

# Capítulo 2

## 2. MARCO TEORICO

En este capitulo se detallarán las etapas que se desarrollarán para llevar a cabo el presente estudio. Las etapas serán explicadas en el orden que se recomienda hacer el diseño de una encuesta, estas etapas son:

### 2.1. Población objetivo

La población objetivo es la que se intenta investigar, como el presente estudio consta de dos objetivos definidos, los cuales son:

1. Determinar el nivel de conocimientos de los estudiantes que han cursado las materias de Estadística, Fundamentos de Computación y Análisis Numérico.

2. Determinar si la enseñanza de estas materias debe readecuarse, en la medida en que las distintas carreras de ingeniería apliquen estos conocimientos.

Tendremos para cada objetivo una población distinta, para el primer objetivo la población objeto de este estudio serán todos los estudiantes de las carreras de Ingeniería de la ESPOL.

Para el segundo objetivo la población objeto del estudio son los profesores, coordinadores, decanos y egresados de las distintas carreras de ingeniería de la ESPOL.

## **2.2. Diseño del cuestionario**

El cuestionario es el medio de comunicación entre el que solicita los datos y el respondiente, así como un documento de trabajo para los codificadores, depuradores y perforistas, que permite el tratamiento informático de los datos.

El cuestionario se suele estructurar en secciones y estas en preguntas que deberían ser fáciles de comprender y contestar.

Atendiendo estos lineamientos, para cumplir con el primer objetivo se diseñaron tres cuestionarios: el primero correspondiente a determinar el nivel de conocimientos de los estudiantes en Estadística, el segundo correspondiente a determinar el nivel de conocimientos de los estudiantes en Fundamentos de Computación y el tercero es el correspondiente a determinar el nivel de conocimientos de los estudiantes en Análisis Numérico.

Para cumplir con el segundo objetivo también se diseñó un cuestionario para los profesores, coordinadores, decanos y egresados.

Los cuestionarios se los encontrará en los anexos.

### **2.3. Marco muestral**

Antes de ser seleccionada la muestra, la población debe ser dividida en partes llamadas unidades de muestreo o unidades, estas deben cubrir la totalidad de la población objeto de estudio y no asumir que todo elemento de la población pertenece a una y solamente una unidad.

Por lo tanto el marco muestral es un listado de unidades a partir del cual se selecciona la muestra.

Tendremos para cada objetivo un marco muestral distinto, para el primer objetivo el marco muestral correspondiente es el listado de los estudiantes de las distintas carreras de ingeniería de la ESPOL.

Para el segundo objetivo el marco muestral correspondiente es el listado de profesores, coordinadores, decanos y egresados de las distintas carreras de ingeniería de la ESPOL.

#### **2.4. Tipo de muestreo**

En muchos casos puede no ser posible o conveniente obtener información de todas las unidades de la población limitándose las observaciones a un subconjunto de la misma a la que denominaremos muestra.

Existen actualmente una gran variedad de planes para seleccionar una muestra por lo cual para el primero y segundo objetivo del presente estudio se utilizará Muestreo Estratificado.

En este tipo de muestreo con objeto de mejorar las estimaciones mediante el previo agrupamiento de los elementos más parecidos entre sí, se divide la población en subpoblaciones o estratos, dentro de cada una de los cuales se hace una selección aleatoria. En realidad el método que acabamos de definir debe denominarse muestreo estratificado aleatorio, ya que el solo hecho de estratificar no determina el procedimiento de selección empleado dentro de cada estrato.

Para los dos objetivos de este estudio se estratificará la población por facultades, es decir la población será dividida en cuatro estratos que son los siguientes:

- Facultad de Ingeniería en Ciencias de la Tierra (FICT).
- Facultad de Ingeniería Marítima y Ciencias del Mar (FIMCM).
- Facultad de Ingeniería en Electricidad y Computación (FIEC).
- Facultad de Ingeniería Mecánica y Ciencias de la Producción (FIMCP).

A continuación se presentará las distintas carreras de ingeniería por facultad con su respectivo número de estudiantes, correspondientes al primer término del año 2000:

- Facultad de Ingeniería en Ciencias de la Tierra (FICT)
  - Ingeniería en Minas 16 estudiantes
  - Ingeniería en Petróleo 38 estudiantes
  - Ingeniería en Geología 25 estudiantes
  - Ingeniería Civil 67 estudiantes
  
- Facultad de Ingeniería Marítima y Ciencias del Mar (FIMCM)
  - Ingeniería Naval
    - Especialización: Construcción y reparación 18 estudiantes
    - Especialización: Civil 13 estudiantes
  
- Facultad de Ingeniería en Electricidad y Computación (FIEC)
  - Ingeniería en Electrónica y Telecomunicación 406 estudiantes
  - Ingeniería en Electricidad
    - Especialización: Potencia 121 estudiantes
    - Especialización: Industrial 189 estudiantes
  - Ingeniería en Computación 307 estudiantes
  
- Facultad de Ingeniería Mecánica y Ciencias de la Producción (FIMCP)
  - Ingeniería Mecánica 187 estudiantes
  - Ingeniería y Administración de la producción

|                           |                 |
|---------------------------|-----------------|
| Industrial                | 282 estudiantes |
| - Ingeniería en Alimentos | 77 estudiantes  |
| - Ingeniería Agropecuaria | 80 estudiantes  |

## 2.5. Muestra piloto

La muestra piloto aparte de que nos ayuda a estimar los parámetros poblacionales para determinar el tamaño de la muestra, también es importante para ensayar el cuestionario en condiciones reales y poner a prueba los aspectos fundamentales de la muestra principal.

La encuesta piloto debe tener una estructura similar a la de la encuesta principal. Por supuesto, los papeles de tiempo y medios juegan un papel importante en su diseño.

Se propone la inclusión de la muestra piloto de algunas cuestiones específicas que a continuación recogeremos en forma breve:

- Idoneidad del marco propuesto para seleccionar la muestra.
- Variabilidad de ciertos caracteres, con relación a la materia objeto de estudio. En general, el tamaño de la muestra piloto no es lo suficientemente grande como para determinar el tamaño de

la muestra de la encuesta principal. La información obtenida en la primera debe suplementarse con cualquier otra información pertinente de que se disponga.

- Tasa esperada de falta de respuesta, con distinción de negativas y ausentes.
- Idoneidad del método de recolección de datos y del cuestionario.
- Eficacia de las instrucciones dadas a los entrevistadores y codificadores, y en general de la organización del trabajo.
- Probable coste y duración de la encuesta principal.

## **2.6. Determinación del tamaño de la muestra**

Al determinar el nivel de conocimientos de los estudiantes de las carreras de ingeniería el tamaño de la muestra diferirá dependiendo si el estudio es para estadística, fundamentos de computación y análisis numérico, debido a que el número de estudiantes que han aprobado cada una de estas materias es diferente.

En el caso de muestreo estratificado aleatorio con proporciones que es el tipo de muestreo que se utilizará en este estudio, con una afijación de mínima varianza, con el error máximo admisible y con la



confianza fijados, mediante la siguiente ecuación podemos calcular n (el tamaño de muestra):

$$n = \frac{n_o}{1 + \frac{n_o}{N}}$$

Donde:

$$n_o = \frac{Z_{\alpha/2}^2 \cdot P \cdot Q}{E^2}$$

N = Tamaño de la población.

E = La precisión prefijada, representada por el error máximo admisible.

$Z_{\alpha/2}$  = El grado de seguridad o confianza.

P = Proporción de elementos en la muestra que tienen una característica particular.

Q = (1 - P). Proporción de elementos en la muestra que no tienen la característica de interés.

Los estimadores de los parámetros P y Q se los obtiene a partir de una muestra piloto.

E y  $Z_{\alpha/2}$  los fija el investigador, de acuerdo a su experiencia.

### 2.6.1. Afijación de la muestra

Luego de seleccionar el tamaño de la muestra es necesario realizar la afijación de la muestra que es el reparto, asignación, adjudicación o distribución del tamaño muestral  $n$  entre los diferentes estratos. Esto es, a la determinación de los valores de  $n_h$  para  $h=1,2,\dots,l$  que verifiquen:

$$n_1+n_2+n_3+\dots+n_l = n$$

donde  $l$  representa el número de estratos.

Pueden establecerse muchas afijaciones o maneras de repartir la muestra, pero las más importantes son: la afijación uniforme, la afijación proporcional, la de mínima varianza y la óptima, de la cual se eligirá para este estudio la afijación de mínima varianza, que consiste en determinar los valores de  $n_h$  de forma que para un tamaño de muestra fijo igual a  $n$  la varianza sea mínima. Esta afijación para cada uno de los estratos se calcula de la siguiente manera:

$$n_h = n \cdot \frac{N_h \cdot \sqrt{P \cdot Q}}{\sum_{h=1}^l N_h \cdot \sqrt{P \cdot Q}}$$

$n$  = Tamaño de la muestra.

$N_h$  = Tamaño de la población en el estrato  $h$ .

$n_h$  = Tamaño de la muestra en el estrato  $h$ .

**2.6.1.1 Determinación del tamaño de muestra en el estudio del nivel de conocimientos de los estudiantes en la materia de Estadística.**

El tamaño de la población aquí es de  $N=1826$  estudiantes, los que se encuentra distribuidos por facultad de la siguiente manera:

**Tabla I**  
**Tamaño de la población de estudiantes de estadística por facultad.**

| <b>Facultad</b> | <b>Tamaño de la población en cada estrato (<math>N_h</math>)</b> |
|-----------------|--|
| FIMCP           | 626 estudiantes  |
| FIEC            | 1023 estudiantes   |
| FIMCM           | 31 estudiantes   |
| FICT            | 146 estudiantes  |

Como se mencionó anteriormente para estimar los parámetros poblacionales  $P$  y  $Q$  es necesario la realización de una muestra piloto, cuyo tamaño queda a criterio del investigador, para este caso se eligió un tamaño de muestra piloto de 40 estudiantes, la cual nos dio los siguientes resultados:

**Tabla II**  
**Estimaciones de la muestra piloto realizada**  
**a los estudiantes de estadística.**

| <b>Estratos</b> | <b>Tamaño de la muestra piloto por estrato</b> | <b>Proporción de estudiantes que responden correctamente</b> | <b>Varianza de la proporción</b> |
|-----------------|--|--|----------------------------------|
| FIMCP           | 13 estudiantes                                 | 0.38   | 0.0196                           |
| FIEC            | 18 estudiantes                                 | 0.33   | 0.013                            |
| FIMCM           | 3 estudiantes                                  | 0.33   | 0.11                             |
| FICT            | 6 estudiantes                                  | 0.33   | 0.044                            |

La proporción de elementos en cada estrato que tienen una característica particular se la calcula de la siguiente manera:

$$\hat{P} = \sum_{i=1}^{n_h} \frac{A_i}{n_h}$$

Siendo  $A_i$  una variable que toma los valores uno o cero, según que la unidad  $u_i$  pertenezca o no a una determinada clase, y  $n_h$  es el tamaño de la muestra en cada estrato.

La media muestral estratificada para proporciones, se lo calcula:

$$\hat{P}_{est} = \sum_{h=1}^l \sum_{i=1}^{n_h} \frac{N_h \cdot A_{hi}}{N \cdot n_h}$$

Siendo  $A_{hi}$  una variable que toma los valores de uno o cero, según que la unidad  $u_{hi}$  pertenezca o no a la muestra de  $n_h$  unidades, obtenida en el estrato h-ésimo.

Así también la varianza para los estimadores de la proporción en el h-ésimo estrato, se lo calcula mediante la siguiente formula:

$$\hat{V}(\hat{P}) = \frac{\hat{P}_h \cdot \hat{Q}_h}{n_h - 1}$$

Para calcular la varianza estratificada para los estimadores de la proporción tenemos:

$$\hat{V}(\hat{P}_{est}) = \sum_{h=1}^l \left( \frac{N_h}{N} \right)^2 \cdot \left( \frac{N_h - n_h}{N_h} \right) \cdot \frac{\hat{P}_h \cdot \hat{Q}_h}{n_h - 1}$$

Teniendo estos estimadores, procedemos a calcular el tamaño de la muestra con una confianza del 95%, lo que nos da un valor de  $Z_{\alpha/2} = 1.96$  (este valor se lo obtiene de la tabla de la distribución normal) y un error máximo admisible de  $E = 0.06$ , lo que al reemplazar estos datos en la fórmula dada anteriormente para calcular el tamaño muestral nos da que:

$$\mathbf{n = 214 \text{ estudiantes}}$$

Al realizar la afijación de mínima varianza tenemos que el tamaño de muestra se va a encontrar distribuido entre los estratos de la siguiente manera:

**Tabla III**  
**Tamaño de muestra de los estudiantes de estadística por estrato.**

| <b>Facultad</b> | <b>Tamaño de la muestra en cada estrato (<math>n_h</math>)</b> |
|-----------------|--|
| FIMCP           | 75 estudiantes   |
| FIEC            | 118 estudiantes  |
| FIMCM           | 4 estudiantes  |
| FICT            | 17 estudiantes   |

**2.6.1.2 Determinación del tamaño de muestra en el estudio del nivel de conocimientos de los estudiantes en la materia de Fundamentos de Computación.**

El tamaño de la población aquí es de  $N = 1113$  estudiantes, los que se encuentra distribuidos por facultad de la siguiente manera:

**Tabla IV**  
**Tamaño de la población de estudiantes de fundamentos de computación por facultad.**

| <b>Facultad</b> | <b>Tamaño de la población en cada estrato (<math>N_h</math>)</b> |
|-----------------|--|
| FIMCP           | 626 estudiantes  |

|       |                 |
|-------|-----------------|
| FIEC  | 310 estudiantes |
| FIMCM | 31 estudiantes  |
| FICT  | 146 estudiantes |

Aquí se omite a la mayor parte de los estudiantes de la Facultad de Ingeniería en Electricidad y Computación (FIEC) excepto a los estudiantes de Ingeniería en Electricidad, debido a que el dictado de esta materia no corresponde al Instituto de Ciencias Matemáticas, ya que por ser materia fundamental es dictada por la propia facultad.

Para estimar los parámetros poblacionales P y Q es necesario la realización de una muestra piloto, cuyo tamaño queda a criterio del investigador, para este caso se eligió un tamaño de muestra piloto de 30 estudiantes, la cual nos dio los siguientes resultados:

**Tabla V**  
**Estimaciones de la muestra piloto realizada**  
**a los estudiantes de fundamentos de computación.**

| <b>Estratos</b> | <b>Tamaño de la muestra piloto por estrato</b> | <b>Proporción de estudiantes que responden correctamente</b> | <b>Varianza de la proporción</b> |
|-----------------|--|--|----------------------------------|
| FIMCP           | 13 estudiantes                                 | 0.77   | 0.0148                           |
| FIEC            | 8 estudiantes                                  | 0.625  | 0.0335                           |
| FIMCM           | 3 estudiantes                                  | 0.67   | 0.11                             |
| FICT            | 6 estudiantes                                  | 0.67   | 0.044                            |

Teniendo estos estimadores, procedemos a calcular el tamaño de la muestra con una confianza del 95%, lo que nos da un valor de  $Z_{\alpha/2} = 1.96$  y un error máximo admisible de  $E = 0.06$ , lo que al reemplazar estos datos en la fórmula dada anteriormente para calcular el tamaño muestral nos da que:

$$n = 186 \text{ estudiantes}$$

Al realizar la afijación de mínima varianza tenemos que el tamaño de muestra se va a encontrar distribuido entre los estratos de la siguiente manera:

**Tabla VI**  
**Tamaño de muestra de los estudiantes**  
**de fundamentos de computación por estrato.**

| Facultad | Tamaño de la muestra en cada estrato ( $n_h$ ) |
|----------|--|
| FIMCP    | 99 estudiantes                                 |
| FIEC     | 56 estudiantes                                 |
| FIMCM    | 5 estudiantes                                  |
| FICT     | 26 estudiantes                                 |

**2.6.1.3. Determinación del tamaño de muestra en el estudio del nivel de conocimientos de los estudiantes en la materia de Análisis Numérico.**



El tamaño de la población aquí es de  $N=1746$  estudiantes, los que se encuentra distribuidos por facultad de la siguiente manera:

**Tabla VII**  
**Tamaño de la población de estudiantes de análisis numérico por facultad.**

| Facultad | Tamaño de la población en cada estrato ( $N_h$ ) |
|----------|--|
| FIMCP    | 546 estudiantes                                  |
| FIEC     | 1023 estudiantes                                 |
| FIMCM    | 31 estudiantes                                   |
| FICT     | 146 estudiantes                                  |

En la materia de Análisis Numérico disminuye el número de estudiantes de la FIMCP debido a que Ingeniería Agropecuaria no recibe esta materia.

Para estimar los parámetros poblacionales  $P$  y  $Q$  es necesario la realización de una muestra piloto, cuyo tamaño queda a criterio del investigador, para este caso se eligió un tamaño de muestra piloto de 40 estudiantes, la cual nos dio los siguientes resultados:

**Tabla VIII**  
**Estimaciones de la muestra piloto realizada a los estudiantes de análisis numérico.**

| <b>Estratos</b> | <b>Tamaño de la muestra piloto por estrato</b> | <b>Proporción de estudiantes que responden correctamente</b> | <b>Varianza de la proporción</b> |
|-----------------|--|--|----------------------------------|
| FIMCP           | 13 estudiantes                                 | 0.77   | 0.0148                           |
| FIEC            | 18 estudiantes                                 | 0.83   | 0.0083                           |
| FIMCM           | 3 estudiantes                                  | 0.67   | 0.11                             |
| FICT            | 6 estudiantes                                  | 0.67   | 0.044                            |

Teniendo estos estimadores, procedemos a calcular el tamaño de la muestra con una confianza del 95%, lo que nos da un valor de  $Z_{\alpha/2} = 1.96$  y un error máximo admisible de  $E = 0.06$ , lo que al reemplazar estos datos en la fórmula dada anteriormente para calcular el tamaño muestral nos da que:

$$n = 168 \text{ estudiantes}$$

Al realizar la afijación de mínima varianza tenemos que el tamaño de muestra se va a encontrar distribuido entre los estratos de la siguiente manera:

**Tabla IX**  
**Tamaño de muestra de los estudiantes de análisis numérico por estrato.**

| <b>Facultad</b> | <b>Tamaño de la muestra en cada estrato (<math>n_h</math>)</b> |
|-----------------|--|
| FIMCP           | 55 estudiantes   |
| FIEC            | 92 estudiantes   |
| FIMCM           | 4 estudiantes  |
| FICT            | 17 estudiantes   |

## 2.7. Evaluación de la calidad de la encuesta.

El programa de evaluación de la calidad de la encuesta pretende medir dos grandes tipos de errores que se producen en toda investigación por muestreo.

1. **Errores de muestreo**, inherentes a la utilización de muestras para obtener la información. Este error se lo calcula de la siguiente manera:

$E = |\hat{\theta} - \theta|$

, donde E es el error de muestreo,  $\theta$  es el valor desconocido de la característica poblacional que tratamos de estimar y  $\hat{\theta}$  es el valor observado.

2. **Errores ajenos al muestreo**, que son comunes a toda investigación estadística, tanto si la información es recogida por muestreo como por censo. Estos errores se presentan en cualquier fase del proceso estadístico: antes de la toma de datos, por deficiencias del marco e insuficiencia en las definiciones y cuestionarios, durante la toma de datos, por defectos en la labor de los informantes; y después de dicha toma, por negligencia en las depuraciones y codificaciones, por errores de perforación, grabación y tabulación, y por erratas de impresión.

Entre los errores ajenos al muestreo se distinguen:

- a) **Errores de cobertura**, que consisten en la omisión o inclusión errónea de personas.
- b) **Errores de contenido**, que se producen por la clasificación incorrecta de una unidad respecto a su carácter de referencia.

Para obtener información acerca de la magnitud de estos errores se utiliza el modelo de entrevista repetida. El procedimiento operativo, muy simple consiste en repetir la entrevista, por agentes especialmente adiestrados, en una submuestra de la muestra original. Mediante el cotejo de los datos proporcionados por la entrevista original y la entrevista repetida se obtiene medida de diferentes errores y diversos indicadores de calidad.

