

Estadística descriptiva

Jessica Nathaly Pulzara Mora
jessica.pulzara@udea.edu.co

Departamento de ingeniería de sistemas



**UNIVERSIDAD
DE ANTIOQUIA**

En muchas áreas del quehacer científico es común la realización de experimentos con la finalidad de verificar afirmaciones acerca de comportamientos de ciertos procesos de la naturaleza, la realización de encuestas para determinar la aceptación hacia un nuevo producto o persona en particular, la realización de estudios de seguimiento a través del tiempo o la recolección de datos de diversas fuentes.

¿Qué es la estadística?

La estadística es la ciencia de describir o hacer inferencias sobre el mundo a partir de una muestra de datos.



- La Estadística se puede definir como la ciencia que estudia cómo debe emplearse la información y cómo dar una guía de acción en situaciones prácticas que entrañan incertidumbre.
- Se usa para realizar **inferencia** con los datos. Esto permite tomar decisiones y formular predicciones..
- Se ocupa de los métodos y procedimientos para recoger, clasificar, resumir, hallar comportamientos y analizar los datos, siempre y cuando la variabilidad e incertidumbre sea una causa intrínseca de los mismos.

Clasificación

Descriptiva

Inferencial



Estadística descriptiva

Estadística Descriptiva: organiza, describe y representa un grupo de datos utilizando métodos numéricos y gráficos que resumen y presentan la información contenida en ellos.



Estadística inferencial

Estadística Inferencial: Apoyándose en el cálculo de probabilidades y a partir de datos muestrales, efectúa estimaciones, decisiones, predicciones u otras generalizaciones sobre un conjunto mayor de datos.



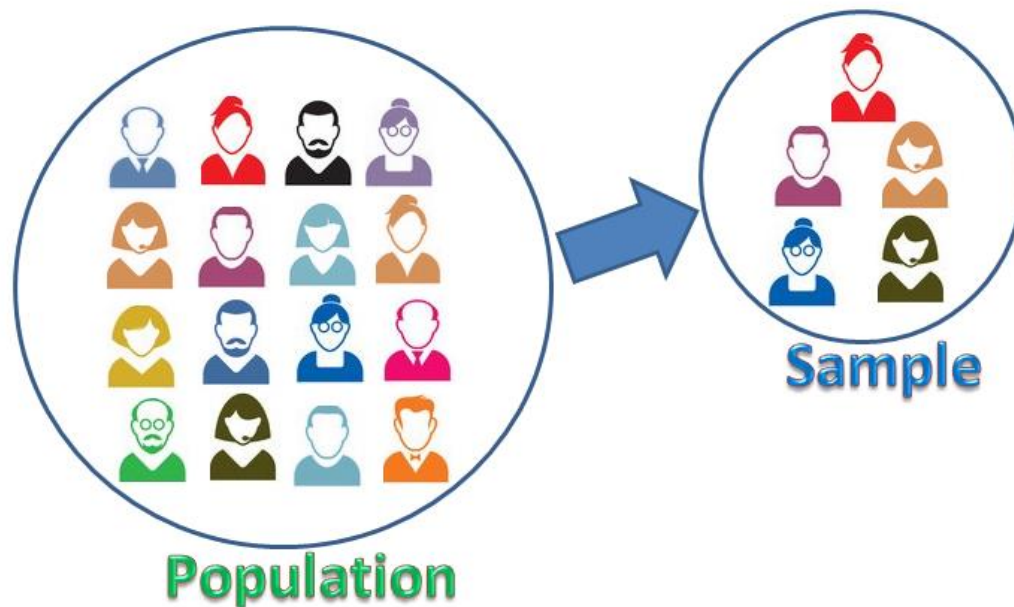
Población

Población: es un conjunto finito o infinito de elementos (personas, animales o cosas) que presentan características comunes, sobre los cuales se quiere efectuar un estudio determinado.

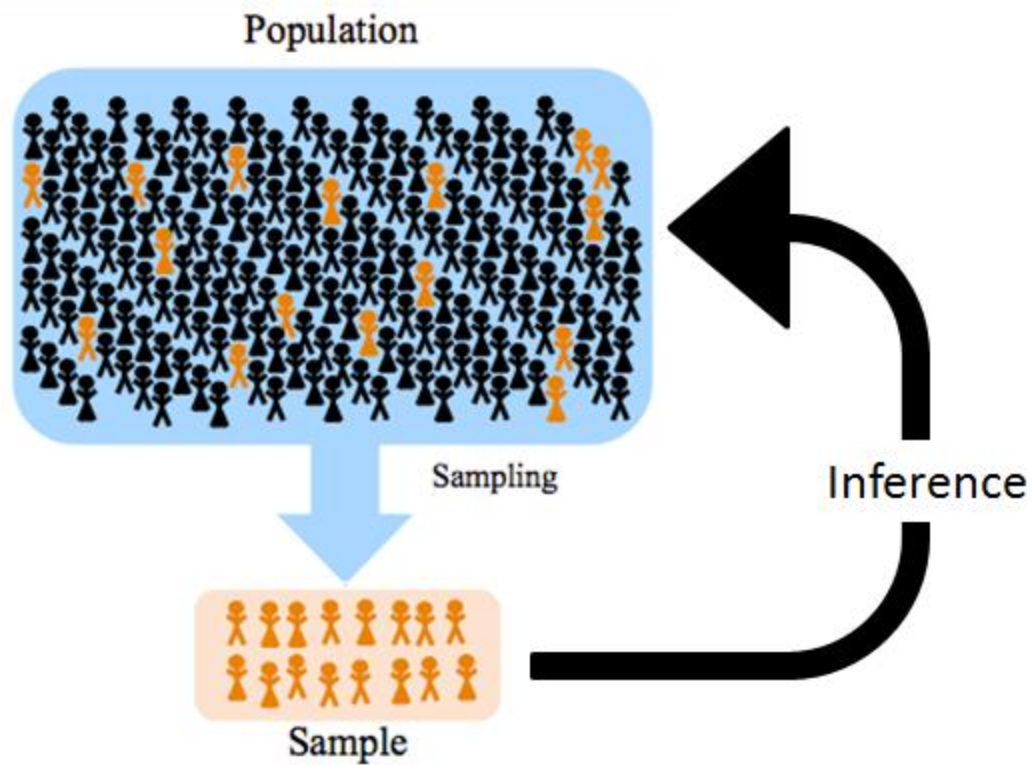


Muestra

Muestra: es un subconjunto de la población, seleccionado de tal forma, que sea representativo de la población en estudio, obteniéndose con el fin de investigar alguna o algunas de las **propiedades** de la población de la cual procede.



¿Por qué se trabaja con la muestra?



¿Por qué se trabaja con la muestra?

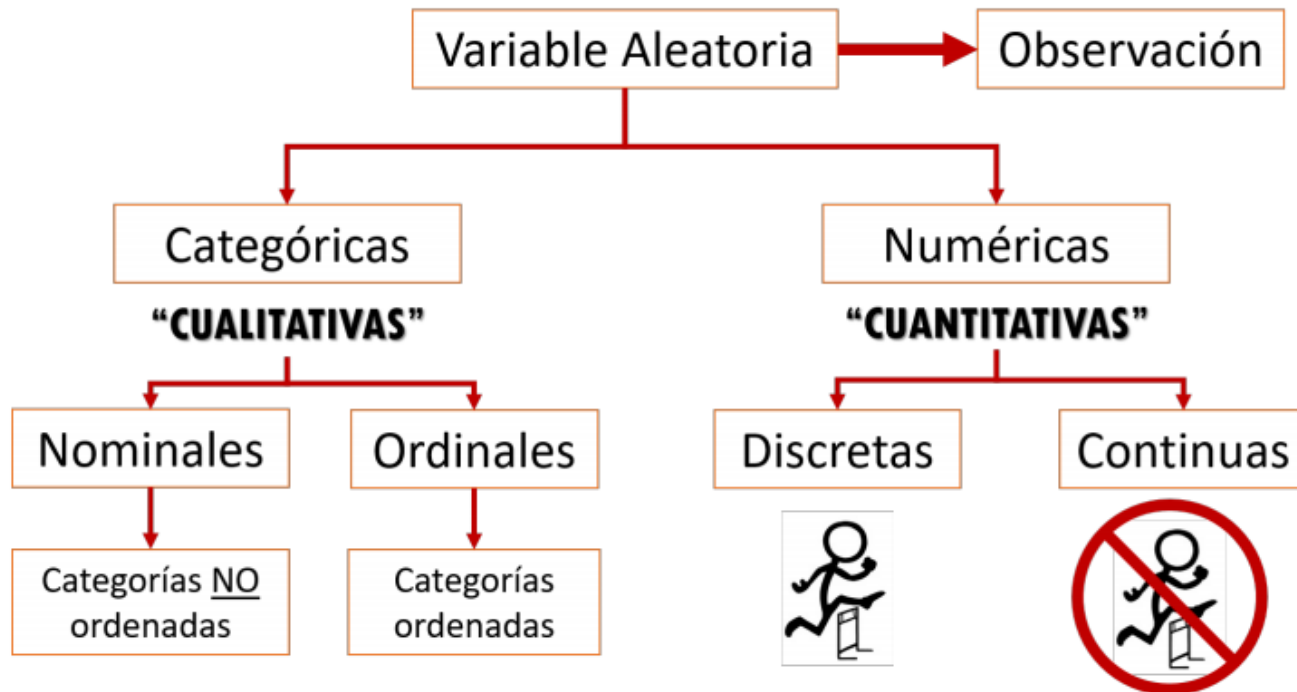
- Medir todas las unidades (árboles, productos, pájaros, etc.) no es práctico, si no imposible.
- El muestreo de unas pocas unidades ahorra dinero.
- El muestreo de unas pocas unidades ahorra tiempo.
- Algunas mediciones son destructivas: talar árboles para inspeccionar patrones de anillos o análisis de tallos capturando vida silvestre para examinar su morfología, etc.

El muestreo hace que los métodos estadísticos sean atractivos y poderosos.

Definiciones

- Una **variable** es una característica o condición que puede cambiar o tomar diferentes valores.
Notación: X, Y, A, B, \dots
- Un **dato** es una observación sobre la variable que se mide.
- Los **datos** son una colección de observaciones.

Tipos de variables



Variables categóricas
Cualitativas

Variable cualitativa Nominal

La variable cualitativa, nominal o categórica son datos que se componen de categorías que no se pueden ordenar por rango; cada categoría es simplemente diferente.

¿Conoce la diferencia entre el machine learning y la inteligencia artificial?

Sí
No

¿Qué aplicación de mensajería prefieres?

Telegram
Whatsapp
Signal



<https://postgrado.ucsp.edu.pe/articulos/machine-learning-inteligencia-artificial-diferencias/>

Nominales:

- Color de la camisa: rojo, verde, negro, blanco, ...
- Sexo del usuario: masculino – femenino
- Profesión: ingeniero, profesor, periodista, ...

Variable cualitativa Ordinal

Es una variable cualitativa que comprende el orden.

How satisfied are you with the level of service you have received? *(please tick)*

Very satisfied

Somewhat satisfied

Neutral

Somewhat dissatisfied

Very dissatisfied

How old are you?

13-19 years

20-30 years

31-50 years

Ordinales:

- Medalla obtenida: oro, plata, bronce.
- Grado escolar: 1º, 2º, 3º, ...
- Frecuencia con la que viaja en vacaciones:

Nunca, Raramente, Ocasionalmente, Frecuentemente, Siempre

Variables Numéricas Cuantitativas

Variables cuantitativas



Variable cuantitativa Discreta

Una variable cuantitativa con valores posibles de solo puntos específicos en una escala se llama variable discreta.

Number of children



$$S=\{0, 1, 2, 3, \dots\}$$

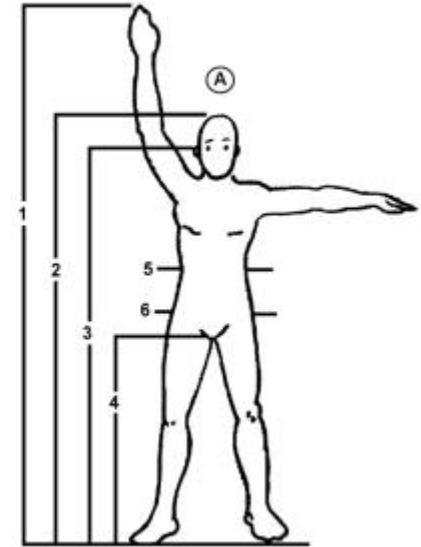
Number of head minus number of tails



$$S=\{-3, -1, 1, 3\}$$

Variable cuantitativa Continua

Esta es una variable en la que la escala es continua y no está formada por pasos discretos.



¿De qué tipo son las variables?

Small



Compact



Sporty



Midsized



Large



Van



¿Cuántos años tiene?

DELL



acer
SAMSUNG

ASUS

lenovo

SONY

COMPAQ

TOSHIBA



FUJITSU

¿Son variables continuas la velocidad y la temperatura?

Depends on instrument



Tarea: ¿De qué tipo son las variables?

Una compañía que brinda servicios informáticos a empresas que deseen tener una presencia firme en la red. Se dedica al tendido de redes LAN (Local área network), instalación de equipos, servidores, etc. Como parte de un estudio, se analiza la información correspondiente a las siguientes variables:

1. Tipos de lenguaje de programación (Python, R, Java, etc..)
2. Cantidad de servidores por empresa
3. Costo de las licencias de software (en pesos colombianos)
4. Versión instalada de Office

¿De qué tipo son las variables?



¿Cómo organizar los datos?



Dataset o conjunto de datos

Nombre	Edad	Sexo	Ingresos	
María	28	F	40.000	Unidad de datos o registro
Carlos	32	M	38.000	Observación numérica del elemento de datos Ingresos
Felipe	41	M	55.000	Observación no numérica, categórica, del elemento de datos Sexo

Dataset

Sex	Age	Smoke	Country	Married
Female	23	Yes	USA	Yes
Male	43	Yes	Colombia	Yes
Male	19	Not	Brazil	Yes
Male	23	Yes	Brazil	Not
Male	54	Not	Spain	Not
Male	76	Yes	Colombia	Not
Female	43	Not	Peru	Yes

7 individuos

Dimensión: 7×5

5 variables

Presentación tabular de la información

Tipos de tablas

- Tablas de frecuencia.
- Tablas de contingencia.
- **Concepto clave:**

Clase

Grupo bajo el cual se clasifican los datos. Puede ser una categoría (cualitativas) o un intervalo (cuantitativas).

Análisis de una variable

Tablas de frecuencia

Tabla de frecuencias

- Cuando se tienen bastantes datos, estos se pueden agrupar en forma de tabla de frecuencia.
- En estas tablas se presentan la frecuencia de los datos.
- Si los datos son categóricos, se agrupan en las categorías definidas.
- Si los datos son cuantitativos, se agrupan por intervalos.

Tabla de frecuencias datos cualitativos y cuantitativos discretos

Valores de La variable X	Frecuencias Absolutas f_i	Frecuencias Relativas f_{Ri}	Frecuencias Porcentajes $f_{Ri}\%$
x_1	f_1		
x_2	f_2		
...	...		
x_k	f_k		
Total	n	1.00	100.00

Tabla de frecuencias datos cuantitativos continuos

Intervalos I_i	Marca de clase	Frecuencias		
		Absolutas f_i	Relativas f_{Ri}	Porcentajes $f_{Ri}\%$
I_1		f_1		
I_2		f_2		
...		...		
I_k		f_k		
Total		n	1.00	100.00

- **Frecuencia absoluta (f_i):** número de datos que pertenece a cada clase.

$$f_i = \# \text{ de datos de la clase } i \text{ (conteo)}$$

- **Frecuencia relativa (f_{Ri}):** fracción o proporción de datos que pertenece a la clase i . Es la frecuencia absoluta dividida el número total de datos (n).

$$f_{Ri} = \frac{f_i}{n}$$

- **Frecuencia relativa porcentual ($f_{Ri} \%$):** porcentaje de datos que pertenece a la clase i . Es la frecuencia relativa multiplicada por 100.

Tabla de frecuencias acumuladas

Valores de La variable X	Frecuencias			Frecuencias Acumuladas		
	Absoluta f_i	Relativa. f_{Ri}	Porcent. $f_{Ri}\%$	Absoluta F_i	Relativa. F_{Ri}	Porcent. $F_{Ri}\%$
	f_1					
	f_2					
	...					
	f_k			n	1.00	100
Total	n	1.0000	100			

Frecuencias acumuladas solo para datos ordinales o cuantitativos

- **Frecuencia acumulada (F_i):** Suma de las frecuencias anteriores hasta la frecuencia i .

$$F_i = \sum_{j=1}^i f_j$$

- **Frecuencia relativa acumulada (F_{Ri}):** Es la frecuencia acumulada dividida el número total de datos (n).

$$F_{Ri} = \frac{F_i}{n}$$

- **Frecuencia relativa porcentual ($f_{Ri} \%$):** porcentaje de datos que pertenece a la clase i . Es la frecuencia relativa multiplicada por 100.

Ejemplo, variable cualitativa

Jorge es un estudiante de ingeniería de sistemas que está desarrollando un proyecto acerca de la conformidad de un programa comercial. Para ello, toma una muestra de 30 empresas que utilizan el programa, a las cuales se les hace una encuesta donde se les pregunta SI están conformes o NO con la aplicación del programa.

Los resultados obtenidos se muestran a continuación.

Si	No	No	No	No	No	No
No	No	No	Si	No	Si	
Si	Si	No	No	No	No	No
No	No	Si	No	Si	No	
No	No	No	No	No	No	Si

Entonces:

- Construir una tabla de frecuencias para esta situación.
- La compañía encargada de evaluar los programas consideran que esta dando malos resultados si hay más del 15 % de empresas inconformes. ¿Qué puede concluir Jorge respecto a dicho programa?

Si No No No No No
No No No Si No Si
Si Si No No No No
No No Si No Si No
No No No No No Si

Clase	f_i	f_{Ri}	$f_{Ri} \%$
Si	8	$\frac{8}{30}$	26.67
No	22	$\frac{22}{30}$	73.33
Total	30	1	100

- De la tabla podemos concluir que 8 de las 30 empresas están conformes con el programa.
- Jorge puede concluir que el programa esta dando malos resultados porque se presentó un 73 % de empresas inconformes en la muestra.