

Una elección importante

La vida está llena de decisiones importantes. Decidimos qué estudiar. A veces, muy pocas, podemos decidir dónde trabajar. Decidimos si nos casamos y con quién (o eso nos parece). Decidimos dónde vivimos, qué coche compramos, etc. Y la cagamos un número de veces considerablemente más elevado del deseable. ¿Creéis que no?. Entonces explicadme el significado de ese lamento que se oye con tanta frecuencia y que reza así: “si se viviese dos veces...”. Pues eso.

Por eso, antes de tomar una decisión hay que valorar cuidadosamente las alternativas de que disponemos. Y esto, válido para la mayor parte de los aspectos de la vida normal, es aplicable también al método científico, con la ventaja añadida de que suele estar más claramente establecido cuál puede ser la elección más correcta.

Y, ya que hablamos de elegir, supongamos que queremos tener una idea de cuál va a ser el resultado de las próximas elecciones. La forma de obtener el dato más aproximado sería preguntar a todos los votantes por su intención de voto, pero a nadie se le escapa que esto puede ser imposible desde un punto de vista práctico. Pensemos en un país grande con cincuenta millones de votantes. En estos casos lo que hacemos es elegir un subconjunto de la **población**, al que llamamos **muestra**, hacemos la encuesta entre sus componentes y estimamos el resultado en la población general.

Pero podemos preguntarnos, ¿es fiable esta estimación?. Y la respuesta es sí, siempre que empleemos una **técnica de muestreo** válida que nos permita obtener una muestra representativa de la población. Todo dependerá de dos características del muestreo: su **exactitud** y su **precisión**.

La exactitud determina la proximidad entre el resultado que obtenemos en la muestra y el valor real inaccesible en la población y depende del tipo de muestra elegido. Para que sea exacta, la muestra ha de ser **representativa**, lo que quiere decir que la distribución de la variable de estudio (y de las variables relacionadas) tiene que ser similar a la de la población. Habitualmente se empieza definiendo el **marco muestral**, que es el listado o forma de identificar a los individuos de la población a los que podemos acceder, llamados **unidades de muestreo**, y sobre los que se aplicará el proceso de selección. Pensemos, por ejemplo, en un censo de población, un listado de historias clínicas, etc. La elección del marco debe hacerse de forma muy cuidadosa, ya que condicionará la interpretación de los resultados.

Por otra parte, la precisión depende del tamaño de la muestra y de la variabilidad entre los participantes, como recordaréis de la fórmula de los **intervalos de confianza** ($IC_{95\%} = \text{valor} \pm 1.96 \times \text{error estándar}$). Como el

error estándar es el cociente de la desviación típica entre la raíz cuadrada del tamaño de la muestra (n), a mayor desviación típica o menor tamaño de muestra, mayor amplitud del intervalo de confianza y menor precisión. Pero esto es una verdad a medias que solo sirve si asumimos que la población tiene un tamaño infinito porque, en realidad, el error estándar debe multiplicarse por un factor de corrección, para muestras finitas, que tenga en cuenta el tamaño de la población (N), con lo que la verdadera fórmula del intervalo sería la siguiente:

$$IC95\% = \text{valor} \pm 1,96 \times \text{error estándar} \times \sqrt{1 - \frac{n}{N}}$$

¡Alto!. Ahora no aceleréis el ritmo de lectura porque haya escrito una fórmula y volved a mirarla para contemplar, una vez más, la magia del [teorema central del límite](#), el Sancta Sanctorum de la inferencia estadística. Si la población (N) es muy grande, el cociente rápidamente se hace casi cero y el error se multiplica casi por uno, con lo cual prácticamente el intervalo no cambia. Y esto no es ninguna tontería, ya que explica por qué con una muestra de 1200 votantes podemos estimar, con muy poco margen de error, tanto las elecciones para alcalde de Nueva York, como las de presidente de EE.UU. o las de emperador del Mundo mundial, si lo tuviésemos (siempre, claro está, que cada muestra sea representativa de cada censo electoral). Por otra parte, si n se va aproximando a N, el factor de corrección se aproxima a cero y el intervalo se va haciendo cada vez más estrecho. Esto explica que, si n=N, el valor que obtenemos coincide con el real de la población, según el teorema de Pero Grullo.

Así que, siendo tal el poder de una insignificante muestra, a nadie puede extrañar que sean diversas las formas de elegirla. Las primeras que vamos a ver son las técnicas de [muestreo probabilístico](#), en las que todos los sujetos tienen una probabilidad conocida, distinta de cero, de ser seleccionados, aunque no es obligatorio que todos tengan la misma probabilidad. Normalmente se utilizan métodos aleatorios para la selección, con lo que se evita la subjetividad del investigador y se minimiza la posibilidad de que, por azar, la muestra no sea representativa, lo que se conoce como [error aleatorio](#) o [de muestreo](#). Como siempre, no podemos desembarazarnos del azar, pero sí cuantificarlo.

El más conocido es el [muestreo aleatorio simple](#), en el que cada unidad muestral del marco tiene la misma probabilidad de ser elegido. Lo más frecuente es realizar un [muestreo sin sustitución](#), que quiere decir que, una vez elegido, el participante no se reintegra a la población para que no se pueda elegir más de una vez. Para hacer las cosas bien, el proceso de selección a partir del marco se hace con una tabla de números aleatorios o un algoritmo informático.

En ocasiones, la variable no se distribuye uniformemente en la población. En estos casos, para que la muestra sea representativa puede dividirse la población en estratos y hacer un muestreo aleatorio en cada estrato. Para poder realizar esta técnica, llamada **muestreo aleatorio estratificado**, necesitamos conocer la distribución de la variable en la población. Además, los estratos deben ser excluyentes entre sí, de forma que la variabilidad dentro de cada uno sea mínima y la variabilidad entre estratos sea la mayor posible. Si los estratos tienen un tamaño similar el muestreo se hace de forma **proporcional**, pero si alguno es más pequeño puede sobre representarse e incluir más unidades de muestreo que el resto. El problema es que el análisis se complica, ya que hay que ponderar los resultados de cada estrato en función de su contribución al resultado global, pero los programas estadísticos hacen estas correcciones sin inmutarse. La ventaja de esta técnica es que las estimaciones que se obtienen, si se analizan los datos correctamente, son más precisas, ya que la varianza global se calcula a partir de la de los estratos, que siempre será menor que la de la población general. Este tipo de muestreo es muy útil cuando la variable de estudio se influye por otras variables de la población. Si queremos estudiar, por ejemplo, la prevalencia de cardiopatía isquémica puede ser útil estratificar por sexo, peso, edad, fumador o no, o lo que pensemos que pueda influir en el resultado.

Un paso más allá de este enfoque está el del **muestreo en etapas múltiples** o **muestreo por conglomerados** (en clusters, para los que lean inglés). En este caso la población se divide en **unidades primarias de muestreo** que, a su vez, se dividen en **unidades secundarias** en las que se lleva a cabo el proceso de selección. Este tipo, con todas las etapas que nos interese en cada caso, es muy usado en estudios escolares, en los que se van separando por nivel socioeconómico, tipo de enseñanza, edad, curso o lo que se nos ocurra. El problema de este diseño, aparte de su complejidad de implementación y análisis de resultados, es que podemos tener sesgos si los miembros de una unidad se parecen mucho. Pensad, por ejemplo, que queremos estudiar la tasa de vacunaciones en una ciudad: dividimos la ciudad en zonas, de cada zona seleccionamos al azar unas familias y vemos cuántos niños están vacunados. Lógicamente, si un niño está vacunado seguramente sus hermanos también lo estarán, con lo que puede sobreestimarse la tasa de vacunación global de la ciudad si en el muestreo nos tocan muchas familias numerosas de las zonas con mejor nivel sanitario.

El **muestreo sistemático** se utiliza con frecuencia en estudios en los que el marco muestral no existe o es incompleto. Por ejemplo, si queremos probar un antigripal, no sabemos quién va a contraer la gripe. Elegimos una constante de aleatorización (k) y esperamos tranquilamente a que lleguen los enfermos a la consulta. Cuando hayan llegado los k primeros, elegimos uno al azar y, a partir de ahí, incluimos a uno de cada k que vengan con gripe hasta completar el tamaño muestral deseado.

En todas las técnicas anteriores se conocía la probabilidad de cada integrante de la población de ser seleccionado. Sin embargo, esta probabilidad se desconoce en los [modelos no probabilísticos](#), en los que no se utilizan métodos aleatorios, por lo que hay que tener especial cuidado con la representatividad de la muestra y la presencia de sesgos.

El [muestreo consecutivo](#) se usa con frecuencia en ensayos clínicos. En el ejemplo anterior de la gripe podríamos enrolar en el ensayo a los n primeros que acudiesen a consulta y que cumplieren los criterios de inclusión y exclusión del ensayo. Otra posibilidad es la de [inclusión de voluntarios](#). Esta no es muy recomendable, ya que los sujetos que aceptan participar en un estudio sin que nadie se lo pida pueden presentar características que afecten la representatividad de la muestra.

Los especialistas en mercado utilizan mucho el [muestreo por cuotas](#), seleccionando a los sujetos según la distribución de las variables que a ellos les interesan, pero este tipo de diseño se emplea poco en medicina. Y, por último, comentemos el uso de [técnicas adaptativas](#), como el llamado [muestreo en bola de nieve](#), [muestreo en paseos aleatorios](#) o [muestreo en red](#). Por ejemplo, pensemos que queremos hacer un estudio con adictos a alguna sustancia ilegal. Nos costará encontrar a los participantes, pero una vez que encontremos al primero podemos preguntarle si conoce a alguien más que pudiese participar. Esta técnica que, aunque pueda parecerlo, no me acabo de inventar, tiene su utilidad para poblaciones muy difíciles de alcanzar.

Y con esto terminamos con las técnicas que tratan de conseguir el tipo de muestra más adecuado para nuestro estudio. Nos quedaría hablar del [tamaño de la muestra](#) y de cómo debe calcularse previamente al inicio del estudio para que no sea ni demasiado grande, ni demasiado pequeña. Pero esa es otra historia...
