

# Tema 1: Fases de una Encuesta por Muestreo

## 1.- Introducción

La investigación por encuestas consiste en establecer reglas que permitan acceder de forma científica a lo que las personas opinan (León y Montero, 1993). Con esta metodología se presentan cuestiones a un conjunto de individuos, de los que se presume que son representativos de su grupo de referencia, para conocer sus actitudes con respecto al tema o temas objeto de estudio. Por tanto, una encuesta no es más que un proceso de recogida de información. No obstante, como dicha recogida de datos conforma una investigación, debe ajustarse a una sistemática que garantice la objetividad de los datos recogidos.

Para recabar la información pertinente de un grupo más o menos numeroso de personas, una forma de proceder puede ser entrevistar a todos y cada uno de los componentes del grupo que se desea estudiar. Sin embargo, resulta obvio que cuando las poblaciones que se quieren estudiar tienen un tamaño medio o grande, obtener información de cada uno de los miembros del grupo resulta inviable, tanto por los costos que conlleva un estudio de tales características, como por el tiempo que requiere recabar una cantidad tan importante de información. Por ello, se entrevista a un subgrupo representativo del colectivo que sea el objetivo de este estudio. Es decir, extraemos la información de una parte de ese grupo, lo que se denomina una muestra, y tales resultados los extrapolamos al resto de la población.

En un planteamiento de investigación de este tipo es conveniente tener presente que la utilización de muestras no supone un procedimiento menos adecuado que si se encuestara a la población total. Existen varias razones que hacen preferible un estudio sobre una muestra en lugar de tomar toda la población:

- a) Encuestar al total de la población puede resultar del todo inviable por razones de esfuerzo y de costes.
- b) El estudio de una muestra es más rápido que si se intentara realizar para toda la población.
- c) El estudio exhaustivo de una población puede acumular muchos más errores que si se toma sólo una muestra.

A modo de resumen, se puede decir que la utilización de una muestra frente a la población total será adecuada en relación a la razón coste / beneficio. Es evidente que si la población es muy pequeña, puede no merecer la pena extraer una muestra de ella. Otra razón que hace poco conveniente la utilización de muestra, es en el caso en que la población sea excesivamente heterogénea; de ser así, la elección de una muestra representativa será muy problemática, con lo cual los resultados obtenidos serán, cuanto menos, cuestionables.

La investigación mediante encuestas se plantea para objetivos de investigación diversos, entre los que se encuentran estudios sobre grupos y organizaciones, cultura y socialización, estructura social, población y familia, medio ambiente, economía y trabajo, política,

problemas y servicios sociales, y un largo etcétera. Es sin lugar a dudas el método de investigación más difundido, lo que justifica que los métodos y técnicas de encuestación sean un contenido frecuente en multitud de planes de estudio universitarios y no universitarios.

## **2.- Elementos de la Investigación por Encuestas**

En la investigación mediante encuestas habremos de tener en cuenta algunos aspectos básicos, para que los objetivos planteados se alcancen de una forma eficiente y sistemática. Algunos de éstos son:

- Definición del objetivo primario, esto es, precisar claramente cuál es el objetivo de la investigación, y determinar si éste es alcanzable. Si lo es, debemos tenerlo presente para no dispersar esfuerzos de tiempo y coste en otros objetivos menos importantes.
- Delimitación de la población y selección de unidades, es decir, qué elementos formarán parte de nuestro estudio, como componentes de una muestra que es representativa de la población. Más concretamente:
  - El procedimiento seguido para la elección de encuestados determinará la validez de la investigación. Los métodos de muestreo probabilístico nos proporcionan un procedimiento científico para obtener estas muestras (aunque hay que decir que ciertos métodos de muestreo no probabilístico proporcionan también muestras representativas, cuyo déficit fundamental estriba en que no es posible obtener con ellas, estimadores de los parámetros poblacionales de los que se conozca ciertas propiedades que garanticen su eficiencia).
  - Al ser el encuestado la base de los datos, es preciso asegurar que sus respuestas carezcan de sesgos. De su implicación personal en el tema depende la sinceridad de las respuestas y, por tanto, la fiabilidad de los datos obtenidos. Este último aspecto ha de tenerse en cuenta a la hora de elaborar el cuestionario.
- Diseño del cuestionario: es el instrumento con el que se obtienen datos sobre la característica en estudio. Por medio del cuestionario homogeneizamos las posibles respuestas, permitiendo que éstas sean comparables, y se facilita el tratamiento informático de los datos.
- Medios y coste: con qué elementos técnicos y personales contamos para que el objetivo de la investigación sea realizable en tiempo y forma. Este factor es crítico en cualquier investigación, pero lo es más en el caso de una investigación por encuestas, ya que, por ejemplo, el no disponer de una Red de Campo que cubra toda la población, imposibilita per sé que ésta se pueda llevar a cabo, sin estar libre de sesgos en las estimaciones.

- Modo de administración: la forma en la que se obtienen los datos (p.e. mediante una entrevista personal) tendrá incidencia tanto en la calidad de los datos obtenidos, como en el coste y tiempo necesarios.
- Información relevante: es conveniente, en la mayor parte de las investigaciones por encuestas, conocer qué estudios previos hay sobre la cuestión investigada, qué dificultades se encuentran en estudios similares, qué metodología es más idónea para emplearse en la cuestión, y qué fuentes de información alternativas podemos usar para mejorar los datos obtenidos. Todo ello redundará en la eficiencia del estudio planteado.

### 3.- Fases de una Encuesta por Muestreo

Hemos comentado que la investigación por encuestas proporciona una forma sistemática de obtener datos sobre cierta característica de interés. Pues bien, a continuación expondremos los pasos que secuencialmente habremos de seguir desde la *concepción* hasta la *publicación de los resultados* de una encuesta. Éstos son:

- Diseño teórico, en el que se tendrá en cuenta:
  - El análisis del problema o definición de los objetivos de la investigación.
  - Ámbito de la investigación: población estudiada, periodo de tiempo del estudio y zona geográfica en que se desenvuelve, entre otros aspectos.
  - Definición de las unidades básicas de las que se obtiene la información: hogares, empresas, personas, etc.
  - Delimitación de las características objeto de estudio: p.e. ingresos, opiniones políticas, preferencias, dolencias físicas, etc.
  - Características generales del cuestionario: extensión, necesidad de pretest.
  - Procedimiento de muestreo: tipo de éste y tamaño de la muestra a considerar.

Por último, también a nivel teórico es conveniente fijar la metodología de recogida, depuración y tratamiento estadístico de la información. También es importante saber si hay encuestas previas igual a la que vamos a realizar.

- Diseño del cuestionario: en el que se tratarán de implementar las líneas básicas de las que se habla en la fase anterior, con especial atención al tipo y número de preguntas que contendrá, así como a la necesidad de un pretest que contraste la adecuación del mismo a la situación en que se lleva a cabo.
- Diseño de la muestra y puesta en práctica del tipo de muestreo ya decidido. Para ello se formará el marco en que se realizará el muestreo o actualizarlo si se dispone de uno anterior. Se establecerá la jerarquía de las unidades de muestreo, si hay que estratificar la población, así como determinar el tamaño de la muestra y el procedimiento de selección. En esta fase también se evaluará la conveniencia de efectuar diseños más complejos que los previamente establecidos, o por el contrario,

en el menor número de casos, simplificar éste, debido entre otros motivos a la homogeneidad de las unidades en estudio.

- Entrenamiento de los encuestadores, asegurando de esta forma que los datos recabados por ellos sean fiables. Este entrenamiento supone, por una parte, la familiarización con el cuestionario que se va a utilizar, el establecimiento de una forma de entrevistar homogénea, que garantice la igualdad de las respuestas, así como el aprender a manejar situaciones más o menos difíciles.
- Recogida de la información: Ésta se podrá efectuar mediante diversos métodos, entre los que encontramos:
  - La entrevista clásica, o sobre papel (PAPI: acrónimo de Paper Assisted Personal Interviewing).
  - Las encuestas por correo.
  - Las entrevistas personales asistidas por ordenador (CAPI: acrónimo de Computer Assisted Personal Interviewing).
  - Las entrevistas telefónicas asistidas por ordenador (CATI: acrónimo de Computer Assisted Telephone Interviewing).
  - Las auto-entrevistas asistidas por ordenador.
- Depuración de datos: en la que se pretende por una parte, mejorar la calidad de los datos (corrigiendo errores, detectando valores anómalos, etc) y por otra, evaluar la calidad de los datos, con el objeto de asegurar que la toma de éstos o el procedimiento de muestreo ha sido adecuado.
- Chequeo y validación de la muestra: en el que se estudiará el cumplimiento de la encuesta en cuanto a cobertura, calidad, número de entrevistas obtenidas, etc., así como a implementar acciones que mejoren ésta como la post-estratificación.
- Análisis estadístico primario. Por éste entendemos la obtención de tablas del plan de explotación, estadísticas descriptivas de variables simples, regresiones simples, un análisis exploratorio de datos que muestre la necesidad de transformación de variables, etc.
- Evaluación de errores: ya que antes de obtener conclusiones definitivas del análisis primario debemos estudiar la fiabilidad de las estimaciones obtenidas y que factores no determinados han podido afectar a tales estimaciones. Para ello se hará un estudio de los errores muestrales, el efecto del diseño de muestreo, y si es necesario utilizar métodos de remuestreo (Bootstrap y Jackknife).
- Análisis estadístico secundario: en el que tendrán su lugar las técnicas habituales del Análisis Multivariante (bien clásico, bien de datos), así como análisis más particulares como modelos log-lineales, métodos de reducción de la dimensión, modelos dinámicos como series temporales, etc.

- Interpretación, presentación y publicación de los resultados: cuyo aspecto más importante es la presentación del informe de resultados junto con las tablas y gráficos obtenidos, si bien éstas últimas, si se desea, se pueden presentar en un apéndice propio.

Una vez realizadas estas fases, la información obtenida en su conjunto debe ser almacenada en bases de datos para facilitar su posterior reexplotación y/o tratamiento. Esta acción, que se puede obviar en encuestas de propósito reducido, es ineludible en otro tipo de encuestas específicas como son los paneles (de la investigación sociológica y comercial), y en otras que en su metodología tengan implícitos análisis estadísticos dinámicos.

## 4.- Muestreo

En la investigación mediante encuestas es preciso tomar decisiones en diversas etapas. Henry (1990) habla de tres tipos de decisiones: previas al muestreo, decisiones de muestreo y decisiones posteriores.

- Las decisiones pre-muestreo se encuadran en algunas de las fases descritas anteriormente, sobre todo en la fase de diseño teórico.
- Las decisiones de muestreo se refieren a:
  - El listado de la población a utilizar.
  - El error tolerable o tamaño del efecto esperado.
  - El tipo de técnica de muestreo a utilizar.
  - Determinar si la probabilidad de selección de sujetos es igual o no.
  - Evaluar el tamaño de la muestra.
- En cuanto a las decisiones post-muestreo, tendremos que observar:
  - Cómo evaluar la ausencia de respuesta.
  - Necesidad de reponderación.
  - Estimación de errores típicos e intervalos de confianza.

En resumen, podríamos señalar tres objetivos básicos que se deben cubrir en un buen diseño cuando se lleva a cabo una investigación mediante encuestas: elegir adecuadamente a los sujetos a encuestar, seleccionar las preguntas para la elaboración de un cuestionario acorde con los fines para abordar el problema planteado en la investigación y, finalmente, organizar las preguntas para su análisis.

Aunque en la actualidad algunos métodos de muestreo no aleatorio como el muestreo por cuotas se están aplicando para la toma de datos en encuestas preferentemente comerciales, en las encuestas por muestreo la toma de datos se realiza a través de algunos de los métodos del muestreo probabilísticos. Si bien no es objeto de este curso desarrollar los diversos aspectos de la *Teoría del Muestreo en Poblaciones Finitas*, se recordarán algunas definiciones, se establecerán comparaciones entre éstos, y se citarán algunos de los métodos habituales de muestreo en la práctica de las encuestas. Los libros de *Cochran (1971)*, *Sánchez-Crespo (1984)*, *Azorín y Sánchez-Crespo (1986)*, *Kish (1967)*, constituyen la bibliografía básica sobre esta materia.

## 4.1.- Terminología Básica de Muestreo

Como hemos visto en el apartado anterior, en la investigación por encuestas es primordial entender los conceptos básicos del muestreo, que definimos a continuación.

- **Población:** conjunto de unidades del que se desea obtener cierta información. Las unidades pueden ser personas, viviendas, escuelas, etc. y la información deseada, el consumo medio por familia, número de personas en paro, número medio de escolares por aula, etc.

Kish (1987) y Henry (1990) distinguen entre diversos tipos de población, de acuerdo con su papel en la investigación:

- Población objetivo: aquel conjunto de individuos al que queremos generalizar los hallazgos obtenidos, excluyendo los elementos a los que sea imposible o costoso acceder.
  - Población de estudio o población marco: la población que realmente estudiamos, con la que se puede operar, de la que hay algún listado o alguna forma de acceder a los sujetos que la conforman.
  - Población inferencial: población a la que el investigador desea hacer inferencias.
  - Población de la encuesta: conjunto de individuos que contestó a la encuesta.
- **Muestreo:** es la disciplina que trata con el conjunto de técnicas para tomar u obtener una muestra.
  - **Muestra:** conjunto de elementos de la población que permiten obtener datos más precisos de las características de dicha población.
  - **Muestra representativa:** subconjunto de una población que tiene las mismas características generales que la población.
  - **Marco:** conjunto muestral a partir del cual se selecciona la muestra. Para determinar el marco muestral es necesario contar con Bases de datos ó Mapas.
  - **Elemento:** unidad básica de la que se busca información y que proporciona los datos para el análisis.
  - **Unidades de muestreo:** Colecciones no solapadas de elementos de la población que cubren la población marco completamente.
  - **Subclase:** Subdivisiones o particiones de la muestra.
  - **Dominio:** Subdivisión de la población para la que se han planificado, diseñado y seleccionado muestras separadas.

## 4.2.- Tipos de Muestreo

Entendemos por muestreo al procedimiento mediante el cual se extrae una muestra de una población dada. Genéricamente, existen dos tipos de muestreo: el muestreo probabilístico y muestreo no probabilístico.

En el primero es posible calcular de antemano la probabilidad de cada una de las muestras que pueden extraerse de la población, para lo cual el procedimiento de elección de la muestra ha de ser aleatorio, y debe asegurar que la selección de cada unidad en la muestra es independiente de la selección de otra unidad; dicho con otras palabras, la selección de cualquier miembro de la población no afecta a la probabilidad de ningún otro miembro para ser seleccionado. Este tipo de muestreo, con la base en la Estadística Matemática, permite evaluar y controlar tanto la precisión como los errores cometidos.

El segundo tipo se puede definir como aquel en el que no hay forma de estimar la probabilidad de que cada elemento sea incluido en la muestra, ya que no se garantiza que cualquier individuo de la muestra tenga alguna probabilidad conocida de ser incluido en la misma. Hay muchos procedimientos de selección no aleatoria de unidades muestrales, que difieren en cuanto a la precisión de los datos que aportan. No obstante, como característica general en todos los casos, al no ser muestras representativas, no se pueden generalizar los resultados a la población. La selección de los sujetos se puede realizar sin seguir ninguna norma, como en el caso de las muestras de conveniencia, o bien es el investigador quien determina la forma de elección de los sujetos y la composición muestral, en las denominadas muestras a propósito.

En el caso probabilístico, a la selección aleatoria de los sujetos que se van a incluir en la muestra presenta algunas ventajas importantes:

1. Elimina el juicio subjetivo en la selección de la muestra y aumenta la credibilidad de los resultados, permitiendo además su generalización.
2. Permite calcular la variabilidad de la muestra, apresando así los posibles errores debidos a tal variabilidad. Puesto que la precisión disminuye conforme aumenta la variabilidad, la teoría muestral facilita el conocimiento de los factores que contribuyen a dicha variabilidad, pudiendo, de esta forma, reducirla y aumentar la precisión.
3. Se trabaja con un margen de error conocido y establecido por el investigador, reduciendo de esta forma la incertidumbre.

En la práctica de las encuestas, los modelos de muestreo empleados son bastante complicados ya que suelen consistir en la aplicación simultánea de varios de los tipos de muestreo básicos que describiremos. Además se establecerá lo que se denomina como *esquema de muestreo* en el que se describe el proceso multietápico del muestreo. En cualquier caso, como es natural la base la constituirán los diversos tipos de muestreo básicos sobre los que también pasamos a hacer un breve recordatorio:

- **Muestreo aleatorio con reemplazamiento:** todas las posibles muestras tienen la misma probabilidad y, por tanto, también todas las unidades de la población tienen la misma probabilidad de ser escogidas.
- **Muestreo aleatorio simple (sin reemplazamiento):** en éste todas las unidades tienen igual probabilidad de ser extraídas, aunque al no reintegrar el elemento a la

población, la probabilidad de obtener un elemento determinado en sucesivas extracciones, depende de las extracciones anteriores en el caso de que la población sea finita.

- **Muestreo sistemático:** consiste en ir eligiendo las unidades muestrales de  $K$  en  $K$  unidades, tomando como origen una de ellas elegida a su vez de forma aleatoria de entre el primer elemento y el  $K$ -ésimo.
- **Muestreo estratificado aleatorio:** los elementos se clasifican en grupos de unidades con características homogéneas que se denominan estratos. Sobre cada estrato se seleccionarán muestras aleatorias con un tamaño dado según cierto criterio denominado *afijación*, entre los que están:
  - Igual tamaño en cada estrato (*afijación uniforme*).
  - Tamaño proporcional al tamaño del estrato en la población que se muestrea (*afijación proporcional*).
  - Afijación proporcional corregida por la varianza del estrato, esto es, seleccionando más muestras de los estratos con más variabilidad (*afijación óptima*).
- **Muestreo por conglomerados:** se denomina así debido a que las unidades objeto de estudio se agrupan en conglomerados o áreas ocupadas por los elementos de la población. Se trata por tanto de seleccionar aquellos conglomerados más representativos de la población, de forma que estén compuestos de unidades lo más heterogéneas posibles.
- **Muestreo bietápico o polietápico:** Cuando se realiza un muestreo por conglomerados pero no se recogen observaciones de todos los elementos, sino que se realiza un muestreo de éstos.
- **Muestreo doble o bifásico:** Se extrae una muestra grande que sirve de base para otra más pequeña. Se utiliza generalmente para muestrear poblaciones raras: en la primera fase se identifican los elementos que pertenecen a la población objetivo y en la segunda se recoge información de la encuesta.
- **Muestras interpenetrantes:** División aleatoria de una muestra aleatoria en un número de subgrupos con el mismo número de elementos. Se usa para determinar sesgos debidos al entrevistador.
- **Otros tipos de muestreo probabilístico:**
  - **Muestreo de recaptura de elementos marcados:** utilizado en la estimación del número de elementos de una población animal móvil.
  - **Muestreo de eliminación,** utilizado para estimar el tamaño de una población cerrada.
  - **Muestreo de posiciones de elementos en el espacio:** conocer si los elementos aparecen en posiciones aleatorias dentro del área de estudio.

Los tipos de muestreo probabilístico usados más frecuentemente son: muestreo aleatorio simple, muestreo estratificado y muestreo por conglomerados.

En cuanto a los tipos de muestreo no probabilístico, los más conocidos utilizados en la práctica de las encuestas son:



- **Muestras de conveniencia o accidentales:** Se forman con un grupo de individuos que se encuentran disponibles para participar en el estudio. Los grupos de alumnos que se presentan voluntarios para participar en un experimento constituyen un ejemplo de las muestras de conveniencia.
- **Muestreo de casos típicos:** Selección de casos que se conoce de antemano que son normales o usuales, eliminando aquellos sujetos que puedan ser considerados únicos o especiales. El método de selección de las unidades muestrales centra la atención en cada caso individual, con un fino escrutinio de la muestra, tratando de asegurarse que no se han incluido casos extremos. Este tipo de selección se utiliza cuando problemas de tiempo y recursos hacen inviables el muestreo probabilístico.
- **Muestreo de casos críticos:** Selección de casos que son esenciales para la aceptación general o para la evaluación. Esto es, la selección de casos se hace en función de que, por conocimiento previo o por lógica, permitan la generalización a la población. Es un procedimiento similar al anterior, con la salvedad de que se elige un grupo dado, que se considera similar, en pequeña escala, a la población con la que se desea comparar.
- **Muestreo de “bola de nieve”:** Se basa en grupos previamente identificados para extraer otros miembros de la población. Se suele utilizar cuando no hay listado de población disponible. Se utiliza en estudios sociológicos, para la determinación de grupos de los que no se tiene censo ni posibilidades de tener un listado, como puede ser el caso, por ejemplo, de consumidores de drogas no permitidas. El procedimiento supone que los mismos miembros que conforman el grupo de interés identifican a otros individuos pertenecientes a la población.
- **Muestreo por cuotas:** También denominado *método empírico* (Martínez Arias, 1995). Los entrevistadores seleccionan la muestra en la misma proporción que está en la población en variables fácilmente identificables. Este tipo de selección de la muestra es en alguna medida una especie de muestra estratificada, y suelen ser muy utilizadas por algunas agencias de investigación de mercado. A los encuestadores se les fija una cuota de individuos que deben entrevistar, especificándoles sus características, como por ejemplo “varones, casados, propietarios de automóviles, con edades entre 25 y 50 años”. Cada entrevistador selecciona por su cuenta a los sujetos basándose en estos criterios, hasta completar la cuota. Este procedimiento es muy popular, y presenta como ventaja principal que se puede obtener la muestra de forma rápida y barata. No es preciso contar con listas de la población, ya que sólo basta con tener información sobre las características de la población con las que definir las cuotas. Sin embargo, se pueden cometer sesgos de muy diverso tipo, como puede ser encuestar sólo a aquellas personas que estén disponibles; por ejemplo, si se realiza una encuesta en horas laborales en los hogares, será mucho más probable encontrar varones en paro que en situación laboral activa.

Las muestras no aleatorias presentan diversos problemas, de difícil solución. El problema principal que cabe destacar es la subjetividad, lo que impide el desarrollo de un marco referencial teórico para ellos. La subjetividad que caracteriza el proceso de selección no ofrece ninguna garantía de representatividad de la muestra, con lo que los sesgos potenciales son muchísimos, siendo muy factible la existencia de variables encubiertas que

expliquen alternativamente los resultados obtenidos, puesto que si las características de la muestra utilizada difieren de las de la población de referencia, esas variables distintas pueden producir efectos sobre las variables de estudio inadvertidamente, amenazando con ello la validez interna de la investigación. Por otra parte, en muestras no aleatorias no se puede determinar de ninguna manera el margen de error con el que se está trabajando (margen de error que si es posible determinar matemáticamente cuando la muestra se ha seleccionado aleatoriamente) por lo que el grado de incertidumbre de los resultados es total.

El siguiente problema que se deriva de la falta de representatividad de la muestra es que los resultados así obtenidos son difícilmente extrapolables, impidiendo la generalización de los hallazgos a la población, por lo que la investigación carece también de validez externa. En presencia de muestras no aleatorias sólo es posible generalizar a una población hipotética, nunca a una población real, acotada y definida.

En suma, pues, los procedimientos de muestreo no aleatorios amenazan la validez interna y externa de la investigación, lo que limita de forma considerable la credibilidad de los resultados obtenidos. No obstante, se pueden producir situaciones en las que las muestras no aleatorias resulten adecuadas en la investigación. Así, en ocasiones, es el único método disponible con el que cuenta el investigador para acceder a la información que le interesa.

Finalmente, este tipo de muestras resultan útiles en investigaciones exploratorias, donde el interés reside en determinar si existe o no un problema concreto, delimitarlo y estudiar las variables relevantes, e incluso las relaciones que pueden haber entre ellas. Toda vez que se ha hecho esta investigación previa y bajo el supuesto que el tema en estudio es importante, será preciso que se lleve a cabo un estudio en profundidad, donde la elección de una muestra representativa (obviamente, enmarcado en un diseño de investigación cuidadoso) es el principal garante de los resultados obtenidos.

### **4.3.- Marcos Muestrales**

Para la selección de la muestra es preciso contar con un marco, a partir del cual se puedan extraer las unidades muestrales. Cuando hablamos de marco estamos haciendo referencia a listas, censos, mapas de ciudad, etc. Dado que el Marco muestral es el conjunto a partir del que se selecciona la muestra:

- Debe proporcionar el medio para identificar y localizar las unidades de muestreo.
- Debe proporcionar la información para permitir estratificación y otros agrupamientos.
- Representa el primer paso para determinar las probabilidades de inclusión en la muestra de las diferentes unidades.

Podemos considerar varios tipos de marcos en términos de su contenido:

- Listas de sujetos: censos, listas electorales.
- Listas de viviendas: realizadas a partir de los censos.
- Listas administrativas: listas de colegios, hospitales, etc.
- Mapas o planos: se subdividen en áreas que serán las unidades de muestreo.

- Guías telefónicas
- Random Digit Dialing (RDD): generación aleatoria de alguno de los dígitos del teléfono.

En cuanto a la calidad del marco, podemos distinguir tres tipos. El marco ideal es el marco perfecto, esto es, aquel donde cada elemento ocurre una sola vez, no aparecen unidades muestrales que no pertenezcan a la población que se desee estudiar y la información suministrada es correcta. Sin embargo, tal tipo de marcos no son demasiados comunes. Una segunda categoría corresponde a los marcos útiles, en los cuales se debe poder calcular la probabilidad de seleccionar las unidades muestrales. Tal probabilidad debe ser distinta de cero. Por último, estarían los marcos imperfectos, en los que aparecen cuestiones tales como unidades extrañas, incorrecciones, duplicaciones, etc.

Como hemos dicho, los marcos para seleccionar a los sujetos pueden ser listas o censos de los mismos. La selección basada en listas presentan la ventaja de que permiten un gran margen de selección. Además, la existencia de un listado posibilita la realización de encuestas por correo. No obstante, como mayor desventaja cabe destacar que suelen contener grandes errores.

Por su parte, la calidad de un marco viene determinada por:

- Su exhaustividad.
- Que se identifique de manera no ambigua a los elementos incluidos en las unidades de muestreo.
- Que exista una regla de asociación determinable entre las unidades del marco muestral y los elementos de la población.
- Que cada unidad tenga la misma o una probabilidad conocida de ser seleccionada.
- No incluya personas que no pertenezcan a la población objetivo.
- No contenga repeticiones.
- Ser fácil de usar.
- Que esté organizado de forma sistemática.
- Contener información auxiliar que pueda usarse en el diseño.

Respecto a las limitaciones que podemos encontrar, citaremos:

- Faltas de cobertura:
  - o Inadecuado: no intenta incluir a toda la población.
  - o Incompleto: no incluye algunos elementos que se supone debería incluir.
 Como solución podemos considerar el uso de *marcos complementarios*, eliminando duplicaciones o procedimientos *enlazados*, ligando los elementos que faltan a listados especificados de forma claramente definida.
- Vinculación a conglomerados: cuando los listados se refieren a grupos de elementos y no a elementos individuales.
- Blancos o elementos extraños: cuando alguno de los elementos no son de la población actual (fallecidos, emigrantes, etc.) o hay elementos que están fuera del objetivo del estudio. Este problema se puede paliar ignorándolos cuando son

extraídos, aunque si son muy numerosos conviene utilizar un diseño muestral en dos fases, para identificar los elementos de la población objetivo.

- Duplicaciones o repeticiones, cuando los elementos son listados más de una vez. Puesto que los elementos duplicados tienen mayor probabilidad de ser seleccionados, habremos de hacer un esfuerzo previo para eliminarlos del marco.

#### **4.4.- Apéndice: tamaño de la muestra en el m.a.s.**

El cálculo del tamaño de la muestra es uno de los aspectos más delicados en el diseño de muestreo. Obviamente, el diseño de muestreo condiciona el tamaño muestral, pero también el coste de la investigación, y el error máximo admisible, serán factores que determinen el tamaño final en la investigación. A continuación proporcionamos las fórmulas necesarias para obtener el número mínimo de sujetos que debe formar parte de una muestra aleatoria simple para estimar los parámetros habituales: proporción, media y total de una población.

Para diseños más complejos, se recomienda la lectura del citado libro de Sánchez-Crespo o la monografía de Pascual et al. "Tamaño de Muestra y Precisión Estadística" del Servicio de Publicaciones de la Universidad de Almería.

Denotando por:

- $n$  = tamaño de la muestra.
- $N$  = tamaño de la población.
- $1-\alpha$  = el nivel de confianza elegido.
- $z_{1-\alpha}$  = el valor de  $z$  (con  $z$  una variable  $N(0,1)$ ), que deja fuera del intervalo  $\pm z_{1-\alpha}$  una proporción  $\alpha$  de los individuos.
- $p$  = proporción en que la variable estudiada se da en la población.
- $q = 1 - p$ .
- $e$  = error de la estimación.

El tamaño de muestra para una población infinita:

$$n = z_{1-\alpha}^2 \frac{pq}{e^2}, \text{ siempre que: } np \geq 5 \text{ y } nq \geq 5$$

En el caso de poblaciones muy grandes (superiores a 10.000) y con las condiciones establecidas (nivel de confianza = 0,95;  $z_{1-\alpha} = 1,96 \approx 2$ ;  $p = q = 0,5$ ), aplicando la fórmula anterior, tenemos que, para un margen de error  $e$ , el tamaño de la muestra es el que se especifica a continuación:

Error (e) en %	Tamaño de la muestra
0,1	1.000.000
0,5	40.000
1,0	10.000
2,5	1.667
5,0	400

Para una población finita:

$$n = \frac{z_{1-\alpha}^2 pqN}{e^2(N-1) + z_{1-\alpha}^2 pq}, \text{ siempre que: } np \geq 5 \text{ y } nq \geq 5$$

En este caso de poblaciones finitas, con mismas condiciones establecidas para el caso anterior, los tamaños muestrales son:

POBLACION	MARGENES DE ERROR					
	1 %	2 %	3 %	4 %	5 %	10 %
500					222	83
1.000				385	286	91
5.000		1.667	909	556	370	98
10.000	5.000	2.000	1.000	588	385	99
50.000	8.333	2.381	1.087	617	397	100
100.000	9.091	2.439	1.099	621	398	100

NOTA: Los espacios que quedan en blanco del cuadro anterior, proporcionan valores superiores a la mitad de la población. Cuando esto ocurre lo más aconsejable es tomar toda la población directamente, a no ser que existan razones de otra índole que aconsejen no trabajar con su totalidad.

En la aplicación de las fórmulas para el cálculo del tamaño de la muestra suele presentarse el problema de determinar el valor de p. Esta dificultad se puede solventar de una de las siguientes formas:

- Si existen estudios anteriores sobre la misma temática, podemos obtener, a través de ellos, una idea aproximada del valor de p.
- Se puede realizar un sondeo previo, con una muestra pequeña, para estimar el valor de p.