

## Capítulo 7

**EL MUESTREO**

## 7.1. La investigación en concreto

La definición y especificación del diseño a desarrollar en una investigación constituye una de las actividades principales de lo que hemos denominado el *momento metodológico* del proceso (v. *supra*, 3.1). Pero ella no agota esta importante etapa, que debe ser completada con una tarea de singular importancia: convertir en operativos, es decir, manejables, a los diversos elementos que intervienen en el problema a investigar. Esta tarea, denominada a veces con el neologismo *operacionalización*, constituye el eslabón indispensable que hace factible la verificación en concreto de lo que se ha postulado en el marco teórico, pues posibilita acercarse a la realidad empírica para adquirir la información de la que depende dicha verificación.

Dos tipos muy diferentes de elementos requieren de lo que llamamos por un lado el universo, en tanto conjunto de unidades o fuentes de datos que es preciso reducir a proporciones manejables para poderlo explorar; por otra parte las variables, en tanto conceptos abstractos a los que es preciso dar contenido concreto para poder efectuar sobre ellos las mediciones correspondientes.

La operacionalización del universo consiste en reducir a proporciones factibles de investigar al conjunto de las unidades que nos interesan, en otras palabras, en la tarea de encontrar una forma de obtener información relevante sin necesidad de acudir a la medición de todo el universo posible de datos. Es el aspecto cuantitativo de la operacionalización y, para resolverlo, habrá de apelarse entre otras disciplinas a la estadística, mediante las técnicas de muestreo. La operacionalización de las variables es, por lo contrario, de naturaleza básicamente cualitativa, y tiene por

objeto encontrar los indicadores a través de los cuales se expresa concretamente el comportamiento de las mismas.

La importancia de una correcta operacionalización se expresa por sí misma: si nuestras variables no pueden ser medidas y evaluadas en la realidad, o si lo que medimos y evaluamos no se corresponde con nuestras formulaciones iniciales, todas nuestras conclusiones estarán vacías de contenido empírico o, a lo sumo, expresarán una realidad divergente de la que nos interesa conocer. En otras palabras, la investigación perderá todo sentido y derivará en un mero ejercicio especulativo o en una serie de confusiones, imprecisiones y errores que le restarán carácter científico. De aquí que el investigador deba atender cuidadosamente a la correcta realización de esta etapa, para evitar desperdiciar todo el esfuerzo hecho hasta ese momento.

La estrecha relación que existe entre operacionalización y tipo de diseño se origina en que, de acuerdo al tipo de método general a utilizar. O sea, de diseño B se requerirá con mayor preponderancia uno u otro tipo de datos, operacionalizar de una cierta manera o de otra. Inversamente, de acuerdo a las posibilidades concretas de operacionalizar en uno u otro sentido las variables y el universo considerado, es que será más o menos adecuado un cierto tipo de diseño. Por esta razón ambas tareas deben encararse casi simultáneamente, atendiendo a las repercusiones que las decisiones en un aspecto puedan acarrear en el otro (v. *supra*, 3.2). La separación que hacemos en este texto es, por lo tanto, algo artificial, y responde más a propósitos pedagógicos de ordenamiento del material que a la dinámica real del proceso de investigación.

En este capítulo nos ocuparemos del problema de obtención de muestras, dejando para el subsiguiente lo relativo a la búsqueda de indicadores para las variables. Antes de exponer los elementos esenciales del muestreo será necesario, sin embargo, detenernos por unos momentos en los conceptos básicos que se necesitan para comprender lo que esta operación significa.

## 7.2. Datos y unidades

Por **dato** entendemos cada uno de los elementos de información que se recoge durante el desarrollo de una investigación y sobre la base de los cuales, convenientemente sintetizados, podrán extraerse conclusiones de relevancia en relación al problema inicial planteado.

Cualquier información, por más pequeña y fragmentaria que sea, puede considerarse como un dato siempre y cuando pueda colaborar, de algún modo, a esclarecer los problemas que nos planteamos en un estudio. Saber, por ejemplo, que la persona NN está a favor del candidato X, es un dato. Esa información, por sí sola, carece prácticamente de valor, pues poco nos dice respecto de las intenciones generales de voto que existen entre los electores. Pero el valor del dato reside no en su alcance individual, en lo que nos expresa por sí mismo, sino en su posibilidad de ser integrado en un conjunto mayor. Es sólo ahí, cuando agrupamos muchas informaciones de carácter similar, cuando en nuestro ejemplo B consultamos a muchas personas respecto a la opinión que tienen sobre el tema, que cada dato individual se hace valioso dentro de una perspectiva más amplia. Así podemos llegar a enunciar que un determinado tanto por ciento de las personas están a favor del candidato X, e integrar esta información, a su vez, en un estudio de opiniones de tipo preelectoral.

En el ejemplo anterior cada dato provenía directamente de alguna pregunta que se formulaba a personas determinadas. En ese sentido puede decirse que cada una de esas personas se constituye en el origen, en la fuente de los datos obtenidos. Estas fuentes, que pueden ser personas, situaciones o hechos que se observan directamente, o materiales bibliográficos de diversa naturaleza, son las que llamamos **unidades de datos** y, a su conjunto, a la suma de todas las unidades, se le da el nombre de **universo**. En general toda investigación puede considerarse como una búsqueda de los datos apropiados que permitan resolver ciertos problemas de conocimiento. Dichos datos son obtenidos a través de un conjunto de unidades que constituyen el universo relevante para la investigación.

Suele ocurrir que una enorme cantidad de unidades resultan relevantes para una investigación: existen universos tan amplios como el conjunto de los ciudadanos de un país, la flora de una región o las innumerables galaxias. Estos universos resultan por lo general demasiado extensos para el investigador, pues éste no tiene ni el tiempo ni los recursos para abordar sistemáticamente el estudio de cada una de las unidades que lo componen. Para resolver este inconveniente se acude a la operacionalización del universo mediante la extracción de muestras, tal como veremos a continuación, a la selección de casos (como vimos en 6.4.6) y a otras técnicas de selección que, por razones de espacio, no será posible tratar.

### 7.3. Universo y muestra

En el caso de que nuestro universo esté compuesto por un número relativamente alto de unidades será prácticamente imposible, por razones de tiempo y de costos, y porque no es en realidad imprescindible, examinar cada una de las unidades que lo componen. En vez de realizar esa fatigosa tarea procederemos a extraer una muestra de ese universo, o sea un conjunto de unidades, una porción del total, que nos represente la conducta del universo en su conjunto. Una muestra, en un sentido amplio, no es más que eso, *una parte del todo que llamamos universo y que sirve para representarlo*.

Sin embargo, no todas las muestras resultan útiles para llevar a cabo un trabajo de investigación. Lo que se busca al emplear una muestra es que, observando una porción relativamente reducida de unidades, se obtengan conclusiones semejantes a las que lograríamos si estudiáramos el universo total. Cuando una muestra cumple con esta condición, es decir, cuando nos refleja en sus unidades lo que ocurre en el universo, la llamamos **muestra representativa**. Sus conclusiones son susceptibles de ser generalizadas al conjunto del universo, aunque para ello debemos añadir un cierto margen de error en nuestras proyecciones. Casi todas las muestras que se utilizan en las investigaciones son muestras representativas no obstante que, en algunos casos, se

empleen muestras no-representativas. Ello ocurre cuando no se pretende hacer un trabajo muy preciso, cuando se buscan conocer apenas algunos indicios generales de un problema o cuando el tiempo impide otra forma de trabajo más rigurosa. De ninguna manera es lícito, sin embargo, proyectar estos conocimientos de tipo sumario hacia el universo, siempre que se tomen como aproximaciones previas para estudios más profundos.

No puede saberse nunca *a priori* si la muestra obtenida es o no representativa, pues para saberlo con absoluta certeza sería necesario investigar todo el universo y luego comparar ambos resultados. Como esto resulta tan absurdo como encender todos los fósforos de una caja para saber si son buenos, lo que se hace es acudir a procedimientos matemáticos que son capaces de decirnos con qué nivel de confianza trabajamos al escoger una muestra determinada. El estudio de estos procedimientos corresponde al campo de la estadística. No es el objeto de nuestro trabajo desarrollar temas que, por su complejidad, han sido tratados extensamente en trabajos especializados. Por ello simplemente nos remitiremos a dar una idea general de los principales tipos de muestras que se utilizan, así como de las técnicas que permiten obtenerlas. [Para todo este capítulo es sumamente importante consultar a Johan Galtung, *Teoría y métodos de la Investigación Social*, Ed. Eudeba, Buenos Aires, 1971, pp. 49 a 75.]

Una primera división que suele hacerse entre las muestras consiste en separarlas en muestras *probabilísticas* y *no probabilísticas*. En el primer tipo la característica fundamental es que todo elemento del universo tiene una determinada probabilidad de integrar la muestra, y que esa probabilidad puede ser calculada matemáticamente con precisión. En las muestras no probabilísticas ocurre lo contrario y el investigador procede en cierta forma a ciegas, pues no tiene una idea del error que puede estar introduciendo en sus apreciaciones. Las muestras no probabilísticas más utilizadas son las llamadas *accidentales*, *por cuotas* o *intencionales*.

Una muestra **accidental** es aquella que se obtiene sin ningún plan preconcebido, resultando las unidades escogidas producto de circunstancias fortuitas. Si, por ejemplo, entrevistamos a los primeros cincuenta transeúntes que pasan por una cierta calle o medimos la profundidad del mar a lo largo de un trayecto entre dos puntos cualesquiera, estaremos en presencia de una muestra accidental. Los datos obtenidos podrán o no representar al universo en estudio, porque corresponderán a circunstancias peculiares del lugar y momento en que se tomó la muestra. El investigador no puede saber, entonces, hasta qué punto sus resultados podrán proyectarse, con confiabilidad, hacia el conjunto más amplio que desea conocer.

Una muestra **por cuotas** consiste en predeterminar la cantidad de elementos de cada categoría que habrán de integrarla. Así podemos asignar una cuota de 50 hombres y 50 mujeres a una muestra de 100 individuos, asumiendo que esa es la distribución de la población total. Por más que esa presunción llegue a ser válida no deja de existir cierta arbitrariedad en este modo de proceder, por lo que la rigurosidad estadística de las muestras por cuotas se reduce considerablemente.

Una muestra **intencional** escoge sus unidades no en forma fortuita sino completamente arbitraria, designando a cada unidad según características que para el investigador resulten de relevancia. Estas muestras son muy útiles y se emplean frecuentemente en los estudios de caso, por más que la posibilidad de generalizar conclusiones, a partir de ellas, sea en rigor nula (Cf. 6.4.5).

#### 7.4. Muestras aleatorias

En las muestras aleatorias cada uno de los elementos del universo tiene una probabilidad determinada y conocida de ser seleccionado. Para que esto suceda así es necesario proceder a la extracción de la muestra mediante ciertas técnicas, capaces de garantizarnos que cada elemento de la misma posea una probabilidad conocida de aparecer en ella. Los procedimientos

más usuales para la obtención de muestras aleatorias son los siguientes:

#### 7.4.1. Al azar simple

Este procedimiento se inicia confeccionando una lista de todas las unidades que configuran el universo, numerando correlativamente cada una de ellas. Luego, mediante cualquier sistema (programas de computación, tabla de números al azar, etc.) se van sorteando al azar estos números hasta completar el total de unidades que deseamos que entren en la muestra. De este modo la probabilidad que cada elemento tiene de aparecer en la muestra es exactamente la misma.

Este método nos garantiza una selección completamente aleatoria pero resulta muy lento y costoso pues nos obliga a elaborar listas completas de todas las unidades de interés, lo cual a veces resulta sencillamente imposible. Además se requiere realizar una por una la selección de cada elemento, lo que es evidentemente muy fatigoso y caro cuando las dimensiones del universo son grandes. Por este motivo sólo se emplea el sistema del azar simple cuando los universos son relativamente pequeños, cuando B por diversas razones B es posible conocer taxativamente cada una de las unidades y cuando además la muestra no tiene dimensiones muy amplias.

Este método no será adecuado si, por ejemplo, queremos sacar una muestra de todas las personas analfabetas que existen en un país, pues por su número y por lo indeterminado de su situación nos será imposible confeccionar la lista de todas las que existen. En cambio, si nuestra intención es extraer una muestra del universo de todos los alumnos que ingresan a una universidad en un determinado año, resultará un método adecuado, puesto que se trata de una cantidad menor de unidades que además están registradas ya en listas pertinentes.

#### 7.4.2. Al azar sistemático

Este tipo de muestra parte de una idea básica muy similar a la de la técnica anterior requiriéndose también, en este caso, de un listado completo de las unidades que integran el universo en estudio. Luego, en vez de proceder a escoger una por una las unidades de acuerdo al método ya señalado, se efectúan las siguientes operaciones:

a) Se calcula la constante  $K$ , que resulta de dividir el número total de unidades que componen el universo por el número de unidades que habrán de integrar la muestra:

$$K = N/n$$

Donde:

$N$  = número total de unidades que componen el universo.

$n$  = número total de unidades que integrarán la muestra.

b) Una vez calculado el valor de  $K$  se efectúa un sorteo para elegir un número que sea inferior o igual a su valor. Como primera unidad a integrar la muestra se elige aquella que, en la lista general, posea idéntico número de orden al sorteado. Si designamos con  $A$  a este primer valor, la segunda unidad escogida será la que lleve el número  $A+K$ , la tercera corresponderá a  $A+2K$ , y así sucesivamente hasta llegar a  $A+(n-1)K$ .

Supongamos un universo constituido por 2.800 elementos del que deseamos obtener una muestra de 70 casos. Tenemos entonces:

$$N = 2.800$$

$$n = 70$$

$$K = 2.800/70 = 40$$

Ahora, mediante cualquier procedimiento, buscamos al azar un número entero cuyo valor figure entre los límites de 1 y 40. Sea el número escogido, en este caso, el 23. Entonces, las unidades que

pasarán a formar parte de la muestra serán las que lleven los siguientes números de orden:

1ª unidad: 23

2ª unidad:  $23 + 40 = 63$

3ª unidad:  $23 + 80 = 103$

.....

70ª unidad:  $23 + 2.760 = 2.783$

Las ventajas y desventajas de este procedimiento son casi idénticas a las de las muestras al azar simple, aunque estas últimas se prefieren ahora más que hace unos años, ya que los procedimientos computacionales hacen mucho más fácil efectuar el sorteo de las unidades y no existe el riesgo de que la muestra quede sesgada por algún tipo de regularidad que no conocemos y que esté presente en el universo.

#### 7.4.3. Muestras por conglomerados

Esta técnica tiene utilidad cuando el universo que se requiere estudiar admite su subdivisión en universos menores Ben partes del mismoB de características similares a las del universo total. Cuando es posible asumir tal cosa se procede a subdividir el universo en un número finito de conglomerados y, entre ellos, se pasa a escoger algunos que serán los únicos que se procederá a investigar. Esta elección puede realizarse ya sea por el método del azar simple o del azar sistemático. Una vez cumplida esta etapa puede efectuarse una segunda selección, dentro de cada uno de los conglomerados elegidos, para llegar a un número aún más reducido de unidades muestrales.

La ventaja de esta técnica es que obvia la tarea de confeccionar el listado con todas las unidades del universo, lo cual, como ya veíamos, resulta imposible de hacer en muchos casos. Su desventaja mayor radica en que, al efectuarse el muestreo en dos etapas, los errores muestrales de cada una se van acumulando,

dado un error algo mayor que para los métodos descritos anteriormente.

La técnica de los conglomerados suele utilizarse cuando queremos extraer muestras de los habitantes de un conjunto geográfico amplio, v.g., una gran ciudad o un conjunto de aldeas, por lo que se procede a tomar cada aldea o grupo de manzanas como un conglomerado independiente. Del mismo modo se la utiliza para investigar reservas forestales y marinas, estrellas, productos industriales y otros universos relativamente semejantes.

#### 7.4.4. Muestras estratificadas

A la inversa de lo que sucedía en el caso anterior, este método supone que el universo pueda desagregarse en sub-conjuntos menores, homogéneos internamente pero heterogéneos entre sí. Es como si fragmentáramos el universo en estratos o categorías de unidades, diferenciándolos de acuerdo a alguna variable que resulte de interés para la investigación. Cada uno de estos estratos se toma luego como un universo particular, ya de tamaño más reducido, y sobre él se seleccionan muestras según cualquiera de los procedimientos anteriores indicados. Este sistema resulta muy práctico y confiable, aunque también aquí el error total se incrementa con respecto a los métodos de azar simple o sistemático.

Por ejemplo, si quisiéramos estudiar las actitudes políticas de los estudiantes de una universidad, podríamos subdividirlos en estratos de acuerdo al tipo de estudios que cursen, asumiendo que estas actitudes van a ser diferentes entre quienes siguen Ingeniería, Letras, Medicina u otras carreras. Luego efectuaríamos un muestreo dentro de cada sub-universo así definido para, finalmente, realizar un análisis integrando los resultados de todas las sub-muestras obtenidas.

Existen otras posibilidades de extracción de muestras, además de las que surgen de la combinación de los tipos ya enunciados, así como diversos problemas técnicos que se plantean en su aplicación y desarrollo. Preferimos, sobre ese tema, remitir al

lector a los tratados especializados que versan sobre tales temas, por cuanto su explicación excedería los límites temáticos de este libro.

### 7.5. Tamaño de la muestra y error muestral

Cuando una muestra es aleatoria o probabilística es posible calcular sobre ella el **error muestral**. Este error indica el porcentaje de incertidumbre, es decir, el riesgo que se corre de que la muestra escogida no sea representativa. De tal modo que, si trabajamos con un error calculado en 5%, ello significa que existe un 95% de probabilidades de que el conjunto muestral represente adecuadamente al universo del cual ha sido extraído.

A medida que incrementamos el tamaño de la muestra el error tiende a reducirse, pues la muestra va acercándose más al tamaño del universo. Del mismo modo, para una muestra determinada, su error será menor cuanto más pequeño sea el universo a partir del cual se la ha seleccionado. Así, para un universo de 10.000 casos, una muestra de 200 unidades tendrá un error mayor que una de 300; una muestra de 200 casos, por otra parte, tendrá un error mayor si el universo tiene 10.000 unidades que si éste posee solamente 2.000.

Debemos advertir a los principiantes que el **error muestral nunca debe calcularse como un porcentaje del tamaño de la muestra respecto al del universo**. La variación de los errores al modificarse estas cantidades se da proporcionalmente, pero no de acuerdo a ecuaciones lineales. Para ello es preciso calcular, en cada caso, el error que podamos cometer, o porcentaje de riesgo, de acuerdo a los datos concretos disponibles y según ciertas ecuaciones bien conocidas en estadística.

Para fijar el tamaño de la muestra adecuado a cada investigación es preciso primero determinar el porcentaje de error que estamos dispuestos a admitir. Una vez hecho esto deberán realizarse las operaciones numéricas correspondientes para poder calcular el tamaño de la muestra que nos permite situarnos dentro del margen de error aceptado. Es decir que *no se fija primero el*

*número de unidades de la muestra para luego proceder a determinar el riesgo que se corre, sino a la inversa, se pone un límite a este riesgo y en función de ese valor se define el tamaño de la muestra que nos garantiza no sobrepasarlo*. A veces, sin embargo, el tamaño de la muestra queda determinado previamente por consideraciones prácticas, fundamentalmente económicas. En tales casos no hay otra alternativa que aceptar el nivel de error que su magnitud acarree.

### Ejercicios

7.1. Defina el universo que habría que estudiar para una investigación acerca de las preferencias de marcas de bebidas alcohólicas.

7.2. ¿Qué método de extracción de muestras elegiría para encontrar una de 500 adultos que viven en zonas rurales?

7.3. Para estudiar el comportamiento de diferentes sindicatos, teniendo como unidad de dato a sus afiliados, ¿qué muestra es preferible: una por conglomerados o una estratificada?

7.4. Analice la composición muestral de alguna encuesta efectuada en el país para conocer las opiniones políticas de los ciudadanos.

7.5. ¿Qué aplicaciones tiene el muestreo en las áreas de:

- 1) control de calidad
- 2) astronomía
- 3) difusión de enfermedades contagiosas?