

Población y muestra en los estudios cuantitativos¹

Uno de los aspectos centrales de la estrategia metodológica cuantitativa es la determinación de la población o grupo en que el estudio será realizado, habida cuenta que las unidades de observación han sido definidas al momento de elaborar los indicadores o referentes empíricos.

Siguiendo una definición estadística, **población** es el conjunto de individuos, objetos o elementos de los que se desea conocer algo en una investigación. Fayad Camel la define como: La totalidad de individuos, elementos o unidades, en los cuales puede presentarse determinada característica susceptible de ser estudiada, observada o medida.

Podría presumirse que si al investigador le interesa obtener explicaciones, predicciones de carácter general o afirmaciones generalizables a toda una población, el estudio debería hacerse abarcando todos los elementos de la población.

Sin embargo, no siempre es posible estudiar la población en su totalidad. Por ejemplo, cuando está constituido por un número inaccesible de elementos por ser éste muy grande, cuando está formado por una cantidad ilimitada de elementos, en cuyo caso se denomina **universo** y la imposibilidad de abordarlo exhaustivamente por no conocer su magnitud, por tratarse de experimentos que suponen la destrucción de los elementos a observar, etc.

Es por ello que en el proceso de definición de esa población en estudio, en la mayoría de los casos, y dependiendo de algunos criterios, es necesario escoger una parte de la misma para llevar a cabo el estudio. Esta parte o subconjunto de la población se denomina muestra. (figura 1).

Entendiendo que una muestra es el medio fáctico para conocer las características de una población, se deduce que la misma debe cumplir ciertos requisitos para que al momento de los resultados, éstos puedan ser generalizados o extrapolados al universo o población, del cual la muestra proviene.

La investigación a partir de muestras tiene muchas ventajas; entre éstas se destacan las siguientes:

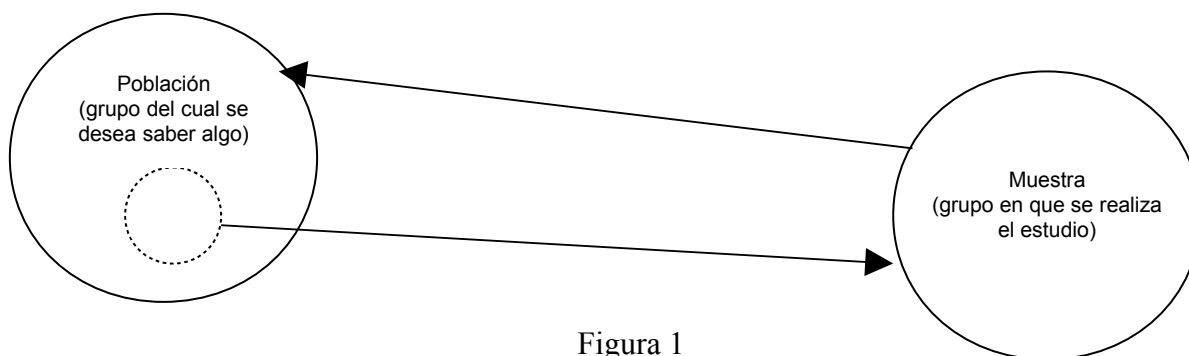


Figura 1

- Permite que el estudio se realice en menor tiempo, al ser menos los elementos observados
- Se reducen los gastos.

¹ Apuntes de clase: Z. T. De Quinteros

- Posibilita profundizar en la captación de otros indicadores empíricos respecto de los estudios exhaustivos o censos.
- Permite tener mayor control sobre la confiabilidad de la información a captar, a través de tareas de supervisión y otros controles de calidad.

En base a las consideraciones anteriores, una vez definida la población (universo), el siguiente paso es decidir si el estudio abarcará toda la población o si por razones de recursos y tiempo o ventajas para investigar el problema, se seleccionará una muestra de esa población.

De adoptarse la segunda alternativa, surgen dos problemas básicos en el diseño metodológico: cuántos individuos/elementos/unidades habremos de estudiar (tamaño de la muestra) y cómo elegiremos a esos individuos/elementos/unidades (método de selección de la muestra) para que los resultados que se obtengan puedan ser generalizados a la población en su totalidad.

Respecto del tamaño de la muestra, el número de casos no es arbitrario ni depende del tamaño de la población, como tampoco existen fórmulas standard para su cálculo. Para facilitar la comprensión de estas afirmaciones, tomemos un ejemplo ficticio en que se intenta valorar la prevalencia de tuberculosis en un área geográfica A, en un momento determinado. La población en este caso estaría conformada por todos los habitantes de dicha área. Una muestra de esa población podría ser cualquier grupo de habitantes del área que se tome para el estudio de prevalencia y la variable principal será "el diagnóstico" que resulte de cada individuo observado, admitiendo dos expresiones que pueden ser "enfermo" y "no enfermo de tuberculosis". Sin embargo, a partir de los hallazgos obtenidos en la muestra, por ejemplo, una tasa de prevalencia de tuberculosis de 15 por mil, pretendemos a posteriori resultados a toda la población, aunque no la hayamos investigado realmente.

Este proceso de generalización, peligroso e incierto desde el punto de vista teórico, se conoce con el nombre de inferencia estadística y consiste en inducir resultados generales a partir de hallazgos particulares proporcionando una medida, expresada en probabilidad, de la duda o confianza que merecen esas generalizaciones.

Es decir que, si a partir de la muestra la proporción de tuberculosis encontrada fue de 15 por mil, ella nos permitirá "estimar" para la población cuál es la verdadera proporción de tuberculosos. Esta estimación la haremos mediante un "intervalo de confianza" cuya técnica de cálculo no puede describirse aquí, y que nos permitirá concluir, por ejemplo que la proporción de enfermos en la población es un número entre 14.3 y 15.7 por mil, dicho con una seguridad de 95%. La precisión con que se ha formulado la estimación es en este caso 1.4 por mil (15.7- 14.3). Esta estimación de la tasa poblacional proviene del estudio de una muestra. La confianza o seguridad de esa estimación, que es de 95%, significa que de cada cien muestras similares a la obtenida, pensando en un muestreo reiterado, habría 95 intervalos que verdaderamente incluyen la tasa poblacional y otros cinco que no. Por lo tanto, a partir de este único intervalo logrado corremos un riesgo de 5% de error en nuestra estimación de la tasa poblacional, o sea que el intervalo calculado constituye una de esas cinco posibilidades que no incluyen la verdadera

proporción de tuberculosos.

Estas consideraciones, que corresponden a la etapa de análisis de resultados, fueron volcadas aquí por cuanto están íntimamente relacionadas con el tamaño de la muestra y con el método de selección de la misma, que son los dos aspectos básicos que deberán considerarse en el diseño de un estudio.

Siguiendo con el problema de definir el tamaño de la muestra, existen tres factores fundamentales que deben tenerse en cuenta, si no incluimos con ellos al costo, que, en buena medida y muy frecuentemente, determina el tamaño de la muestra. Estos tres factores son: "precisión, "confianza" y "variabilidad de la variable".

Al mencionar el término "precisión" nos referimos concretamente a " la amplitud que debe tener el intervalo de confianza en la estimación de la tasa o valor poblacional. En el ejemplo dado la precisión fue de 1.4 por mil. Esta precisión es impuesta por el propio investigador antes de realizarse la investigación.

Este dato, que se incorpora en el cálculo del tamaño de la muestra, está en relación directa con el número de elementos o individuos que se someterán a estudio, pues la muestra habrá de ser mayor cuando la precisión exigida también sea mayor (menor amplitud del intervalo).

Otro factor que afecta al tamaño de la muestra es el grado de confianza o precisión. También aquí hay relación directa entre confianza y tamaño de la muestra, pues a mayor confianza exigida, para una precisión dada, será necesaria una muestra de mayor tamaño. Es obvio que este aspecto también es impuesto por el propio investigador antes de realizar el estudio.

El tercer factor, variabilidad de la variable, incide también directamente sobre el tamaño de la muestra. Cuanto más heterogénea sea la variable habrá que habrá que extraer mas elementos, para que la muestra reúna en sí las diferentes expresiones de esa variable y puedan inferirse los resultados con la precisión y confianza establecidas. (En el ejemplo que se sigue, la variable diagnóstico positivo o negativo de tuberculosis, tendremos la máxima variabilidad, cuando la tasa de tuberculosos sea de 500 por mil, variabilidad que irá decreciendo a medida que la tasa disminuye de 500 a 0 por mil o aumente de 500 a 1000 por mil.)

La variabilidad es una característica propia de la variable que está en juego en el estudio. Es necesario entonces conocer esa información antes del cálculo definitivo de la muestra, ya que la condiciona en cuanto a su tamaño. Para ello se realizan estudios pilotos, que se conforman con pocos elementos de la población y que sirven además para ensayar la técnica de recolección de datos, formularios, etc.

Con respecto a los métodos de selección de la muestra, diremos que del procedimiento elegido dependerá no sólo la factibilidad de la generalización posterior a la población sino también la posibilidad de medición de la probabilidad de error de la inferencia. Esto puede hacerse sólo y únicamente cuando la muestra es elegida por algún procedimiento aleatorio, o sea, cuando la muestra es probabilística, ya que no se puede medir el grado de confianza que pueda atribuirse a una inferencia obtenida a partir de una

muestra no aleatoria.

La expresión aleatoria se refiere más concretamente a la manera de elegir la muestra misma. Toda posible muestra es una muestra aleatoria. Lo que sucede es que, recurriendo al azar para la selección de la muestra, se logra con más frecuencia aquello que es representativo de la población. O sea que la "aleatoriedad en la selección de la muestra" está directamente asociada a una condición de la misma, la de ser "representativa" de la población.

Otros mecanismos de selección de muestras (muestreo dirigido, por cuotas, etc.) a los que se apela por razones operativas o bien porque se carece de la información básica para un muestreo probabilístico conducen a estimaciones puntuales, en general a resultados más inciertos o débilmente aproximados a los valores reales de la población en estudio.

En relación con el muestreo probabilístico o aleatorio, se reconocen diferentes métodos según el diseño particular del estudio para controlar distintos factores condicionantes, intervinientes, o asociados con el problema investigado, como muestreo "estratificado", por "conglomerados", a "etapas múltiples", etc.

Todos ellos implican la utilización del método "aleatorio simple", en alguna etapa del muestreo y en la mayoría de los casos. Una noción básica en este tipo de muestreo es que cada elemento o individuo de la población tiene la misma probabilidad de ser incluido en la muestra. A través de un ejemplo se detallan los pasos a seguir en este método y donde N simboliza el tamaño de la población y n el correspondiente a la muestra.

Muestreo aleatorio simple

Supongamos que la población a investigar está constituida por 2.000 (N) historias clínicas de un hospital del cual se ha pensado extraer 100 (n) historias para proceder a cierto estudio.

El primer paso que deberá realizarse es enumerar todas las historias, lo que se hará en la forma que resulte más práctica y que irá de la 1 a la 2000. En general la numeración irá, entonces, de 1 a N . Para elegir las 100 historias al azar se podría utilizar un bolillero, o bien una tabla de números al azar, que resulta más rápido y práctico.

Una tabla de números al azar consta de una gran cantidad de dígitos (números 0, 1, 2 hasta el 9), como la de Fisher y Yates, que tiene 15.000 dígitos y están dispuestos totalmente al azar, como si se hubieran hecho 15.000 extracciones sucesivas con reposición de un bolillero en el que constan los números 0, 1, 2 hasta el 9. Estos dígitos aparecen en la tabla aproximadamente con la misma frecuencia.

Si se combinan en pares se puede leer números que van desde 00 hasta 99; si se usan de a tres, se tienen números que van desde 000 al 999, etc.

Para el ejemplo en que se deben extraer 100 historias clínicas de un total de 2000 habrá que combinar números de cuatro cifras, ya que pueden ser elegidas cualesquiera de las siguientes historias: 0001, 0002, 0003...1999 ó 2000. En general, el número de cifras que se incluirán en los números es igual al de las cifras que tiene el total de la población, N .

Para extraer los números de una tabla al azar se comienza en un lugar cualquiera, también elegido al azar. La tabla se lee de izquierda a derecha o de arriba hacia abajo. Lo importante es que partiendo de un determinado lugar se lean todos los números sin omitir ninguno.

Al hacer una determinada extracción, por azar, pueden repetirse algunos números. Estos habrán de ser sustituidos por otros de la tabla, que deberán ser tomados a continuación del último de la serie extraída al principio, de manera que finalmente se tendrán n números diferentes.