**CAPITULO 4**

4. PRINCIPIOS DE LAS TÉCNICAS DE MUESTREO UTILIZADO.

En toda investigación estadística existe un conjunto de elementos sobre los que se toma información. Este conjunto de elementos es lo que se denomina población o universo estadístico. Cuando el estadístico o el investigador toma información de todos y cada uno de los elementos de la población estadística, se dice que está realizando un censo. Sin embargo, esto no es muchas veces posible, ya sea por el costo alto que resulta de la toma de información, o que la población tenga infinitos elementos, o por otras causas. Este problema lleva al investigador a tomar la información sólo de una parte de los elementos de la población estadística, proceso que recibe el nombre de muestreo.

**4.1 Concepto de Muestreo.**

La teoría del muestreo proporciona una técnica estadística de carácter muy práctico, que sencillamente busca obtener datos de una población (hogares, empresas, árboles, etc) en su totalidad, utilizando tan solo una parte reducida de la misma, denominada muestra.

El conjunto de elementos de los que se toma la información en el proceso de muestreo se llama ***muestra*** y el número de elementos que la componen ***tamaño muestral***. Existen varios tipos de muestreo, dependiendo de que la población estadística sea finita o infinita. El investigador utiliza la muestra para la toma de información, pero lo importante es que dicha muestra sea representativa. Por lo tanto, entenderemos por muestra a todo subconjunto con lo más representativo posible de una población.

**4.2 Las Ventajas del Método por Muestreo.**

En la mayoría de las aplicaciones para las que se edificó esta teoría, el conjunto del que se desea obtener información es finito y bien delimitado, como sería por ejemplo el conjunto de habitantes de una ciudad, las máquinas de una fábrica o los peces de un lago.

A continuación se mencionarán las ventajas que se obtienen al emplear encuestas por muestreo:

### Si los datos se obtienen únicamente de una pequeña fracción del total, los gastos son menores que los que se realizarán si se lleva a cabo un censo completo.

### Por la misma razón, los datos pueden ser recolectados y resumidos más rápidamente con una muestra que con una enumeración completa.

* Debido a que al reducir el volumen de trabajo se puede emplear personal más capacitado y someterlo a un entrenamiento intensivo y debido también a que en estas condiciones será factible la supervisión cuidadosa del trabajo de campo y del procesamiento de los resultados, una muestra puede producir ***resultados más exactos*** que la enumeración completa.

**4.3 Etapas Principales en una Encuesta por Muestreo.**

Los pasos principales en una encuesta están agrupados más o menos arbitrariamente bajo 11 encabezados y se citan a continuación:

### 4.3.1 Objetivos de la encuesta

Una exposición clara de los objetivos es lo más útil. Sin esto, es fácil olvidarlos en una encuesta compleja al preocuparse por los detalles de la planeación y por lo tanto tomar decisiones que varían de los objetivos.

### 4.3.2 Población bajo muestreo

La palabra ***población*** se emplea para denominar el conjunto del que se elige la muestra.

La población que se muestrea ( la población muestreada ) debe coincidir con la población sobre la cual se desea información ( la población objetivo ). En ocasiones, por razones de factibilidad, o simple conveniencia, la población muestreada es más restringida que la población objetivo.

### 4.3.3 Los datos recogidos

Es conveniente cerciorarse que todos los datos son pertinentes a la encuesta y que no se omiten datos esenciales. Un cuestionario demasiado largo produce una baja general de la calidad de las respuestas, tanto a las preguntas importantes como a las otras.

### 4.3.4 Grado de precisión deseado

Los resultados de una encuesta por muestreo están siempre sujetos a cierta incertidumbre, por que sólo se mide una parte de la población, y por los errores en las mediciones realizadas. La especificación del grado de precisión deseado es responsabilidad de la persona que va a utilizar los datos y puede presentar dificultades, porque los administradores no están acostumbrados a pensar en términos de la magnitud del error tolerable en las estimaciones, compatible con una buena decisión.

### 4.3.5 Métodos de medición

La encuesta puede emplear un cuestionario autoadministrado, o un proceso de entrevistas en la que los entrevistadores simplemente leen un cuestionario prescrito, o bien, un proceso en el que se permite mucha libertad en la forma y el orden de las preguntas. En cuestionarios sencillos a veces es posible precodificar las respuestas; es decir, colocarlas de tal modo que puedan transferirse rutinariamente a un equipo mecánico.

### 4.3.6 El Marco

Antes de seleccionar la muestra, la población debe ser dividida en partes llamadas ***unidades de muestreo o unidades***. Estas deben cubrir la totalidad de la población y no traslaparse en el sentido de que todo elemento de la población pertenezca a una y solamente a una unidad. En el muestreo de los residentes de una ciudad, la unidad puede ser una persona, los miembros de una familia, o las personas que viven en una manzana.

Frecuentemente, la construcción de esta lista de unidades de muestreo, llamada marco, es uno de los principales problemas prácticos.

### 4.3.7 Selección de la muestra

Existe, actualmente, una gran variedad de planes para seleccionar una muestra. Por cada plan considerado, se pueden hacer, a groso modo, estimaciones del tamaño de la muestra, partiendo de un conocimiento del nivel de precisión deseado.

### 4.3.8 La encuesta piloto

Es de gran utilidad probar el cuestionario y los métodos de campo en pequeña escala. Esto casi siempre da por resultado mejoras al cuestionario y puede evitar otros problemas que serían serios a mayor escala, por ejemplo, que el costo fuera mucho mayor que el esperado.

### 4.3.9 Organización del trabajo de campo

El personal debe recibir un entrenamiento sobre el propósito de la encuesta y los métodos de medición que se emplearán. Además, se debe supervisar adecuadamente su trabajo.

### 4.3.10 Resumen y análisis de los datos

El primer paso después de realizar la encuesta es el editar los cuestionarios obtenidos, con la esperanza de corregir errores o cuando menos desechar los datos que obviamente están equivocados. Después se realizarán los cálculos que conduzcan a las estimaciones. Puede haber diferentes métodos de estimación para los mismos datos.

### 4.3.11 Información conseguida para encuestas futuras

Cuanta más información de una población se tenga inicialmente, más fácil será el diseño de una muestra que proporcione estimaciones exactas. Toda muestra obtenida es una guía potencial de futuros muestreos. Un muestreador hábil aprende a reconocer los errores de ejecución y a evitar que se repitan.

### 4.4 La estimación del Tamaño de la Muestra.

Al planear una encuesta por muestreo, siempre se alcanza una etapa en donde hay que tomar una decisión respecto al tamaño de la muestra. Esta decisión es importante. Una muestra demasiado grande implica un despilfarro de recursos y una muy pequeña disminuye la utilidad de los resultados. La decisión no siempre puede tomarse satisfactoriamente; a menudo no disponemos de la suficiente información para saber si el tamaño de la muestra seleccionada, es el óptimo. La teoría del muestreo proporciona un marco dentro del cual se puede pensar inteligentemente respecto a este problema.

### 4.5 La Especificación de la Precisión.

La precisión deseada se puede establecer, al definir la cantidad de error tolerable en las estimaciones muestrales. En ocasiones, es difícil saber que tanto error debería tolerarse, particularmente cuando los resultados se destinan a varios fines. De modo que el límite de error que se escoge es en cierta medida arbitrario.

Parte de la dificultad es que no se sabe lo suficiente sobre las consecuencias que producirían errores de diferentes magnitudes en las decisiones prácticas que se toman con base a los resultados de la encuesta.

Aun cuando estas consecuencias se conozcan, los resultados de muchas encuestas importantes, los utilizan diferentes personas con diversos propósitos, y algunos de ellos no se han previsto al planear la encuesta.

### 4.6 Métodos de Muestreo.

Podemos definir los ***métodos de muestreo*** como el 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 y acotable***.

**4.6.1 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.

**4.6.2 Muestreo aleatorio sin reposición o muestreo aleatorio simple**.

Como en el método anterior, todas las unidades de la población tienen la misma probabilidad de que salgan; 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.

**4.6.3 Muestreo 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.

**4.6.4 Muestreo por conglomerados o áreas.**

Consiste en un método de sustituir las unidades físicas o elementales, 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 ha dividido el terreno o ámbito ocupado por la población en estudio.

**4.6.5 Muestreo sistemático.**

Este método cosiste 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: por ejemplo, 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 suele denominarse rígido o estrictamente sistemático y deja de ser probabilístico.

**4.7 Muestreo Aleatorio Estratificado.**

En el muestreo estratificado, una ***población heterogénea*** con  unidades  se subdivide en  ***subpoblaciones, lo más homogéneas posibles,*** no solapadas denominadas ***estratos***  de tamaños . La muestra estratificada de tamaño  se obtiene seleccionando  elementos  de cada uno de los  estratos en que se subdivide la población de forma independiente. Si la muestra estratificada se obtiene seleccionando una muestra aleatoria simple en cada estrato de forma independiente, el muestreo se denomina ***muestreo aleatorio estratificado***, pero en general, nada impide utilizar diferentes tipos de selección en cada estrato.

Para un estrato en particular pueden pertenecer todas sus unidades a la muestra, parte de ellas o ninguna. También puede ocurrir que para formar la muestra estratificada se obtengan elementos de todos los estratos o sólo parte de ellos. Si sabemos seguro que un determinado estrato aporta unidades para la muestra, dicho estrato se denomina ***estrato correpresentado***. Por otra parte, las unidades de la población que con certeza van a pertenecer a la muestra se denominan ***autorepresentadas***.

Podemos representar gráficamente la población dividida en  estratos de tamaño  de cada uno de los cuales seleccionamos de modo independiente  unidades ( mediante muestreo aleatorio simple si no se especifica otra cosa ) para la muestra estratificada de tamaño .

Podemos expresar la formación de estratos en la población y la formación de la muestra estratificada de la forma siguiente:

**POBLACIÓN**



**MUESTRA**



**4.7.1 Razones para el uso de Muestreo Estratificado**

Son diversos los motivos que aconsejan efectuar una partición de nuestra población  en  subpoblaciones, no solapadas, , entre los que destacan los siguientes:

* El muestreo estratificado puede aportar información más precisa de algunas subpoblaciones que varían bastante en tamaño y propiedades entre sí, pero que son homogéneas dentro de sí.
* El uso adecuado del muestreo estratificado puede generar ganancia en precisión, pues al dividir una población heterogénea en estratos homogéneos, el muestreo en estos estratos tiene poco error debido precisamente a la homogeneidad.
* En otros casos la estratificación viene motivada por el requerimiento de estimaciones para ciertas áreas o regiones geográficas. En esta situación cada estrato será un área compacta, como por ejemplo un municipio, una provincia, una colonia de una ciudad, etc.
* También es una razón para utilizar muestreo estratificado la existencia de una variable precisa para la estratificación cuyos valores permitan dividir convenientemente la población en estratos homogéneos. Las variables utilizadas para la estratificación deberán estar correlacionadas con las variables objeto de la investigación. Por ejemplo, si se quiere realizar estadísticas en el sector educativo puede utilizarse la variable de estratificación nivel de enseñanza, tomando como estratos los niveles de enseñanza infantil, enseñanza primaria, enseñanza secundaria obligatoria, bachillerato y enseñanza universitaria (cada estrato tiene así unas características muy peculiares que lo hacen homogéneo).

**4.7.2 Afijación de la Muestra**

Se llama afijación de la muestra al reparto, asignación, adjudicación, adscripción o distribución del tamaño muestral  entre los diferentes estratos. Esto es, a la determinación de los valores de  que verifiquen . Pueden establecerse muchas afijaciones o maneras de repartir la muestra entre los estratos, pero las más importantes son: la afijación uniforme, la afijación proporcional, la afijación de varianza mínima y la afijación óptima.

**4.7.2.1 Afijación Uniforme**

Consiste en asignar el mismo número de unidades muestrales a cada estrato, con lo que se tomarán todos los  iguales a , aumentando o disminuyendo este tamaño en una unidad si  no fuese múltiplo de , esto es .



Este tipo de afijación da la misma importancia a todos los estratos, en cuanto a tamaño de la muestra, con lo cual favorecerá a los estratos de menor tamaño y perjudicará a los grandes en cuanto a precisión. Sólo es conveniente en poblaciones con estratos de tamaño similar.

**4.7.2.2 Afijación Proporcional**

Consiste en asignar a cada estrato un número de unidades muestrales proporcional a su tamaño. Las  unidades de la muestra se distribuyen proporcionalmente a los tamaños de los estratos expresados en número de unidades. Tenemos:







A la vista de los resultados anteriores podemos asegurar lo siguiente:

* Las fracciones de muestreo en los estratos son iguales y coinciden con la fracción global de muestreo, siendo su valor la constante de proporcionalidad.
* Los coeficientes de ponderación  se obtienen exclusivamente a partir de la muestra, pues para su cálculo sólo son necesarios valores muestrales .
* Como  todas las unidades de la población tienen la misma probabilidad de figurar en la muestra de  unidades, es decir, estamos en el caso de *muestras autoponderadas*.

**4.7.2.3 Afijación de Mínima Varianza (o Afijación de Neyman)**

La afijación de mínima varianza, o afijación de Neyman, consiste en determinar los valores de  (número de unidades que se extraen del estrato h-ésimo para la muestra) de forma que para un tamaño de muestra fijo, igual a , la varianza de los estimadores sea mínima.

Donde:



Otra expresión para  es :



Vemos que los valores de  son proporcionales a los productos  y en el supuesto de que , esta afijación de mínima varianza coincidiría con la proporcional tal y como se ve a continuación:



**4.7.2.4 Afijación Óptima**

La afijación óptima consiste en determinar los valores de  (número de unidades que se extraen del estrato h-ésimo para la muestra) de forma que para un coste fijo C la varianza de los estimadores sea mínima. El coste fijo C será la suma de los costes derivados de la selección de las unidades muestrales de los estratos, es decir, si  es el coste de unidad por muestreo en el estrato , el coste total de selección de las  unidades muestrales en este estrato será . Sumando los costes  para los  estratos tenemos el coste total de selección de la muestra estratificada.



Vemos que los valores de  son proporcionales a los productos  y en el supuesto de que  (coste constante en todos los estratos) la afijación óptima coincide con la de mínima varianza, y si además  la afijación óptima coincidirá con la de mínima varianza y con la proporcional.

**4.7.3 Comparación de Eficiencias según los distintos tipos de Afijación.**

Se realizará un estudio comparativo de la conveniencia de los distintos tipos de afijación en términos de su eficiencia medida a través el error de muestreo, o lo que es lo mismo, a través de la varianza. Por lo tanto será más eficiente aquel tipo de afijación que presente menos varianza.

El muestreo estratificado con afijación proporcional es más preciso que el muestreo aleatorio simple, produciéndose la igualdad de precisiones cuando las medias de los estratos son todas iguales. Por lo tanto la ganancia en precisión del muestreo estratificado respecto del aleatorio simple será mayor cuanto más distintas entre sí sean las medias de los estratos, es decir, para que el muestreo estratificado sea preciso es conveniente que los estratos sean heterogéneos entre sí en media.

El muestreo estratificado con afijación de mínima varianza es más preciso que el muestreo estratificado con afijación proporcional, produciéndose la igualdad de precisiones cuando las cuasidesviaciones típicas de los estratos son todas iguales.

Por lo tanto la ganancia en precisión del muestreo estratificado con afijación de mínima varianza respecto del muestreo estratificado con afijación proporcional será mayor cuanto más distintas entre sí sean las cuasidesviaciones típicas de los estratos, es decir, para que el muestreo estratificado sea más preciso es conveniente que los estratos sean heterogéneos entre sí en desviación típica.

En realidad podemos ver que:



Lo que permite asegurar que en general el muestreo estratificado con afijación de mínima varianza es más preciso que el muestreo estratificado con afijación proporcional y que el aleatorio simple, siendo además el estratificado con afijación proporcional más preciso que el aleatorio simple.

**4.7.4 Tamaño de la Muestra**

Vamos a analizar ahora, el tamaño de la muestra estratificada necesario para cometer un determinado error de muestreo conocido de antemano. Distinguiremos los casos de error de muestreo dado con coeficiente de confianza adicional para los tipos de afijación proporcional y de varianza mínima.

**4.7.4.1 Error de muestreo y Coeficiente de Confianza dados  para la media con Afijación Proporcional.**

Tenemos que:



Los tamaños de la muestra en los casos de la estimación de la proporción y el total de clase se calculan sustituyendo  por  en la fórmula del tamaño de la muestra para la estimación de la media.

**4.7.4.2 Error de muestreo y Coeficiente de Confianza dados  para la media con Afijación de Mínima Varianza.**

Tenemos que:



Los tamaños de la muestra en los casos de la estimación de la proporción y el total de clase se calculan sustituyendo  por  en la fórmula del tamaño de la muestra para la estimación de la media.

**4.7.5 Tamaño de la Muestra sin especificar el tipo de Afijación.**

En el caso más general de muestreo estratificado aleatorio, sin especificar el tipo de afijación empleado, tenemos, fijados el error máximo admisible y el coeficiente de confianza , la siguiente ecuación fundamental cuando se trata de estimar la media:



Ya tenemos una fórmula general para calcular el tamaño de muestra en el caso de la estimación de la media para cualquier afijación. Según esta fórmula, para calcular  necesitamos conocer:

* Los tamaños de los estratos, , que nos permiten obtener los coeficientes .
* La precisión prefijada, representada por el error máximo admisible .
* El grado de seguridad o confianza , representado por el coeficiente .
* La variabilidad de cada estrato, representada por la cuasivarianza estratal  y que suele conocerse de una encuesta similar anterior o a través de encuesta piloto.
* El peso  correspondiente a cada estrato en la muestra.
* De aquí se pueden deducir las fórmulas del tamaño muestral  para los distintos tipos de afijación a través del valor que toma .

Para ***Afijación Proporcional:***



Para ***Afijación de Mínima Varianza:***



Si en la fórmula general para n sustituímos estos valores de  para cada tipo de afijación, se obtienen las mismas fórmulas que fueron deducidas anteriormente para el tamaño muestral en el caso de estimación de la media.