

DEFINICIÓN TÉCNICA

Una muestra estadística es un subconjunto de datos perteneciente a una población de datos. Estadísticamente hablando, debe estar constituido por un cierto número de observaciones que representen adecuadamente el total de los datos.

La estadística, como rama de las matemáticas, se encarga de recoger datos, ordenarlos y analizarlos. Es decir, cuando queremos estudiar un determinado fenómeno recurrimos a la estadística. Un buen ejemplo de fenómeno que estudia la estadística es el salario medio de los ciudadanos de un país

En este sentido, por cuestiones de tiempo y coste, no podemos recoger la totalidad de los datos. Esta totalidad de los datos es lo que se conoce como **población** de datos o, simplemente, **población**.

¿Por qué se trabaja con muestras estadísticas?

Para explicar por qué se utiliza una muestra estadística en lugar de la **población** total, vamos a recurrir al ejemplo planteado anteriormente.

Supongamos que queremos estudiar un fenómeno cualquiera. En nuestro caso, ese fenómeno es el salario medio de los ciudadanos de un país. La **población** de datos está formada por todos y cada uno de los trabajadores del país. Claro que por razones de tiempo y coste sería imposible ir preguntando a cada trabajador cuál es su salario anual. Tardaríamos mucho tiempo o necesitaríamos muchos recursos.

En este punto aparece el concepto de muestra estadística. En lugar de preguntar a los millones de trabajadores de un país o región, tan solo recogemos una pequeña cantidad de datos. Por ejemplo, preguntamos a 100.000 personas. Esta tarea sigue siendo complicada, pero es mucho más asequible preguntar a 100.000 personas que preguntar a 45 millones.

Esta pequeña cantidad de datos ha de ser representativa. Es decir, debe representar adecuadamente a la **población**. Si las 100.000 personas a las que preguntamos se concentran en barrios ricos, obtendremos datos que no son representativos. El salario medio nos saldría mucho más alto de lo que es en realidad.

Características de una muestra estadística representativa

Si se quiere hacer una buena investigación, la calidad de la muestra estadística es esencial. De nada sirve realizar las métricas estadísticas más complejas con los modelos más sofisticados si la muestra estadística está sesgada. Es decir, si la muestra no es representativa.

A la hora de obtener una muestra representativa existen ciertos aspectos que el investigador debe conocer de antemano. Entre esos aspectos se encuentran las

características de una muestra representativa. Las características de una muestra representativa son las siguientes:

- **Tamaño suficientemente grande:** Cuando trabajamos con muestras estamos, normalmente, trabajando con una cantidad de datos inferior a la **población**. Ahora bien, para que una muestra estadística sea representativa deberá ser lo suficientemente grande como para considerarse representativa. Por ejemplo, si nuestra **población** está formada por 10 millones de datos y escogemos 10, es difícil que sea representativa. Eso sí, no siempre a mayor tamaño la muestra es más representativa.
- **Aleatoriedad:** La selección de los datos de una muestra estadística debe ser aleatoria. Es decir, debe ser totalmente al azar. Si en lugar de realizarlo al azar, realizamos un proceso de selección de datos planificado, estamos introduciendo un sesgo a la obtención de datos. Por tanto, para evitar que la muestra sea sesgada y, por tanto, conseguir que sea una muestra representativa, debemos hacer una selección aleatoria.

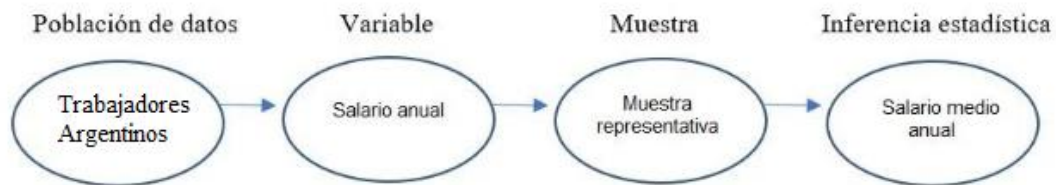
Tipos de muestra estadística

A continuación, te explicamos los diferentes tipos de muestra estadística que hay. Antes de nada, cabe destacar que se pueden dividir en dos grandes grupos, muestra probabilística y muestra no probabilística:

- **Muestra probabilística:** En este tipo de muestras todos los sujetos disponibles tienen las mismas probabilidades de ser incluidos.
 - **Muestra Aleatoria Simple:** Es un conjunto de variables aleatorias independientes e idénticamente distribuidas, obtenidas a partir de la variable aleatoria X y que se distribuyen igual que la misma.
 - **Muestra Aleatoria Sistemática** En este caso la **población** se enumera y se agrupa en grupos de 10 personas. Posteriormente, se selecciona a un miembro de cada grupo para elaborar la muestra.
 - **Muestra Aleatoria por Conglomerados:** La **población** se encuentra ya agrupada previamente y de estos grupos se extraen los individuos para conformar la muestra.
 - **Muestra Estratificadas** En este caso la **población** se divide en subgrupos o estratos con base en las variables de estratificación.
- **Muestra no probabilística:** En este tipo de selección de muestra todos los elementos no tienen la misma probabilidad de ser elegidos, ya que depende del procedimiento escogido para seleccionarlos.
 - **Bola de nieve:** En primer lugar, se seleccionan a diferentes sujetos. A partir de ahí estos sujetos colaboran para encontrar a más sujetos que tengan relación con ellos.
 - **Muestra por cuotas:** La **población** es elegida en función a unas características determinadas.
 - **Muestra discrecional:** La selección de la **población** la realizan los investigadores en función a su propio criterio.
 - **Muestra por conveniencia:** Es una muestra elegida por los propios investigadores según su interés o cercanía.

Inferencia estadística

Una vez obtenidos tenemos la muestra representativa, entonces toca inferir ciertas métricas. A menudo, lo que nos interesa es saber cierta medida de una variable. En el ejemplo inicial, la variable sería el salario de los ciudadanos de un país. En este sentido, la métrica que queremos analizar es la media del salario de los ciudadanos de un país.



Es decir, tenemos una **población** de datos formada por todos los trabajadores de Argentina. De dicha **población** obtenemos una variable, es decir, el salario anual. Utilizando las técnicas adecuadas obtenemos una muestra representativa. Y, por último, una vez tenemos un conjunto de datos con el que podemos trabajar utilizamos técnicas de inferencia estadística para calcular el salario medio.

Por supuesto, una vez tenemos el conjunto de datos, podríamos inferir otras medidas. Por ejemplo, cómo se distribuye el salario, qué porcentaje de trabajadores se encuentra por debajo de cierto salario o de qué tamaño es la brecha salarial.

Ejemplo de muestra estadística

Supongamos que queremos realizar un estudio sobre el gasto medio de las familias de Colombia en el mes de enero. Para ello tenemos dos opciones:

1. Entrar en las cuentas bancarias de todas las familias de Colombia
2. Preguntar a una cantidad de personas representativa

La primera opción no es viable por varias razones. Primero que las familias no van a ceder sus datos y segundo que tampoco podíamos ir familia por familia mirando los datos.

Principalmente, porque la **población** de Colombia se encuentra cerca de los 50 millones. Mientras, la segunda es la opción para recoger una muestra estadística.

Lo que haremos, siguiendo las características mencionadas anteriormente, será preguntar a 100.000 familias. Es algo complicado, pero mucho más fácil que preguntar a 50 millones de colombianos. La diferencia es considerable. Así pues, a partir de esa muestra de 100.000 familias, intentaremos calcular el gasto medio de las familias en enero.

Los datos extraídos serán más o menos fiables según una serie de métricas que se tienen en cuenta en las investigaciones estadísticas. Claro que, ese tipo de métricas son más avanzadas y, por ello, no las trataremos aquí.

Otro ejemplo de muestra estadística podría ser conocer la opinión de la gente de un determinado barrio acerca de que se cree un restaurante de comida mexicana en la zona en

la que residen. Las opciones para realizar este muestreo pueden ser consultarle a la gente en la calle para saber si irían a comer o cenar a este nuevo restaurante.

En cambio, otra opción sería que nuestro grupo de amigos enviase entre sus grupos de conocidos un cuestionario consultando esta información. La gente que respondiese sería la muestra.