CAPÍTULO 3

3. TÉCNICAS DE MUESTREO

En el presente capítulo se muestran las definiciones de los términos estadísticos a utilizar, además se expone la población objeto de estudio, así como: el marco muestral, la muestra piloto y el proceso que se efectuó para obtener el tamaño de la muestra a investigar.

3.1 Definiciones Básicas

- **Población**

Se denomina población a una colección finita o infinita de unidades (individuos o elementos), que presentan características comunes de la cual se desea obtener cierta información.

- **Muestra**

La muestra es un subconjunto de la población, seleccionado de tal forma, que sea representativo de la población en estudio, obteniéndose con el fin de

investigar alguna o algunas de las propiedades de la población de la cual procede.

- **Muestreo**

Es el procedimiento mediante el cual se obtiene una o más muestras de una población determinada. Esta muestra puede ser: No probabilística y probabilística.

- **Muestreo no probabilístico**

En el muestreo no probabilístico se usa el criterio del investigador, suele presentar grandes sesgos y es poco fiable; no garantizan la representatividad de la muestra y por lo tanto no permiten realizar estimaciones inferenciales sobre la población.

Se utilizan a veces, para estudios exploratorios, ya que el muestreo probabilístico resulta excesivamente costoso y se acude a métodos no probabilísticos, aún estando conscientes de que no sirven para realizar generalizaciones, pues no se tiene certeza de que la muestra extraída sea representativa, ya que no todos los sujetos de la población tienen la misma probabilidad de ser elegidos. También se recurre a este tipo de muestreo cuando es difícil enumerar o precisar el universo objeto de estudio o cuando no existen registros de los datos.

- **Muestreo probabilística**

Este muestreo esta basado en la teoría de la aleatoriedad o del azar; es cuando, se puede determinar de antemano la probabilidad de selección de cada uno de los elementos de la población.

- **Marco Muestral**

Es el proceso de definir y enumerar los elementos sobre los cuales se realizan las inferencias estadísticas en el muestreo probabilístico. Es importante la construcción de un marco muestral lo más perfecto posible a fin de que exista una correspondencia biunívoca entre las unidades muestrales poblacionales y las listas físicas que lo conforman. Entre los factores que contribuyen a distorsionar la calidad de un buen marco muestral están:

a. Elementos faltantes
b. Unidades ocultas por estar pareadas con otras
c. Unidades maestrales repetidas y,
d. Elementos extraños

Tipos de Muestreo

Entre los tipos de muestreo probabilístico tenemos:

- Aleatorio Simple
- Aleatorio Estratificado
- Sistemático
- Por conglomerados

A continuación se exponen los utilizados en este trabajo:

- **Muestreo Aleatorio Simple**

Es un procedimiento en donde al tomar un elemento para la muestra, cada uno de ellos tiene la misma posibilidad de ser elegido. Una muestra obtenida por este procedimiento se llama muestra aleatoria simple.

Uno de los métodos utilizados para lograr que la muestra, de una población finita sea aleatoria, enumerar todos los N elementos y una vez fijado el tamaño n de la muestra se toman al azar n números.

- **Muestreo Aleatorio Estratificado**

Consiste en considerar categorías típicas diferentes entre sí (estratos) que poseen gran homogeneidad respecto a alguna característica. Lo que se pretende con este tipo de muestreo es asegurarse de que todos los estratos de interés estarán representados adecuadamente en la muestra. Cada estrato funciona independientemente, pudiendo aplicarse dentro de ellos el muestreo aleatorio simple o el estratificado para elegir los elementos concretos que formarán parte de la muestra. En ocasiones las dificultades que plantean son demasiado grandes, pues exige un conocimiento detallado de la población. (Tamaño geográfico, sexos, edades).

La distribución de la muestra en función de los diferentes estratos se denomina afijación, y puede ser de diferentes tipos:

- ***Afijación Simple***: A cada estrato le corresponde igual número de elementos muéstrales.
- ***Afijación Proporcional:*** La distribución se hace de acuerdo con el peso (tamaño) de la población en cada estrato.
- ***Afijación Óptima:*** Se tiene en cuenta la previsible dispersión de los resultados, de modo que se considera la proporción y la desviación típica. Tiene poca aplicación ya que no se suele conocer la desviación.

Muestreo por conglomerados

Consideramos una población finita con M unidades elementales ó últimas agrupadas en N unidades mayores llamadas conglomerados ó unidades primarias, de tal forma que no existan solapamientos entre los conglomerados y que éstos contengan en todo caso a la población de estudio. Consideramos como unidad de muestreo al conglomerado y extraemos de la población una muestra de n conglomerados a partir de la cual estimaremos los parámetros poblacionales.

El número de unidades elementales de un conglomerado se denomina tamaño del conglomerado. Los conglomerados pueden ser de igual ó de

distinto tamaño y han de ser lo más heterogéneos posible dentro de ellos, de tal forma que la situación ideal sería que un solo conglomerado pueda representar fielmente a la población. Los casos más frecuentes en los que se utiliza muestreo por conglomerado son: la selección aleatoria de una población para efectuar un estudio de individuos dentro de ellas, la selección de árboles o matas en una plantación donde las unidades últimas fuesen los frutos.

Se presenta a continuación algunos de los componentes que son comunes en las diferentes fórmulas que se utilizan en el estudio tal como: para calcular el tamaño de la muestra.

- **Grado de confianza**

Es fijado por el investigador de acuerdo con su experiencia o conocimiento que tenga de la población a investigar. Sin embargo, por lo general se trabaja con el 95% o 95.5 % correspondiendo un valor Z=1.96 y Z=2.00, respectivamente.

- **Grado de variabilidad**

Está dado por la varianza. Entre más variabilidad presente la característica, mayor será el tamaño de muestra, necesario para que represente a la población.

- **Error de Muestreo**

Simbolizado por E o e, es determinado por el investigador teniendo en cuenta que a mayor error menor será el tamaño y, lo contrario, a menor error mayor será el tamaño de muestra.

3.2 Población Objetivo

La población es un conjunto de datos que representa el centro de interés a investigar; para el presente estudio se tiene como población objetivo a los estudiantes del tercer año de bachillerato que se encuentran registrados en las entidades educativas del sector Urbano de la ciudad de Guayaquil.

Por las investigaciones realizadas [(2)], se obtuvo el número de estudiantes de tercer año de diversificado de los colegios Fiscales, Particulares y Fiscomisionales (19.973). Ante la imposibilidad de recoger la opinión de todas las unidades de investigación que forman parte de la población objetivo, se escoge una muestra que es representativa, para realizar estimaciones sobre el tipo de formación que brindan los colegios a sus estudiantes.

[(2)]Dirección Provincial de Educación del Guayas (Departamento de Estadística).

Es importante conocer el número de instituciones de educación media, ya que contienen las unidades de exploración. En Tabla XV se puede apreciar el número de establecimientos que poseen el tercer año de diversificado.

TABLA XV
COLEGIOS FISCALES, PARTICULARES Y FISCOMISIONALES CON TERCER AÑO DE DIVERSIFICADO

CICLO	***COLEGIOS***		
	FISCALES	**PARTICULARES**	**FISCOMISIONALES**
Básico y diversificado	81	225	3
Básico, diversificado y post bachillerato	9	30	0
Básico, diversificado y post básico	2	10	0
Diversificado	3	0	0
Diversificado y post básico	1	0	0
TOTAL	**96**	**265**	**3**

Fuente: *Dirección Provincial de Educación del Guayas (Departamento de Estadística)*

3.3 Marco Muestral

Es una representación simbólica de la población, la cual para este estudio, es el listado de establecimientos educativos que posean el último año de bachillerato y que se encuentren localizados en la zona urbana de la ciudad de Guayaquil.

En el ANEXO I, podrá verse el nombre, ciclo y jornada de los establecimientos educativos (año lectivo 2003-2004).[(2)]

[(2)]Dirección Provincial de Educación del Guayas (Departamento de Estadística).

3.4 Muestra Piloto

Se escoge una muestra piloto, primero para comparar si el cuestionario es fácilmente entendido por los estudiantes y además para obtener cual de las variables investigadas posee mayor varianza, la que es utilizada para hallar el tamaño de la muestra real, el mismo, que se calcula a partir de las fórmulas siguientes:

$$n_O = \frac{\hat{s}^2 Z^2_{\alpha/2}}{e^2} \quad \textbf{(1.1)} \qquad n = \frac{n_O}{1 + \frac{n_O}{N}} \quad \textbf{(1.2)}$$

Donde:

N: tamaño de la población.

s^2: estimador de la varianza obtenida de la muestra piloto.

e: error o margen esperado de incertidumbre.

Z: prueba utilizada cuando el tamaño de la muestra es (>30) con 95% de confianza $Z_{a/2} = 1.96$.

n: tamaño de la muestra real.

n_0: tamaño de la muestra real si el N (tamaño de la población es muy grande).

El tipo de muestreo a utilizar en esta parte es *aleatorio simple*.

La muestra piloto se realizó a 51 estudiantes pertenecientes a colegios particulares y fiscales. La variable que se consideró, por obtener la mayor

varianza y ser de interés fue la siguiente: Tu colegio te formó para el trabajo; cuyas respuestas podían ser: 1. Total Desacuerdo, 2. Parcial Desacuerdo, 3. Indiferente, 4. Parcial Acuerdo, 5. Total Acuerdo. Las dos últimas opciones fueron tomadas como $\hat{p}$ (proporción de estudiantes que están de acuerdo) y las restantes $\hat{q}$ (proporción de alumnos que no están de acuerdo o se encuentran indiferentes); obteniendo como resultado para las proporciones estimadas los siguientes valores: $\hat{p}$ = 0.78 y $\hat{q}$ = 0.22.

3.5 Cálculo del Tamaño de la muestra

A la hora de determinar el tamaño de la muestra hay que tomar en cuenta varios factores como: la variabilidad inherente en la población de mediciones, la precisión que el estimar requiere y el grado de confianza. Todos estos criterios están relacionados, a mayor dispersión de los datos con respecto a la media, mayor es el tamaño de la muestra para obtener una mejor precisión y viceversa.

Para obtener el tamaño de la muestra, es necesario conocer:

- Error máximo admisible (error de diseño) = 0.04
- Nivel de Confianza: (1-a)100% con a=0.05, se tiene entonces 95% de confianza.

- $Z_{a/2}$ = 1.96, cuyo valor es obtenido de la Tabla de la Distribución Normal.
- Tamaño de la población, N=19.973.

Los valores de las estimaciones de $\hat{p}$ = 0.78 y $\hat{q}$ = 0.22 (obtenidos en la muestra piloto); así como la estimación del valor de la varianza ($\hat{\sigma}^2 = \hat{p} * \hat{q}$), la cual es: 0.1716.

Después de obtener estos valores se utiliza el muestreo aleatorio simple, para conocer el tamaño de la muestra n. Al reemplazar los valores en las fórmulas **(1.1) (1.2)**, se obtiene lo siguiente:

$$n_o = \frac{1.96^2}{0.04^2}(0.78 * 0.22) = 412.01$$

$$n = \frac{412.01}{1 + \frac{412.01}{19973}} = 403.68$$

Por lo tanto el tamaño de muestra resultante para esta investigación es n= 403 estudiantes.

El total de muestra obtenida se lo distribuye en los tipos de colegios: Fiscales, Particulares y Fiscomisionales, para esto se utiliza el muestreo aleatorio estratificado en el cual, las unidades de investigación que se

encuentran dentro de cada estrato presentan características homogéneas, mientras que los que se sitúan entre ellos son heterogéneos.

Se tendrá tres estratos (los cuales se llamaran: FISCALES, PARTICULARES Y FISCOMISIONALES); el primero representará a las instituciones fiscales, el segundo a las instituciones particulares y el tercero a las instituciones fiscomisionales que posean el tercer año de bachillerato, para determinar el número de unidades de cada estrato que serán seleccionadas aleatoriamente, se utilizará la afijación proporcional.

En la siguiente tabla, se muestra la cantidad de estudiantes en cada uno de los estratos distribuidos proporcionalmente.

TABLA XVI
TAMAÑO DE MUESTRA DISTRIBUIDO POR ESTRATOS

h	*ESTRATO*	*Total de alumnos (N_h)*	*Peso del estrato $W_h = (N_h/N)$*	*Tamaño de la muestra $n_h = (n*W_h)$*
1	Fiscales	11045	0.55	222
2	Particulares	8731	0.44	177
3	Fiscomisionales	197	0.01	4

Fuente: *Dirección Provincial de Educación del Guayas (Departamento de Estadística)*

A continuación se determina el número de instituciones de las cuales se extraerán las unidades de investigación, para esto se utilizará el muestreo por conglomerado. Los establecimientos educativos muestran las unidades

de estudio con características heterogéneas dentro de estos y homogéneas fuera de los mismos.

Como las unidades de investigación se encuentran agrupadas en conglomerados, se necesita conocer, la cantidad de conglomerados a entrevistar. Para lo cual se utilizan las fórmulas [3] que se presentan a continuación:

$$\delta = \frac{\sum_{ii}^{N} \sum_{j \neq l}^{\overline{M}} (X_{ij} - \overline{\overline{X}}) * (X_{lj} - \overline{\overline{X}})}{N * (\overline{M} - l) * \overline{M} * \hat{\sigma}^2} \qquad \textbf{(1.3)}$$

Donde:

- $\overline{M}$: Número de unidades elementales por conglomerado.
- N: número de conglomerados en la muestra.
- $\overline{\overline{X}}$: Media de todos los datos de la muestra.

El símbolo d representa el coeficiente de correlación intraconglomerado o (media de homogeneidad). El índice i (i=1…N), el índice j (j=1…$\overline{M}$), el índice i toma números diferentes de j.

---

[3] Azorín Francisco. Métodos y Aplicaciones de Muestreo. Madrid. Alianza Editorial S.A. (1986)

Finalmente se calcula:

$$n_c = n_a(1+ d (\overline{M} -1)) \qquad \textbf{(1.4)}$$

donde:

- n_a: es el tamaño de la muestra obtenida con muestreo aleatorio simple.
- n_c: tamaño de la muestra con muestreo por conglomerados.
- $\overline{M}$: número de unidades elementales por conglomerado.

El coeficiente (1+ d ($\overline{M}$ -1)) denominado “efecto de diseño”, es el factor por el que es necesario multiplicar al tamaño de la muestra obtenida con muestreo aleatorio simple para obtener el correspondiente al muestreo por conglomerado que proporcione la misma precisión.

Para obtener el tamaño del conglomerado, es necesario conocer los siguientes valores:

- El coeficiente de correlación intraconglomerado es: d= - 0.015. Como podemos observar el valor del coeficiente es negativo, lo cual nos indica que la precisión del muestreo por conglomerado es superior a la del muestreo aleatorio simple.

- Al no tener información sobre el número de estudiantes que se encuentran en cada curso, se decide calcular el promedio de estudiantes registrados en los dos cursos de la muestra piloto (((30+34)/2= 32 estudiantes), esta es la cantidad de adolescentes que se desea investigar en cada curso (subconglomerado) de la muestra real; se decide entrevistar en cada conglomerado a dos subconglomerados. Por lo que tenemos que $\overline{M}$ = 64.

Se sustituyen los valores en la fórmula 1.4 y se obtiene lo siguiente:

$$n_c: 403^* (1-(0.01558^*((64-1)) = 7.43$$

El número de conglomerados a entrevistar es de 7, pero estos se encuentran estratificados en colegios Fiscales, Particulares y Fiscomisionales. Para determinar el tamaño de cada estrato, se aplica el método de afijación proporcional. Con este método las n unidades de la muestra se distribuyen proporcionalmente a los tamaños de cada estrato, expresados en números de unidades.

La ponderación o peso de cada estrato se obtuvo a partir del siguiente cálculo:

$$W_h = \frac{n_h \text{ (tamaño de cada estrato)}}{N\text{(tamaño de la población)}} * n_c \qquad \textbf{(1.5)}$$

Según los registros [(2)], el número de estudiantes de tercer año de bachillerato en los Colegios Fiscales es: 11045, mientras que en los Colegios Particulares es: 8731 y en los Colegios Fiscomisionales es: 197.

Utilizamos la Fórmula **1.5** para obtener el número de colegios en cada estrato. El cuál es:

Número de Colegios Fiscales = (7) * (11045/19973) = 4

Número de Colegios Particulares = (7) * (8731/19973) = 3

Número de Colegios Fiscomisionales = (7) * (197/19973) = 1

Para seleccionar los colegios que forman parte de la muestra, se tuvo que enumerar en serie a las instituciones que integran parte del marco muestral, para luego generar números aleatorios y así obtener los establecimientos educativos en los cuales, se debe investigar. El método de muestreo utilizado es el aleatorio simple.

[(2)]Dirección Provincial de Educación del Guayas (Departamento de Estadística).

La Tabla XVII nos muestra las instituciones educativas con su respectivo número de estudiantes que fueron seleccionados aleatoriamente, para formar parte de la muestra en el presente estudio.

TABLA XVII
NOMBRE DE INSTITUCIONES Y NÚMERO DE ALUMNOS SELECCIONADOS PARA LA MUESTRA

Colegios Fiscales	Número de estudiantes
Emilio Estrada	15
Joaquín Gallegos Lara	77
Provincia de Carchi	37
Rita Lecumberri	70
Teodoro Maldonado	34
Colegios Particulares	
Cayetano Tarruel	60
Liceo Naval	67
O'NEIL	39
Colegios Fiscomisionales	
Maria Mazarrello	4

Fuente y elaboración: *Johanna Valencia Flores*