

Experimentando con muestras

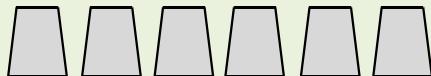
(Actividad didáctica recomendada para Primaria y Secundaria)

El **muestreo** es una técnica muy eficaz cuando no podemos medir a toda la población que pretendemos estudiar, ya sea por una limitación de tiempo o de costes. Consiste en estudiar solo a una parte de los individuos de manera que las conclusiones que obtengamos puedan ser aplicables a toda la población.

Vamos a hacer un pequeño experimento en el que veremos cómo **se obtienen distintos resultados dependiendo de la forma de obtener la muestra**.

Material necesario

6 vasos desechables, no transparentes



6 trozos de papel numerados del 1 al 6



12 bolitas de papel pequeñas (tamaño garbanzo)



1 dado



Supongamos que los vasos representan viviendas y las bolitas son personas. Sabemos que hay 6 viviendas pero vamos a suponer que no conocemos cuantas personas hay viviendo en ellas.

Queremos estudiar cuánta población hay residiendo en esas viviendas pero no disponemos de tiempo y ni recursos para ir consultando cada vivienda por lo que tenemos que obtener una muestra. Podemos permitirnos estudiar 3 viviendas, por lo tanto, **el tamaño de la muestra será 3**.

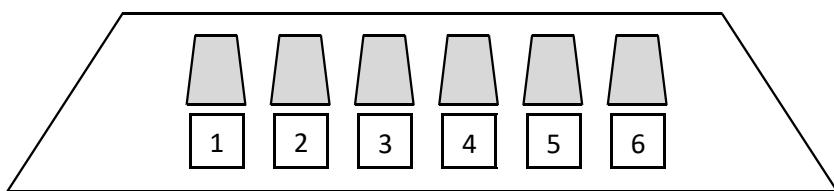
Paso 1

Distribuimos las bolitas en grupos tapándolas con los vasos bocabajo de la siguiente forma:

- 2 vasos con 3 bolas
- 2 vasos con 2 bolas
- 2 vasos con 1 bola

Paso 2

Una vez tenemos todos los vasos bocabajo, los mezclamos para perder de referencia su ocupación de bolas y los disponemos en fila. Asignamos un número a cada uno de los vasos haciendo uso de los trozos de papel numerados.



Paso 3

Tiramos el dado tres veces hasta obtener tres números, se repite el lanzamiento tantas veces como sea necesario en caso de repetición de resultado.

Paso 4

Levantamos la muestra obtenida de tres vasos, los tres que ha seleccionado el dado, para observar cuantos individuos habitan cada vivienda. Anotamos los resultados.

Paso 5

Estimamos el total de la población a partir de la muestra. Como el número total de viviendas es 6 y la muestra es de 3, el total estimado se calcularía:

$$(\text{Suma de individuos en la muestra}) \times (6/3)$$

Paso 6

Vamos a repetir el experimento 10 veces, anotando los resultados en una tabla y calculando la diferencia respecto al valor real. Así comprobaremos cómo de acertada o no ha sido nuestra técnica. Al tratarse de una estimación, lo normal es que siempre se cometa un error mayor o menor.

Un ejemplo de repeticiones del experimento puede ser el siguiente:

Repetición	Individuos en las viviendas de la muestra			Suma	Estimación	Total real	Error
	1	2	3				
1	3	2	3	8	16	12	+4
2	2	2	1	5	10	12	-2
3	1	1	2	4	8	12	-4
4	3	2	1	6	12	12	0
5	1	2	1	4	8	12	-4
6	2	2	3	7	14	12	+2
7	2	1	2	5	10	12	-2
8	3	1	1	5	10	12	-2
9	1	1	2	4	8	12	-4
10	3	1	3	7	14	12	+2

En este ejemplo de realizaciones del experimento se observa cómo se suele cometer un error que puede llegar a +4 o -4 respecto al valor real, es decir, la estimación del total puede ir de 8 a 16. Es un error bastante grande teniendo en cuenta que el real es 12.

El tipo de muestreo empleado se denomina **muestreo aleatorio simple**, ir seleccionando individuos al azar.

Las **formas de mejorar una estimación** pueden ser: ampliar el tamaño de la muestra, en lugar de escoger 3 individuos escoger por ejemplo 4, pero eso supone un esfuerzo de tiempo o dinero que en la vida real puede no ser viable; o mejorar la técnica de muestreo.

Para obtener buenos resultados es necesaria una planificación previa, **adecuar el tipo de muestreo a la población que estamos estudiando**, esto suele requerir una observación general en busca de pistas antes de ejecutar la acción, o acceder a información auxiliar.

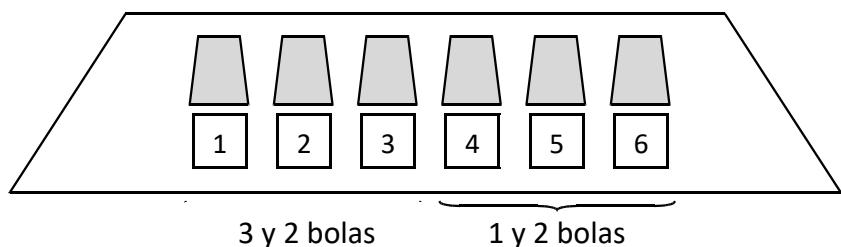
Mejorando la técnica

Vamos a suponer que creemos, por lo que se observa a simple vista, que las viviendas menos ocupadas están en el centro de la localidad y que en las más alejadas viven más personas.

Procedemos de la siguiente forma desde el paso 2:

Paso 2

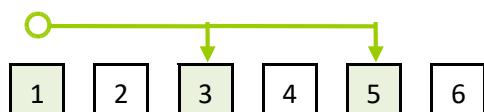
Hacemos dos grupos: los dos vasos que tienen 3 bolas y uno de 2 bolas; y los otros dos con 1 bola y el de 2 bolas restante. Mezclamos los vasos de cada uno de los grupos por separado, es decir, sin mezclar los vasos de distintos grupos. Los ponemos todos en fila de manera que la población de vasos queda partida entre los que más población de personas tienen y los que menos, las que viven en las afueras y en el centro respectivamente.



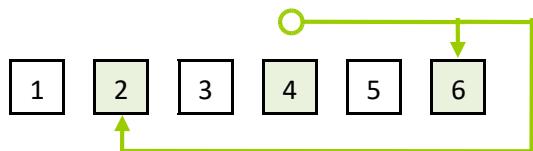
Paso 3

Tiramos el dado una sola vez. Seleccionamos el número de vaso que se corresponde con el número aleatorio que sale y a partir de él se selecciona el resto de la muestra (Es lo que se conoce como **muestreo sistemático**). Hay que ir cogiendo vasos dando saltos de $Tamaño de la población / Tamaño de la muestra = 6/3 = 2$, de esta forma se acaba seleccionando una muestra del tamaño necesario sin repetir elementos. Ejemplos:

Sale el 1:



Sale el 4:



Paso 4

Levantamos la muestra obtenida de tres vasos para observar cuantos individuos habitan cada vivienda. Anotamos los resultados.

Paso 5

El total se sigue estimando como:

$$(\text{Suma de individuos en la muestra}) \times (6/3)$$

Paso 6

De nuevo repetimos el experimento 10 veces, anotando los resultados en una tabla y calculando la diferencia respecto al valor real.

Un ejemplo de repeticiones del experimento puede ser el siguiente:

Repetición	Individuos en las viviendas de la muestra			Suma	Estimación	Total real	Error
	1	2	3				
1	3	2	2	7	14	12	+2
2	3	1	1	5	10	12	-2
3	2	2	1	5	10	12	-2
4	3	1	1	5	10	12	-2
5	2	2	1	5	10	12	-2
6	3	3	1	7	14	12	+2
7	3	2	1	6	12	12	0
8	2	3	2	7	14	12	+2
9	2	3	1	6	12	12	0
10	3	1	2	6	12	12	0

Con el conocimiento previo y una nueva técnica de muestreo, hemos reducido el error de estimación a +2 o -2 respecto al valor real, es decir, la estimación del total puede ir de 10 a 14.

Consideraciones

En este ejemplo trabajamos con una población muy pequeña, eso hace que las oscilaciones en el valor estimado sean grandes respecto al valor real. Cuanto mayor sea la población y considerando un tamaño de muestra aceptable, tendremos mejores estimaciones, es decir, más aproximadas al valor real (Ley de los grandes números).

Existen otros tipos de muestreo básicos que puedes tratar de aplicar en tus propios experimentos, pero necesitaríamos una población mayor que la de este sencillo ejemplo para poder aplicarlos con efectividad. A continuación te hacemos una propuesta que puedes poner en práctica en clase.

Propuesta: Altura media de clase

Vamos a calcular la altura media de los alumnos de clase. Para ello vamos a aplicar distintos tipos de muestreo y comparar las estimaciones obtenidas. Podemos tomar como el tamaño de la muestra a una tercera parte de la clase.

Procedemos a aplicar estos tipos de muestreo y calculamos la media para cada conjunto de datos obtenidos:

- **Muestreo aleatorio simple.** Se escogen los alumnos al azar por cualquier procedimiento aleatorio objetivo. Con este método, puede que seleccionemos a todos los alumnos de una determinada zona del aula: puede que los compañeros del equipo de baloncesto se sienten juntos, o los más altos se sienten detrás para no molestar la visión de la pizarra,...
- **Muestreo sistemático.** Para asegurarnos de una distribución homogénea aplicamos el muestreo sistemático. Se selecciona un elemento y el resto a saltos de tamaño 3 (*Tamaño de población/Tamaño de muestra*) hasta completar la muestra.
- **Muestreo estratificado.** Se divide a la clase en grupos, que llamamos estratos, de forma que los elementos dentro de cada estrato sean parecidos entre sí. Podemos dividir la clase previamente por sexos, considerando dos estratos, ya que globalmente suelen darse diferencias de altura entre mujeres y hombres, y seleccionar para la muestra obligatoriamente elementos de los dos grupos, en la misma proporción que hay de cada sexo en el global de clase. Así nos aseguramos de que el valor estimado se aproxime más al real que si no tuviéramos el sexo en cuenta.
- **Muestreo por conglomerados.** También se hace la división en grupos, que llamamos conglomerados, de manera que los elementos dentro de cada grupo sean distintos entre sí. Lo que hacemos es seleccionar uno de los conglomerados y examinar todos los elementos de ese conglomerado. Podemos hacer tres grupos que se compondrán de chicos y chicas en la misma proporción que en el global de la clase. Se selecciona un solo grupo y se miden a todos los elementos del mismo.

Cuando terminemos con todos los tipos de muestreo podemos recoger los datos de todos y comparar los resultados de cada tipo con el valor real y así obtener el error que hemos cometido en cada estimación. ¿Qué método ha sido el mejor?

Nota: Cuando se realiza una encuesta debe preservarse la información recogida, ya que así se evita que se haga un mal uso o uno distinto del que estaba previsto y del que se dijo al informante. Siempre que se midan características físicas de alumnos jóvenes se debe tener cuidado con no exponerlos frente a complejos o sensibilidades. En este caso, para no destacar a nadie con valores altos o bajos de altura, se les puede pedir que anoten su altura y sexo en un papel, que entregaran en caso de ser seleccionados. El valor se registra anónimamente junto con el resto de la muestra. Aparte del anonimato, se agiliza el experimento al no tener que medir explícitamente a nadie, ni siquiera importa que sea el valor real, solo coherente.