

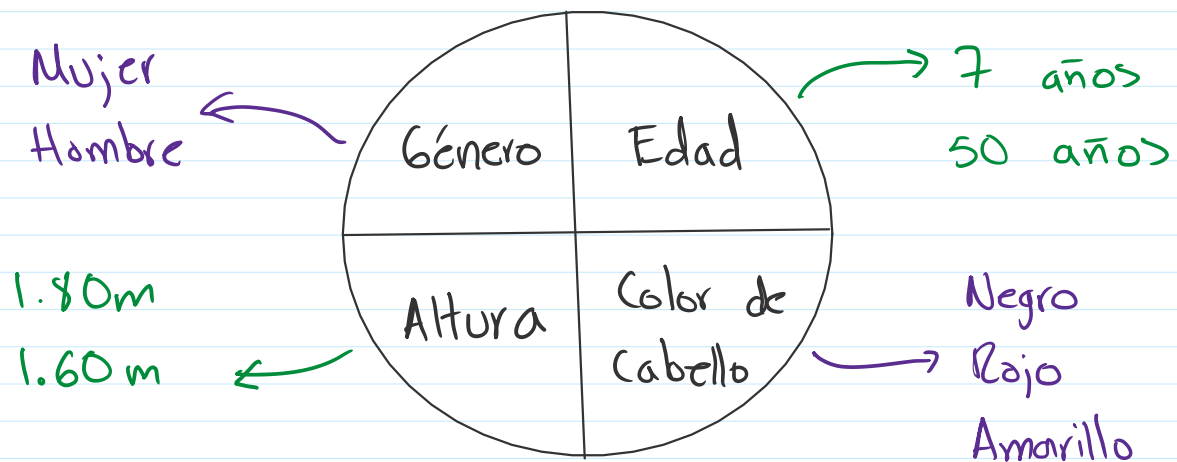
La Estadística se ocupa fundamentalmente de la recolección, organización, representación y estudio de datos para obtener conclusiones de cierto fenómeno o situación. A menudo la cantidad de datos disponibles es tan grande que es poco práctico o casi imposible recolectarlos todos. Por tal motivo, es necesario introducir los conceptos de **población** y de **muestra** que se analizarán más adelante.

Empecemos pues, con los conceptos fundamentales. En primer lugar, ¿qué datos vamos a recolectar? Bueno, dichos datos deben ser característicos de la situación o fenómeno a estudiar (es decir, deben ser datos **significativos**). Dichas características tienen un nombre especial.

En estadística, una **variable** tiene dos características que la definen

- 1) Una variable es un atributo que describe a una persona, cosa, lugar, idea.
- 2) El valor de la variable puede cambiar de una entidad a otra.

Por ejemplo, variables que pueden describir a una persona son



Las variables pueden ser clasificadas en dos grupos: **cualitativas** (o "categóricas") y **cuantitativas** (o "numéricas"):

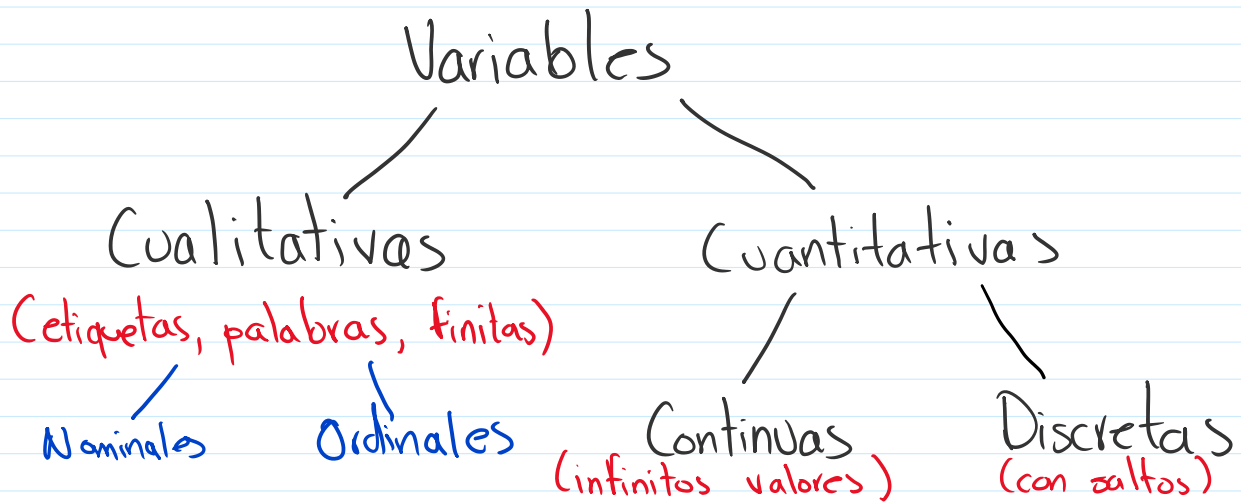
- **Cualitativas.** - Son aquellas variables cuyos valores son palabras o etiquetas. Por ejemplo el color de cabello de una persona, el género, tipo de sangre, nacionalidad.
- **Cuantitativas.** - Son aquellas cuyos valores son numéricos. Representan una cantidad medible. Por ejemplo la población de una ciudad o la estatura de una persona.

Nota que en general las variables cualitativas adquieren sólo una cantidad finita de valores, a los cuales se les llama **niveles**.

Género	Tipo de Sangre	Color de Cabello
Niveles → M ₂ F	A B AB ₄ U	R C N ₄ Ro

Las variables cuantitativas pueden clasificarse a su vez en **continuas** o **discretas**.

- **Continuas.** - Son aquellas variables numéricas que pueden tomar TODOS los valores entre un valor máximo y un valor mínimo. Por ejemplo: el peso de un boxeador gallo, la altura de una persona, el precio del dólar. Así pues, la cantidad de valores posibles son infinitos.
- **Discretas.** - Son aquellas variables numéricas que no pueden tomar todos los valores en un rango. Por ejemplo: la edad de una persona, la población de una ciudad.
Típicamente, las variables discretas tienen "saltos" entre sus valores.
Otro ejemplo clásico de este tipo de variable, es aquella que sólo tenga valores enteros.



Como mencionamos anteriormente, la estadística analiza conjuntos de datos. Dos tipos muy importantes de conjuntos de datos son los siguientes:

- **Población.** - Incluye TODOS los elementos del conjunto de datos.
- **Muestra.** - Consiste de un subconjunto de observaciones de la población. Por tal motivo, múltiples muestras pueden ser obtenidas de una sola población.

Así, la población y la muestra se diferencian por la cantidad de elementos del conjunto de datos que se analizan.

- **Ejemplo.** - Una empresa farmacéutica quiere analizar los posibles efectos secundarios que uno de sus medicamentos pudiera tener en las personas. Para tal motivo, la empresa administra el medicamento a 1000 voluntarios diariamente durante 30 días. En este caso

Población: Todos los seres humanos vivos a la fecha.
Muestra: 1000.

¿Por qué usar una muestra? Como se mencionó al principio, podría ser complicado o prácticamente imposible considerar TODOS los datos. De hecho, uno de los objetivos de la Estadística es obtener conclusiones válidas para TODA la población, a partir de una o varias muestras. Otra razón para usar una muestra, es que podría ser "peligroso" usar toda la población (como en este ejemplo).

Entre población y muestra existen también diferencias entre nomenclatura, fórmulas y notación. En cuestión de nomenclatura, una característica medible de una población se denomina **parámetro**, mientras que en una muestra se denomina **estadística**. Por ejemplo si calculamos el promedio de un conjunto de datos, éste es un "parámetro de la población", mientras que el promedio de una muestra es una "estadística de la muestra".

En cuanto a fórmulas, más adelante se verá que, por ejemplo, las fórmulas para la varianza de la población y la muestra son distintas.

En cuanto a nomenclatura, los atributos de la población se representan con letras griegas y los de la muestra con letras latinas.

Ahora surge la pregunta: ¿cómo obtener muestras de una población? Existen muchas maneras de hacerlo, pero nos enfocaremos en una particular: **muestreo simple aleatorio**.

Este "método de muestreo" se caracteriza por tres propiedades:

1) La población consiste de N elementos.

2) La muestra consiste de n elementos.

3) Todas las muestras de n elementos son igual de probables de ocurrir.

→ Importante ⇒ Representatividad

Debe ser obvio por qué usamos el muestreo aleatorio. Queremos que cualquier dato de la población sea igual de probable de elegir que cualquier otro. ¿Por qué? Porque queremos que la muestra sea **representativa** de la población. De esta manera, las conclusiones obtenidas de la muestra podrán generalizarse a toda la población.

• **Ejemplo.** - ¿Qué pasaría si quisiéramos estudiar el desempeño académico de los estudiantes del Tec de Monterrey y usamos:

Población
Todos los estudiantes
del Tec.

Muestra
100 mejores promedios?

Es obvio que obtendríamos conclusiones que no podrían generalizarse a toda la población.

Es obvio que obtendríamos conclusiones que no podrían generalizarse a toda la población.

Una manera para obtener muestras simples aleatorias es el "método de la lotería", el cual consiste en asignar números a cada elemento de la población y seleccionar n números de manera aleatoria (por ejemplo, por computadora). Un ejemplo relacionado es la manera en la que se asignan funcionarios de casilla por apellido.