# Capítulo IV

4. PRELIMINARES MATEMÁTICOS: TEORÍA DEL MUESTREO

### 4.1. Introducción

Este capítulo expone la teoría estadística que se utilizará en el estudio; además, se define la técnica de muestreo, el proceso de diseño de la muestra y la encuesta preliminar, así como también la estimación del tamaño muestral, distribución proporcional de la muestra y determinación de la muestra final que se realizará en el Sector denominado Centro-Sur-Oeste el mismo que comprende el conjunto de las Parroquias 9 de Octubre, Sucre, Urdaneta y García Moreno.

**4.2. Definiciones Básicas**

La teoría del muestreo proporciona una técnica estadística práctica que busca obtener datos de una población (hogares, empresas, árboles, etc.) en su totalidad, utilizando tan sólo una parte reducida de la misma denominada ***muestra***, aunque como es lógico pagando algún coste (calculable) en cuanto a la precisión de las medidas poblacionales inferidas.

En los problemas estadísticos, por lo general se hace necesario investigar una o varias particularidades de cada fenómeno individual, para conocer las características más sobresalientes del colectivo. Si deseamos investigar que parte del ingreso se dedica al consumo y cual al ahorro, necesitamos levantar un censo; es decir obtener esas relaciones en cada una de las unidades del conjunto.

En muchos casos, por motivos de orden económico, por falta de personal o de tiempo, no es posible realizar las investigaciones completas, y, en tales casos, recurrimos al***método representativo o Muestreo****,* que consiste en seleccionar una fracción únicamente de las unidades del conjunto, de manera que esa parte represente al colectivo en sus características más sobresalientes, incluyendo las medidas de variabilidad.

Sin embargo siempre existirá una diferencia entre el resultado obtenido por la muestra y el correspondiente al universo. Esta diferencia entre los dos resultados, se denomina ***error del muestreo***. Mientras más grande es la muestra, es menor, en promedio, el error de muestreo y, por lo tanto, es mayor la confianza que se tiene en los resultados. Para que una muestra sea útil, se debe tener una idea del error del muestreo.

Las principales ventajas del muestreo sobre un censo de todos los elementos de un colectivo, son: la reducción de los costos de la información, puesto que se recogen datos de un grupo reducido de elementos del universo; por esta razón, se logra mayor rapidez en la recolección de los datos, lo que es sumamente importante cuando la información que se pide es urgente. Además como se estudian pocos elementos del colectivo, las informaciones se recopilan con mayor cuidado, se logran resultados con menos errores de observación.

En un sentido general la teoría del muestreo puede considerarse como coexistentes con los métodos estadísticos, la mayoría de los desarrollos modernos en estadística se refieren a inferencias sobre la población teniendo como única información disponible una muestra de los elementos que componen dicha población.

En toda investigación estadística existe un conjunto de elementos sobre los que se toma información, este conjunto de elementos es lo que se denota con el nombre de ***población o universo estadístico***.

La ***población o universo*** es el conjunto completo de todas las ***unidades******de muestreo*** cuyas características se van a estimar, si el investigador toma información de todos y cada uno de los elementos de la población estadística, se dice que se está realizando un ***censo***, pero en la realidad esta clase de estudio no es factible, ya sea por diversas razones tales como: el costo que resulta tomar toda la información, eso si la información se compone de elementos finitos, lo que no sucede en todos los casos, o porque la toma de la información lleva consigo la destrucción de los elementos en cuestión.

Colección de elementos acerca de los cuales se desea hacer alguna inferencia se denomina ***Población Objetivo.***

El investigador deberá definir la población objetivo de forma cuidadosa y completa antes de recolectar la muestra, pero como mencionamos anteriormente el muestreo de toda la población objetivo es algunas veces imposible, ya que se presentarán problemas ajenos al muestreo. Por lo tanto la población objetivo se ve restringida a la hora de obtener información de sus elementos o unidades, de lo que derivamos la ***población investigada*** que no es más que la población que realmente es objeto de estudio.

La población que se muestrea (población investigada) debe coincidir con la población sobre la cual se desea información (población objetivo).

La ***unidad de análisis*** es la unidad para la que deseamos obtener información estadística, en las ***encuestas*** de tipo usual, pueden ser personas, hogares, fincas, firmas comerciales, tarjetas perforadas o productos surgidos de algún proceso mecánico.

La unidad de análisis se denomina frecuentemente elemento de población, en una misma encuesta pueden existir más de un elemento, por ejemplo, familias y personas, o número de fincas y hectáreas (o acres) cultivadas.

El conjunto de elementos representativos de una población, de los que se toma información en el proceso de muestreo, se denomina ***muestra***.

***Tamaño Muestral*** es el número de elementos que componen la muestra.

La ***unidad de Muestreo*** es una unidad seleccionada del ***marco muestral***, puede ser la unidad de análisis aún cuando no es necesario, si se trata del caso de un simple elemento de la población estamos ante una ***unidad elemental de muestreo***, pero también pueden considerarse unidades de muestreo que sean grupos no solapados (de intersección vacía) de elementos de la población que cubren la población completa (población objetivo), en cuyo caso estamos ante una ***unidad de muestreo compuesta*** de varias unidades elementales, también denominada ***unidad primaria.***

Ahora sería necesario plantear una jerarquía de las unidades de muestreo en que el primer nivel estaría ocupado por las unidades elementales, el segundo nivel lo formarían grupos de unidades elementales, el tercer nivel grupos de unidades del segundo nivel, y así sucesivamente. Por lo tanto las unidades de muestreo son grupos no solapados (de intersección vacía) de elementos de la población que cubran la población objetivo, en el caso de que las unidades de muestreo sean elementales, una unidad de muestreo y un elemento de la población son idénticos.

La población investigada para este trabajo es la que habita en las viviendas de las Parroquias: Sucre, García Moreno, Nueve de Octubre y Urdaneta, las mismas que se ubican en el centro sur de la Ciudad de Guayaquil. Se excluye, pues; el sector industrial, empresas y negocios, que no formarán parte del estudio así como los que habitan en viviendas colectivas como Hoteles, Pensiones, Residenciales, Cuarteles Militares, Cuarteles de Policía, Hospitales, Clínicas, Conventos o Instituciones religiosas.

La unidad de muestreo que se utilizó en este estudio fueron las viviendas particulares de cada una de las parroquias en estudio.

El ***Marco Muestral*** es la totalidad de las unidades de muestreo de donde se extraerá la muestra, el marco de muestreo puede ser una lista material de unidades de muestreo, por ejemplo; para obtener información acerca de personas podríamos usar una lista completa de un censo, o un registro de personas y seleccionar directamente una muestra de personas. Sin embargo, también podríamos seleccionar una muestra de familias e incluir en la encuesta todas las personas de las familias seleccionadas, en forma similar podríamos seleccionar edificios completos e incluir todas las personas que viven en las estructura seleccionadas.

Lo ideal sería disponer de un marco tal que la lista de unidades muestrales que lo componen coincida con la población objetivo, pero en la práctica esto no sucede ya que el marco contiene impurezas debidas a desactualizaciones, duplicaciones de algunas unidades de muestreo, omisiones de otras y presencia de **unidades extrañas y vacías**, factores que hacen que el marco no coincida con la población objetivo, lo que no impide que el marco sea la contrapartida en el mundo real de la población objetivo.

Sin embargo la separación entre el marco y la población objetivo ha de ser lo suficientemente pequeña como para permitir que se hagan inferencias acerca de la población basándose en una muestra obtenida del marco.

***Unidades vacías*:** son unidades de muestreo erróneamente incluidas en el marco y que no pertenecen a la población objetivo.

***Unidades extrañas*:** son Unidades que aparecen en el marco pero que no son realmente del marco y que no se relacionan con la población objetivo.

La presencia de unidades vacías y extrañas se produce principalmente en dos tipos de situación práctica que son conceptualmente equivalentes. En la primera situación, la lista, por no estar actualizada, incluye unidades que han dejado de pertenecer al colectivo que se desea muestrear. En la segunda situación la población que se desea muestrear es una subpoblación de la cubierta por el marco.

Una forma de solucionar estos inconvenientes es eliminar del marco las unidades erróneamente incluidas (unidades extrañas, vacías y duplicaciones) y a su vez incluir las omisiones, a este proceso se lo llama ***Depuración de marcos imperfectos.***

Cuando se tiene el listado de unidades compuestas y adicional a éste el listado de unidades simples dentro de cada unidad compuesta, se está haciendo uso de los ***marcos múltiples***

Por ejemplo, los residentes de una ciudad pueden ser muestreados de una lista de manzanas de la ciudad relacionada con una lista de residentes dentro de las manzanas, el segundo marco puede no estar disponible hasta que las manzanas sean seleccionadas y estudiadas con cierto detalle.

El marco muestral que se usará para el análisis es la cartografía de las Parroquias Urbanas obtenidos en el Censo de Población y Vivienda –1990, fragmentados en zonas censales, sectores censales y manzanas, planos de la Ciudad de Guayaquil –1990 estructurado de 14 Parroquias y 246 zonas censales, proporcionadas por el INEC.

**4.3. Etapas de un Estudio por Muestreo.**

Los pasos más importantes en todo estudio de muestreo, son análogos, en general.

* La fijación del universo del que se tomará la muestra.
* Determinar cuál ha de ser la unidad de muestra. Para esto, hay que dividir el colectivo en sus partes iguales, siendo cada una de ellas la unidad de muestreo. En muchos casos la unidad es obvia; por ejemplo, si se trata de tomar muestras de una población, la unidad del muestreo puede ser cada habitante, cada familia o cada una de las manzanas de que consta la población.
* Determinar el tamaño de la muestra. Para este paso hay que tomar en cuenta la estimación del costo del estudio, para evitar un fracaso por causa de error en el presupuesto.
* Determinar el método a seguir para selección de los elementos que han de integrar a la muestra.

Son cuatro los procedimientos que pueden seguirse para la obtención de la unidades del colectivo que han de formar una muestra:

* Escoger cuidadosamente del universo, cada una de esas unidades.
* Obtener al azar las unidades que han de integrarse a la muestra.
* Formar la muestra combinando los dos procedimientos anteriores.
* El muestreo por conglomerados, consiste en tomar todos los elementos de la muestra, con cualesquiera de los tres procedimientos anteriores, de un colectivo que se ha dividido geográficamente.

**4.4. Tipos de Muestreo**

Existen diferentes planes de muestreo de acuerdo a las condiciones que presenta la población. Por cada plan que es considerado, se pueden hacer estimaciones de tamaño partiendo del conocimiento del nivel de precisión deseado. Los costos relativos y el tiempo involucrado para cada plan también son comparados antes de tomar la decisión.

**Muestreo Probabilístico**

El método otorga una probabilidad conocida de integrar la muestra a cada elemento de la población, y dicha probabilidad no es nula para ningún elemento.

Los métodos de ***muestreo no probabilisticos*** no garantizan la representatividad de la muestra y por lo tanto no permiten realizar estimaciones inferenciales sobre la población.

Entre los métodos de muestreo probabilísticos más utilizados en investigación encontramos:

* Muestreo Aleatorio Simple
* Muestreo Estratificado
* Muestreo Sistemático
* Muestreo Polietápico o por conglomerados

Entenderemos por representatividad, a la medida en que la muestra es imagen de la población bajo estudio. Sea **N** el numero de unidades de la población (tamaño de la población), y **n** el número de unidades de la muestra (tamaño muestral).

**4.4.1. Muestreo Aleatorio Simple o sin reposición (M.A.S.).**

Este es un método de selección de n unidades sacadas de N, de tal manera que cada una de las muestras tienen la misma probabilidad de ser seleccionada pero al ser la población finita, la probabilidad de que salga un elemento dependerá de los que fueron separados anteriormente y dejaron, por tanto de pertenecer a los seleccionables.

**4.4.2. Muestreo Aleatorio con Reposición.**

En este tipo de muestreo todas las muestras y en consecuencia todas las unidades de la población, tienen la misma probabilidad de ser seleccionadas para formar parte de la muestra.

##### 4.4.3. Muestreo Aleatorio Simple para Proporciones

La estimación de la proporción de unidades que tienen una característica constituye una clase importante de estadística para la que resultan particularmente sencilla las fórmulas de la varianza y de la determinación del tamaño de la muestra.

##### Tipos de estadísticas para las que se usan proporciones

En el análisis estadístico las proporciones pueden presentarse en dos formas:

Primero, interés en conocer más una proporción que un total o un promedio, por ejemplo la proporción de desempleados en la población, el porcentaje de familias con ingresos mayores que una suma dada.

Segundo, podemos desear clasificar a una población en ciertos grupos diferentes y saber que porcentajes de la población constituye cada grupo.

**4.4.4. Muestreo Aleatorio Estratificado.**

Con el 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 uno de los cuales se hace una selección aleatoria, como en los casos anteriores. 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 empleando dentro de cada uno de los estratos.

Los estratos pueden ser de igual o diferente tamaño, en este ultimo caso, una manera posible de determinar el tamaño de la muestra al interior de cada estrato, es proporcional al tamaño de la misma. Este tipo de asignación se conoce como afijación proporcional, la que no siempre resulta ser la afijación más conveniente por cuanto los costos de muestreo en cada uno de los estratos pueden ser distintos. Otro tipo es dependiendo de la variabilidad que presente la población.

# 4.4.5. Muestreo Sistemático

Este método de selección es bastante diferente a los anteriores. Se usa cuando los individuos de la población están de algún modo totalmente ordenados. Para seleccionar una muestra se aprovecha la ordenación de los individuos.

Este método consiste en tomar los elementos poblacionales que formarán la muestra de k en k a partir de uno de ellos elegido aleatoriamente entre los que ocupan el lugar primero y el k-ésimo, ambos inclusive. Una aproximación del mismo puede conseguirse en el muestreo de ficheros sustituyendo el recuento de k en k, por la medida efectuada en una reglilla graduada que se lleva a lo largo de la población para seleccionarlos: p. ej, de decímetro en decímetro. Cuando el orígen no es aleatorio, sino que se toman los elementos centrales de cada grupo de k consecutivos, el muestreo puede denominarse rígido o estrictamente sistemático y deja de ser probabilístico.

**4.4.6. Muestreo por Conglomerado**

El método consiste en sustituir las unidades físicas, elementales o últimas, a las cuales se refiere el estudio, por unidades de muestreo que comprendan un grupo de aquellas. Como la selección es de conglomerados, interesa que cada uno de ellos represente en lo posible a la población y, por tanto, que esté constituido por unidades últimas heterogéneas. Con frecuencia estos conglomerados son superficies o áreas en que se han dividido el terreno o ámbito ocupado por la población en estudio.

Hay dos razones principales para la extensa aplicación de este plan de muestreo: falta de una lista confiable de elementos de en la población y consideraciones de tipo económica.

#### 4.5. Criterios de aceptación de un Método de Muestreo

Los métodos de muestreo modernos pueden proporcionar datos de confiabilidad conocida en forma eficaz y económica, si bien es cierto que una muestra es una parte de una población, implicaría tener un concepto equivoco el llamar “muestra” a cualquier conjunto de números simplemente porque se trata de una parte de una población. Para que una muestra sea aceptable es necesario desde el punto de vista del análisis estadístico que represente a la población, que tenga una confiabilidad susceptible de medición y que responda a un plan práctico y eficaz.

##### Probabilidad de selección de cada unidad

Se debe atribuir a cada unidad una probabilidad de ser elegida la que deberá ser siempre distinta de cero.

##### Confiabilidad susceptible de medir

La confiablidad de las estimaciones derivadas de la muestra debe ser susceptible de medir, es decir la muestra además de dar las estimaciones de las características de la población debe proporcionar medidas de la precisión de tales estimaciones, dichas medidas se podrán usar para determinar el error máximo que puede esperarse es esas estimaciones si el procedimiento se cumple en forma específica y la muestra es moderadamente grande.

##### Viabilidad

Otra característica es que el plan de muestreo sea lo suficientemente simple y directo.

##### Economía y Eficiencia

Un diseño debe ser eficiente, es decir será aquel que en la medida de nuestro conocimiento sea capaz de producir la mayor cantidad de información al menor costo, es decir se hará el uso más efectivo de todos los recursos y medios disponibles, por ejemplo, mapas información estadística al alcance, preparación que posee el personal, teoría del muestreo, etc.

Siguiendo estos criterios empezamos entonces determinando el método de muestreo a usar.

#### 4.6. Técnica de selección de muestra a utilizarse

Una técnica o método de muestreo es un proceso o mecanismo mediante el cual se seleccionan las muestras de modo que cada una tenga una determinada probabilidad de ser elegida, por lo tanto el método aleatorio empleado para seleccionar la muestra define el espacio muestral S una función de probabilidad P tal que:

P(Si) ≥ 0 ∀i

ΣsP(S) = 1

En general puede ocurrir que no todas las muestras del espacio muestral sean no elegidas.

El método de muestreo a usar siguiendo cada una de las fases anteriormente mencionadas es Muestreo Aleatorio Simple para Proporciones.

##### 4.6.1. Razones por las que se utiliza la técnica seleccionada

Para este trabajo deseamos estimar el número total, la proporción o el porcentaje de unidades en la población que poseen alguna característica o atributo o que caen dentro de alguna clase definida. se trata de una población cuyas unidades son uniformes ya que las características que poseen son muy parecidas, empezando con la delimitación geográfica pues se ubican en un nivel socioeconóminco equilibrado y situado en el centro de la Ciudad por lo que se entiende que los servicios básicos que se brindan a las parroquias agrupadas son muy semejantes. Es por esta razón que convenientemente se las agrupó denominándolas Sector Centro Sur-Oeste, para lo que se decide usar Muestreo Aleatorio Simple por Proporciones, por tratarse de un muestreo que proporciona igual probabilidad de selección para cada uno de los entes de la población, además de la confiabilidad de las estimaciones derivadas de esta y por tratarse de un método de muestreo que posee un diseño eficiente y que incurre un bajo coste.

##### En este caso este tipo de muestreo es el ideal para obtener los resultados requeridos, ya que lo que se quiere es medir el nivel de satisfacción de los servicios básicos, las variables expuestas en el cuestionario miden estas características al procesar los datos podremos obtener, por ejemplo, el porcentaje de viviendas del sector Centro Sur-Oeste que posee medidor de energía eléctrica o el porcentaje de personas por viviendas de la Parroquia Urdaneta, que han presentado algún reclamo ante la Empresa de Telefonía Fija.

##### 4.7. Diseño de la Muestra

En la práctica un muestreo aleatorio simple se realiza unidad por unidad, se enumeran las unidades de 1 a N, posteriormente se extraen una serie de n números aleatorios o mediante un programa de computación que genere una tabla semejante, en cada extracción el proceso debe otorgar la misma oportunidad de selección a todos y a cada uno de los números que no hayan salido., las unidades que llevan estos no constituyen la muestra.

Para tomar una muestra dentro de la población objetivo se hizo uso de la cartografía correspondiente a las parroquias, estructurada por manzanas, a pesar de que estos mapas corresponden al último Censo de Población y Vivienda de 1990, ya que actualmente se está empezando con el diseño del censo que se efectuará casi a fines del 2001, no habría ningún inconveniente puesto que geográficamente el centro de la ciudad no ha experimentado cambios sustanciales.

**4.7.1. Selección de la Unidad de Muestreo**

A continuación se detalla las actividades que se realizaron para la selección de las unidades de muestreo.

1. Obtención del marco muestral, es decir, la cartografía de cada una de las Parroquias en estudio con el mayor número de detalles acerca de las manzanas. Si el mapa no es reciente, se deben tomar las medidas convenientes, a través de consultas en la misma localidad, para actualizarlos (por ejemplo, marcar nuevas calles que se hayan podido abrir desde la fecha en que se imprimió el mapa).
2. Enumerar en serie las manzanas, anotando los números directamente en el mapa; se aconseja adoptar un sistema de numeración serpentino a fin de que no se omita alguna manzana.
3. Generar números aleatorios por medio de la función Random.
4. Seleccionar una muestra simple al azar de manzanas utilizando una muestra del 0.525%.
5. Particionar cada una de las manzanas muestrales en 4 segmentos. Si se cuenta con mapas que muestren la estructura interna de cada manzana (pasajes, edificios, etc.), pueden usarse los mismos. Si no es así, preparar un bosquejo rápido y esquemático de las manzanas muestrales señalando cada edificio; usar ese bosquejo como base de la segmentación. Los 4 segmentos dentro de una manzana deben tener aproximadamente el mismo número de unidades de vivienda.
6. Enumerar los segmentos en cada manzana de 1 a 4.
7. Seleccionar los segmentos muestrales tomando un número aleatorio entre 1 y 4 para cada manzana.
8. Entrevistar una vivienda de cada uno de los segmentos muestrales.
9. En el caso de edificios con varios departamentos se lo contará como una sola vivienda y si el edificio resulta seleccionado para la encuesta el departamento sería escogido al azar, habiendo generado previamente números aleatorios por medio de la función random. Los pisos se contarán de abajo hacia arriba, los departamentos se contarán de izquierda a derecha, en caso de que se presentase un piso con más de dos departamentos.
10. En caso de tener falta de respuesta luego de los intentos establecidos, imputar este dato por la vivienda que pertenezca al grupo de unidades que fueron seleccionadas para reserva.

##### 4.8. Diseño e Implementación de la Encuesta Preliminar

Cuando se realizan encuestas de gran dimensión suele ser muy útil seleccionar una pequeña muestra para una prueba piloto, esta prueba piloto puede ser crucial ya que permite probar en campo el cuestionario y otros métodos de medición, calificar a los encuestadores y verificar el manejo de las operaciones generales de campo.

De la encuesta preliminar también se puede obtener estimaciones de determinadas características poblacionales que pueden utilizarse posteriormente en cálculos sobre tamaños muestrales y estimaciones de los errores de muestreo.

“El uso del submuestreo descrito en el procedimiento de selección de las unidades de muestreo puede considerarse como esencialmente equivalente a usar una muestra de pequeños conglomerados de unidades de vivienda (en los que se enumeraría cada unidad de vivienda) pero como un muestreo bietápico como un recurso para reducir la cantidad de trabajo exigida por la extracción de una muestra de pequeños conglomerados”.

Los resultados de la encuesta piloto siempre sugieren modificaciones en la planificación en la encuesta general que van a mejorar la calidad de los resultados de la encuesta a escala completa.

**4.8.1. Características críticas de una encuesta piloto:**

* Ensayar el cuestionario en condiciones reales.
* Probar los aspectos principales de la encuesta real.
* Contrastar la idoneidad del marco.
* Resaltar la variabilidad de determinados caracteres
* Intuir la tasa esperada de falta de respuesta.
* Comprobar que el método de recogida de datos sea el adecuado.
* Contribuir con datos sobre el probable coste y duración de la encuesta real.

El cuestionario es el medio de comunicación entre el que solicita los datos y el respondiente.

Es además el instrumento de trabajo para la posterior codificación de la información. Ha de estar, por tanto, estructurado convenientemente en secciones y preguntas, siguiendo un orden o secuencia preestablecida para que sea fácilmente manejable y codificable informáticamente. Además, es conveniente que el cuestionario mantenga en todo momento el interés del encuestado, siendo el vocabulario utilizado adecuado a su nivel. Por otra parte, el cuestionario ha de diseñarse para que la entrevista no supere la duración de una hora.

Antes de realizar la encuesta real se debe realizar una encuesta piloto, para la cual se elaborará un cuestionario piloto, el cual nos indicará posibles errores que se podrían cometer durante la encuesta real.

**4.9. Estimación del Tamaño Muestral**

Para calcular el tamaño de una muestra hay que tomar en cuenta tres factores:

1. El porcentaje de confianza con el cual se quiere generalizar los datos desde la muestra hacia la población total.
2. El porcentaje de error que se pretende aceptar al momento de hacer la generalización.
3. El nivel de variabilidad que se calcula para comprobar la hipótesis.

La ***confianza*** *o el* ***porcentaje de confianza*** es el porcentaje de seguridad que existe para generalizar los resultados obtenidos. Esto quiere decir que un porcentaje del 100% equivale a decir que no existe ninguna duda para generalizar tales resultados, pero también implica estudiar a la totalidad de los casos de la población.

Para evitar un costo muy alto para el estudio o debido a que en ocasiones llega a ser prácticamente imposible el estudio de todos los casos, entonces se busca un porcentaje de confianza menor.

Comúnmente en las investigaciones sociales se busca un 95% y es precisamente este el porcentaje de confianza que se usará para el análisis.

El ***error*** *o* ***porcentaje de error*** equivale a elegir una probabilidad de aceptar una hipótesis que sea falsa como si fuera verdadera, o la inversa: rechazar a hipótesis verdadera por considerarla falsa. Al igual que en el caso de la confianza, si se quiere eliminar el riesgo del error y considerarlo como 0%, entonces la muestra es del mismo tamaño que la población, por lo que conviene correr un cierto riesgo de equivocarse.

Comúnmente se aceptan entre el 4% y el 6% como error, tomando en cuenta de que **no** son complementarios la confianza y el error.

La ***variabilidad*** es la probabilidad (o porcentaje) con el que se aceptó y se rechazó la hipótesis que se quiere investigar en alguna investigación anterior o en un ensayo previo a la investigación actual.

El porcentaje con que se aceptó tal hipótesis se denomina ***variabilidad positiva*** y se denota por ***p*,** y el porcentaje con el que se rechazó se la hipótesis es la ***variabilidad megativa***, denotada por ***q*.**

Hay que considerar que *p* y *q* son complementarios, es decir, que su suma es igual a la unidad: *p* + *q* =1. Además, cuando se habla de la máxima variabilidad, en el caso de no existir antecedentes sobre la investigación (no hay otras o no se pudo aplicar una prueba previa), entonces los valores de variabilidad es *p*=*q*=0.5.

Una vez que se han determinado estos tres factores, entonces se puede calcular el tamaño de la muestra como a continuación se expone.

### 4.9.1. Distribución Muestral de Proporciones.

Supóngase una población infinita y que la probabilidad de que ocurra un suceso (conocido como un éxito) es p, mientras que la probabilidad de que no ocurra (fracaso) es q = 1 - p. Por ejemplo, la población podría ser todos los posibles lanzamientos de una moneda, en este caso, la probabilidad del suceso "cara", es : p=1/2.

Se consideran todas las posibles muestras de tamaño N extraidas de esta población y para cada muestra se determina la proporción P de éxitos. En el caso del lanzamiento de la moneda. P sería la proporción de caras obtenidas en N lanzamientos. Entonces se obtiene una distribución normal de las P de cada muestra, que se llama distribución normal de proporciones cuya media y desviación típica son:



**4.9.2. Tamaño de Muestra para estimar la Proporción de la Población**

Para calcular el tamaño de muestra para la estimación de proporciones poblacionales, se pueden usar dos fórmulas:

La primera se la aplica en el caso de que **no se conozca con precisión el tamaño de la población**, y es:



donde:

*n:* es el tamaño de la muestra;
*Z*: es el nivel de confianza;
*p:* es la variabilidad positiva;
*q*: es la variabilidad negativa;
*E:* es la precisión o error.

En el caso de que **sí se conozca el tamaño de la población** entonces se aplica la siguiente fórmula:



donde

*n:* es el tamaño de la muestra

: corresponde al nivel de confianza elegido
P: proporción de una categoría de la variable
e: es la precisión o el error máximo
N: tamaño de la población

La ventaja sobre la primera fórmula es que al conocer exactamente el tamaño de la población, el tamaño de la muestra resulta con mayor precisión y se pueden incluso ahorrarse recursos y tiempo para la aplicación y desarrollo de una investigación, para el análisis haremos uso de la segunda fórmula.

La muestra piloto representa el 0.525% del total de la población, la misma que se constituye del grupo de Parroquias que se muestran en la Tabla IV, el total de viviendas de este grupo es de 29373, se obtuvo de forma proporcional al tamaño de cada parroquia agrupada, en el Sector denominado Centro Sur-Oeste, el tamaño de la muestra piloto es de 155 viviendas, para llegar a las viviendas se siguieron los pasos que se detallaron en la Selección de Unidades de Muestreo.

Las 155 viviendas fueron tomadas de la población como lo muestran los pasos de Selección de la unidad de Muestreo.

**Tabla III**

**Distribución del Tamaño Poblacional**

Fuente: INEC

|  |  |
| --- | --- |
| PARROQUIAS | **TOTAL DE VIVIENDAS**  |
| **Sucre****García Moreno****9 de Octubre****Urdaneta**Total Sector Centro Sur-Oeste | 4.87414.1842.7187.597**29.373** |

4.10. Determinación de la Muestra Final

Se Tomó como variable de interés la proporción de personas que calificaron el servicio de recolección de basura como bueno, ya que ésta fue la opción de mayor proporción.

La fórmula que se usará es:



Donde:

P = 0.54 Proporción de personas que calificaron al servicio de recolección de basura como Bueno.

Q = 0.46 Proporción de personas que no calificaron al servicio de recolección de basura como Bueno.

E = 0.0419 Límite deseado de error para la media estimada

N = 29373 Tamaño de la población: Total de viviendas

Entonces se tiene:

El tamaño de la muestra real es 534, pero para tomar las unidades de muestreo más dispersos se procede a distribuir este tamaño de forma proporcional entre todas las parroquias que conforman el Sector Centro Sur-Oeste, usando la siguiente fórmula:

Los tamaños de las muestras para cada una de las parroquias resultan tal como se observa en la siguiente tabla:

**TABLA IV**

**Distribución del Tamaño Muestral**

|  |  |  |
| --- | --- | --- |
| PARROQUIAS | Ni | **ni** |
| **Sucre****García Moreno****9 de Octubre****Urdaneta**Total | 4.87414.1842.7187.567**29.373** | 8925849138**534** |