 31-14	10,000	30,641	no
21	ANDRADE DIAZ MARIA ELENA	-CARRERA 9 # 63N-18	10,028	35,381	si
22	CABRERA SALAS HUGO HERNANDO	-CARRERA 6 # 12-51	10,500	521,794	si
23	ALVAREZ LOPEZ LTDA	-CALLE 11N # 9-44	14,384	75,990	no
24	CASTANEDA CASTANEDA MARIA NIDIA	-CARRERA 7 # 12 -106	18,000	58,535	no
25	CASTANO CAMPO CLAUDIA MARCELA	-CARRERA 6A # 9N-92 B/BOLIVAR	23,178	111,957	si
26	AGREDO CIFUENTES ANA LUCIA	-CARRERA 4 # 7-35	29,469	747,544	si

Fuente: Cámara de Comercio del Cauca

Nomenclatura para calcular las proporciones del muestreo estratificado

N	Total de unidades que constituyen la población objetivo.
N_h	Total de unidades que contiene cada estrato poblacional
N_1, N_2, N_3 etc.	Serán los tamaños poblacionales en los estratos 1, 2, 3 etc.

$\Sigma N_h = N = N_1 + N_2 + N_3 + \ldots N_M$

$$\overline{Y}_h = \frac{\Sigma Y_{hj}}{N_h}$$ Media aritmética poblacional para cada estrato

$$\overline{Y}_{st} = \frac{\Sigma Y_h N_h}{N} = \Sigma \overline{Y}_h W_h$$ Media aritmética poblacional estratificada ponderada

$$W_h = \frac{N_h}{N} \qquad W_1 = \frac{N_1}{N}$$

$$W_2 = \frac{N_2}{N} \qquad W_3 = \frac{N_3}{N}$$

Proporción de elementos de cada estrato

$$\Sigma W_h = W_1 + W_2 + W_2 + ...W_M = 1$$

$$S^2{}_h = \frac{\Sigma Y^2{}_{hj} - N_h \overline{Y}^2{}_h}{N_h - 1}$$ Varianza poblacional en cada estrato.

n =	Número de unidades que contiene la muestra total
nh =	Número de unidades que contiene la muestra en cada estrato muestral.

$$\Sigma n_h = n = n_1 + n_2 + n_3 + ...$$

$$\overline{y}_h = \frac{\Sigma \overline{y}_{hj}}{n_h} \qquad \overline{y}_1 = \frac{\Sigma \overline{y}_{1j}}{n_1}$$

$$\overline{y}_2 = \frac{\Sigma \overline{y}_{2j}}{n_2} \qquad \overline{y}_3 = \frac{\Sigma \overline{y}_{3j}}{n_3}$$

$\overline{y}_h = Media\ \ aritmética\ \ muestral\ \ para\ \ cada\ \ estrato.$

$$\overline{y}_{st} = \frac{\Sigma N_h \overline{y}_h}{N} \ldots\ldots\ldots \overline{y}_{st} = \Sigma \overline{y}_h W_h$$ Media aritmética muestral estratificada.

$$s^2{}_h = \frac{\Sigma y^2{}_{hj} - n_h \overline{y}^2{}_h}{n_h - 1}$$ Varianza muestral en cada estrato

3.19.2. Cálculo del tamaño de la muestra. Se debe calcular la proporción de unidades en cada estrato y peso relativo:

ESTRATO	CRITERIO DE SELECCIÓN	No. Unid.	PROPORCIÓN
ESTRATO I	Activos menores o iguales a $ 2.000.000	24	$W_1 = \frac{N_1}{N} = \frac{24}{50} = 0{,}48$
ESTRATO II	Activos mayores o guales a $ 2.000.0001.	26	$W_2 = \frac{N_2}{N} = \frac{26}{50} = 0{,}52$

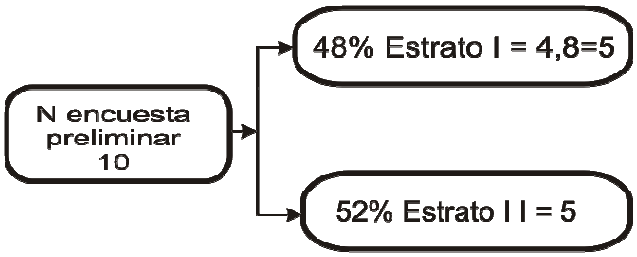
$$W_1 + W_2 = 0,48 + 0,52 = 1$$

$$\sum W_h = 1 \ldots ó \ldots 100\ \%$$

El tamaño de la muestra piloto es 10 encuestas.

En la asignación proporcional las 10 encuestas se distribuyen de la siguiente manera:

Figura 3.11. Distribución de número de encuestas por estrato
(Variables continuas)



Determinados los tamaños muestrales para la encuesta preliminar, se procede a la selección de las unidades requeridas, que permitirán calcular las varianzas y el error. Haciendo uso de una tabla de números aleatorios para los dos estratos se seleccionó de la siguiente forma:

Tabla 3.12. Muestra piloto Estrato I.

No. De orden	Números aleatorios	Local propio	Activos en miles de pesos	$(X - \overline{X})^2$
1	24	no	1,900	0,072
2	20	si	1,650	0,000
3	14	no	1,414	0,047
4	22	no	1,690	0,004
5	17	si	1,500	0,017
Totales			8,154	46,879
			Media	Varianza
Media y Varianza			1,631	11,720

Tabla 3.13. Muestra piloto Estrato II.

No. De orden	Números aleatorios	Local propio	Activos en miles de pesos	$(X - \overline{X})^2$
1	24	no	18,000	98,804
2	6	no	2,800	27,668
3	13	si	5,000	9,364
4	10	si	4,000	16,484
5	22	si	10,500	5,954
Totales			40,300	186,956
			Media	Varianza
Media y Varianza			8,060	46,739

La media ponderada es igual a:

$$\bar{x}_{st} = \sum W_h \bar{x}_h = 0{,}48(1{,}631) + 0{,}52(8.060) = 4.974$$

El error de muestreo corresponde a 0,05 $E = 0{,}05(\bar{x}) = 0{,}05(4.974) = 0{,}2487$

Se calcula de la varianza ponderada:

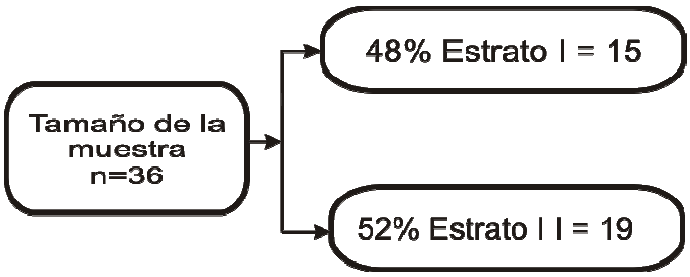
$$S^2{}_{st} = \sum W_h S^2{}_h = 0{,}48(11.72) + 0{,}52(0{,}46.739) = 29.93$$

El tamaño de la muestra es igual a:

$$n = \frac{Z^2 NS^2}{NE^2 + Z^2 S^2} = \frac{1{,}645^2 \times 50 \times 29.93}{(50 \times 0{,}806^2) + (1{,}96^2 \times 29.93)} = 35.687 \cong 36$$

Una vez calculado el tamaño de la muestra se procede a distribuir el tamaño de la muestra en forma proporcional a los estratos así:

Figura 3.12. Distribución de número de encuestas por estrato
(Variables discretas)



Ahora veamos como sería el cálculo de n para variables ***discretas o atributos***. Para ello consideramos como característica cualitativa la propiedad del local comercial, se establece un nivel confianza del 95% (Z=1,96) y un error del 5%.

Tabla 3.14. Calculo de proporciones de la muestra pilo (Variables discretas).

ESTRATO I	ESTRATO II
$P=\frac{2}{5}=0,4$	$P=\frac{3}{5}=0,6$
$Q=\frac{3}{5}=0,6$	$Q=\frac{2}{5}=0,4$
$P+Q=1$	$P+Q=1$

Calculamos PQ ponderado así:

$$PQ_{st}=\sum W_h P_h Q_h = 0,48(0,40)(0,60)+0,52(0,60)(0,40)=0,24$$

Se procede a calcular el tamaño de la muestra así:

$$n=\frac{Z^2 NPQ}{NE^2+Z^2 PQ}=\frac{1,96^2\times 50\times 0,24}{50\times 0,05^2+1,96^2\times 0,24}=44$$