

# Inferencia Estadística

## 2. ESTADÍSTICA DESCRIPTIVA

# Agenda

1. Introducción
2. Tipos de variables
3. Escalas de medición
4. Distribuciones de frecuencias
5. Análisis exploratorio para datos cualitativos
6. Análisis exploratorio para datos cuantitativos
7. Análisis exploratorio bivariado

# Introducción

---

La estadística descriptiva es una parte fundamental del proceso para conocer una población a partir de una muestra (¡representativa!) de la misma.

Permite identificar patrones, encontrar inconsistencias, o datos atípicos.

Útil para plantear el contexto y los supuestos adecuados en los cuales se basará el proceso inferencial (modelación/estimación).

Comúnmente le llamamos también «**análisis exploratorio de los datos**» (EDA por sus siglas en inglés).

# Introducción

---

Para hacerlo de forma apropiada, primero debemos identificar adecuadamente la naturaleza de nuestro conjunto de datos, pues dependiendo del «tipo» de datos se emplearán diferentes estadísticas descriptivas.

Puede ocurrir que el EDA sea suficiente para conocer bien a una población.

Se apoya en recursos gráficos y no gráficos o cuantitativos (EDA vs GEDA).

Dependiendo de los objetivos del análisis se opta por cualquiera de las dos opciones, aunque, generalmente, los recursos gráficos son más útiles para presentar los resultados.

# Introducción

---

Formalmente, un **conjunto de datos** es una **realización** de una variable aleatoria  $X_1, X_2, \dots, X_n$  de tamaño  $n$  de una distribución multivariada; es decir:

$$\mathbf{X}_i^T = (X_{i1}, X_{i2}, \dots, X_{ip}) \quad \text{para } i = 1, 2, \dots, n.$$

Cada  $\mathbf{X}_i$  es un vector aleatorio de dimensión  $p \times 1$ .

Finalmente, el conjunto de datos  $\mathbf{X}$  se puede escribir como

$$\mathbf{X} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix},$$

donde  $\mathbf{X}$  es una matriz aleatoria.

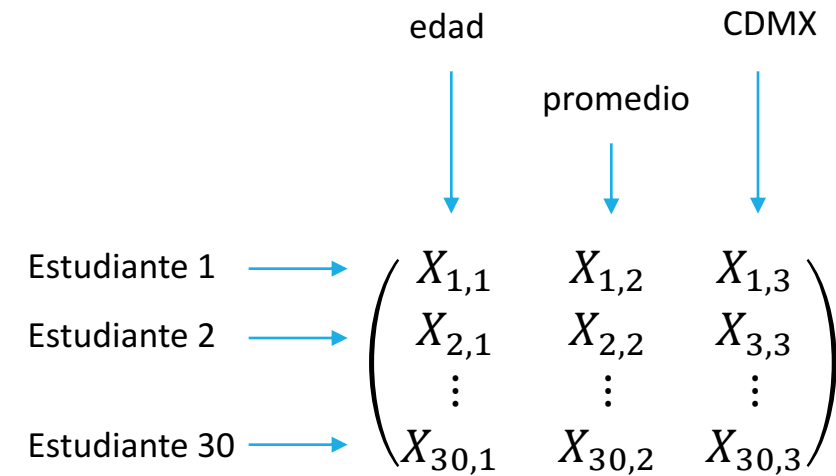
# Introducción

Ejemplo:

En este salón de clases hay 30 estudiantes  
( $n = 30$ )

De cada estudiante se tiene información de  
las siguientes  $p = 3$  variables:

- $X_1$  = edad en años
- $X_2$  = promedio acumulado de la carrera
- $X_3 = \begin{cases} 1 & \text{si el estudiante es de la CDMX} \\ 0 & \text{en otro caso} \end{cases}$



# Introducción

¿Qué ocurre si consideramos ahora la carrera que estudian?

¿Cómo la “convertimos” en un valor numérico?

	estudiante	...	carrera
1	eco		2
2	d.f.		3
3	IN		1
4	II		4
			4
			1
			2
			<u>12</u>
			$\emptyset = 3.1$

estudiante	edad	prom_acum	cdmx	eco	dir_fin	ing_neg	ing_ind
1	22	8	1	1	0	0	0
2	21	8.4	1	0	1	0	0
3	23	7.7	0	0	0	1	0
4	21	7.4	1	0	1	0	0
5	20	7.4	1	0	0	0	1
6	20	8.9	0	0	0	1	0
7	22	8.1	0	1	0	0	0
8	21	8.1	0	1	0	0	0
9	22	8.9	1	0	0	0	1
10	21	8	0	1	0	0	0
11	22	7.4	0	0	0	0	1
12	21	7.8	1	0	0	0	1
13	21	7.6	1	1	0	0	0
14	22	8.3	0	1	0	0	0
15	20	8.2	0	0	1	0	0
16	20	9	1	0	1	0	0
17	22	7.3	1	1	0	0	0
18	23	7.6	1	0	0	1	0
19	22	7.8	1	1	0	0	0
20	23	7.5	0	0	1	0	0
21	23	8.5	1	1	0	0	0
22	20	7.7	0	0	0	1	0
23	20	8.9	1	0	0	0	1
24	20	7.9	1	0	1	0	0
25	21	8.1	0	0	0	0	1
26	23	7.5	1	1	0	0	0
27	21	8.6	1	0	1	0	0
28	21	8	0	0	1	0	0
29	21	9.1	1	1	0	0	0
30	22	8.4	1	1	0	0	0

$$\emptyset = \frac{12}{30} \approx 35\%$$

# Tipos de variables

---

En general, podemos clasificar a los datos de acuerdo al nivel de precisión con el cual estos nos aportan información. Los datos pueden ser de tipo

## Cualitativo

Representan características o atributos. Por ejemplo el color de ojos, el género, nivel socioeconómico.

## Cuantitativo

Se obtiene a partir de un proceso de conteo o bien de un proceso de medición.

Si se obtienen de un proceso de conteo se denominan discretos, si provienen de un proceso de medición se denominan continuos.



# Escalas de medición

Una forma de identificar el nivel de precisión con el cual los datos son medidos es a partir de las escalas de medición. Las escalas nos permiten clasificar a los datos de la siguiente manera:

## Nominal

→ Ej: color de ojos = { azul, verde, café } = { A, V, C }  
 $\Omega = A \cup V \cup C$        $A \cap V = A \cap C = V \cap C = \emptyset$

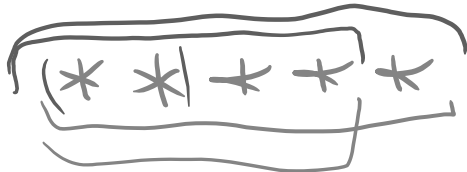
Corresponde a categorías exhaustivas y mutuamente excluyentes que denotan un atributo.

Corresponde al nivel más bajo de medición y no es posible establecer una relación de orden entre las categorías, pero sí de igualdad.

$$\cancel{A < B < V} \quad \underline{A = A}$$

## Ordinal

Los valores de las variables están clasificados en categorías ordenadas de acuerdo con el grado de posesión del atributo. No es posible definir una distancia entre categorías.



# Escalas de medición

## De intervalo

Ej. grados Celsius  $-30^{\circ}\text{C}$ ,  $0^{\circ}\text{C}$ ,  $45^{\circ}\text{C}$

La distancia entre las unidades de medida sí es uniforme, por lo que permite realizar operaciones matemáticas. El cero es arbitrario, no indica la ausencia de atributo.

## De escala o razón

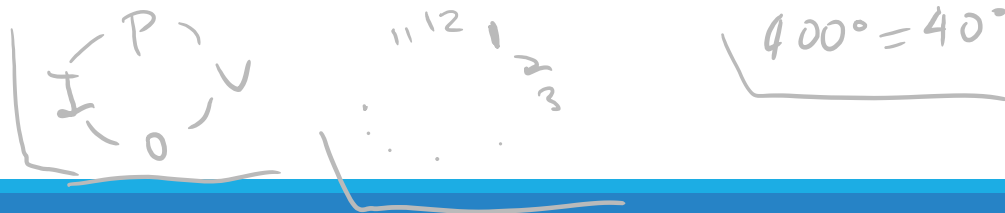
Ej. grados Kelvin  $0^{\circ}\text{K}$ ,  $-3^{\circ}\text{K}$

Similar a la de intervalo, con la única diferencia que el cero en esta escala sí indica la ausencia de atributo, es cero absoluto.

## Cíclica o periódica

Forma alternativa de interpretar las anteriores

Variables ordinales o de intervalo con un valor inicial o mínimo que, después de tomar un valor final o máximo, se reinician en el valor mínimo.



# Escalas de medición

Antes de realizar cualquier tipo de análisis, el primer paso es identificar con qué tipo de información contamos en términos de

1. procedencia,  
escala de medición,
2. temporalidad,
3. periodicidad,
4. nivel de desagregación.

1. P.ej. si proviene de un humano o de instrumento de medición. Nos ayuda a conocer el motivo de posibles errores o inconsistencias.

2. P.ej. temperatura corporal

Una vez que hayamos identificado todas las características inherentes a nuestros datos, podemos proceder al análisis descriptivo de los mismos.

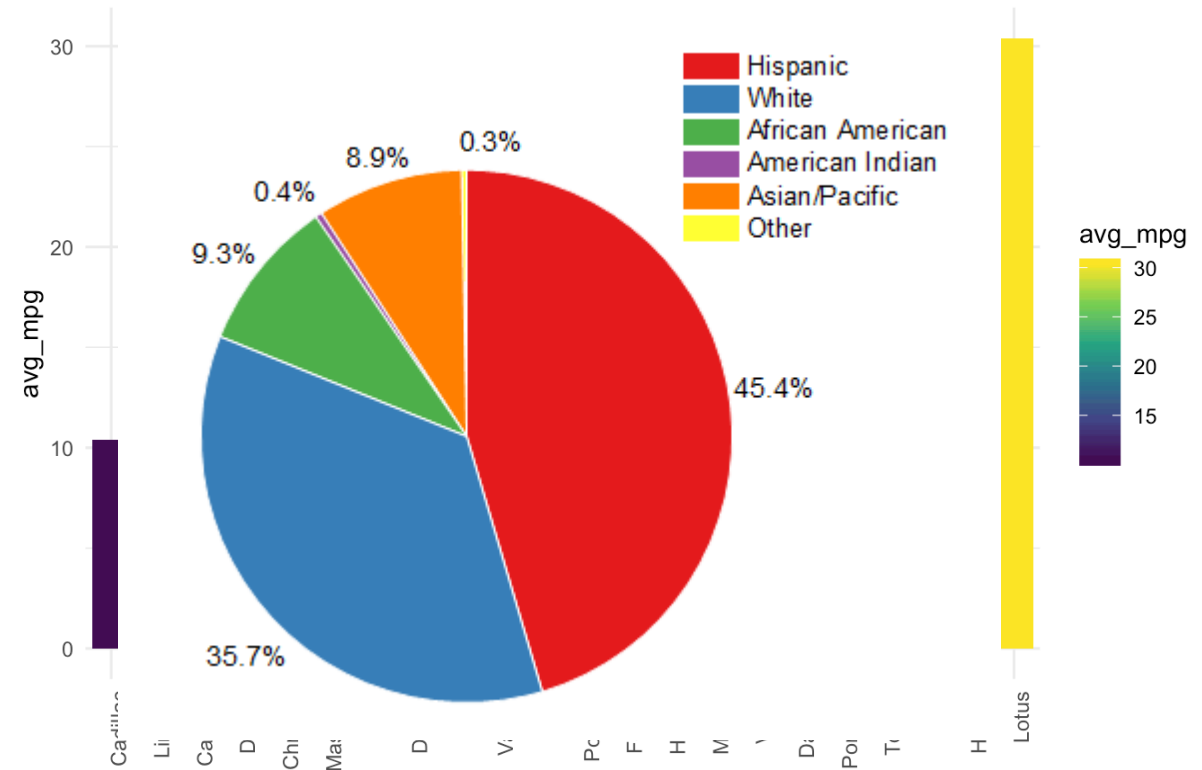
3. P.ej. ventas semanales vs. mensuales vs. anuales

4. SKU (stock keeping unit)

Sabritas Originales / SKU 32 gr  
SKU 78 gr

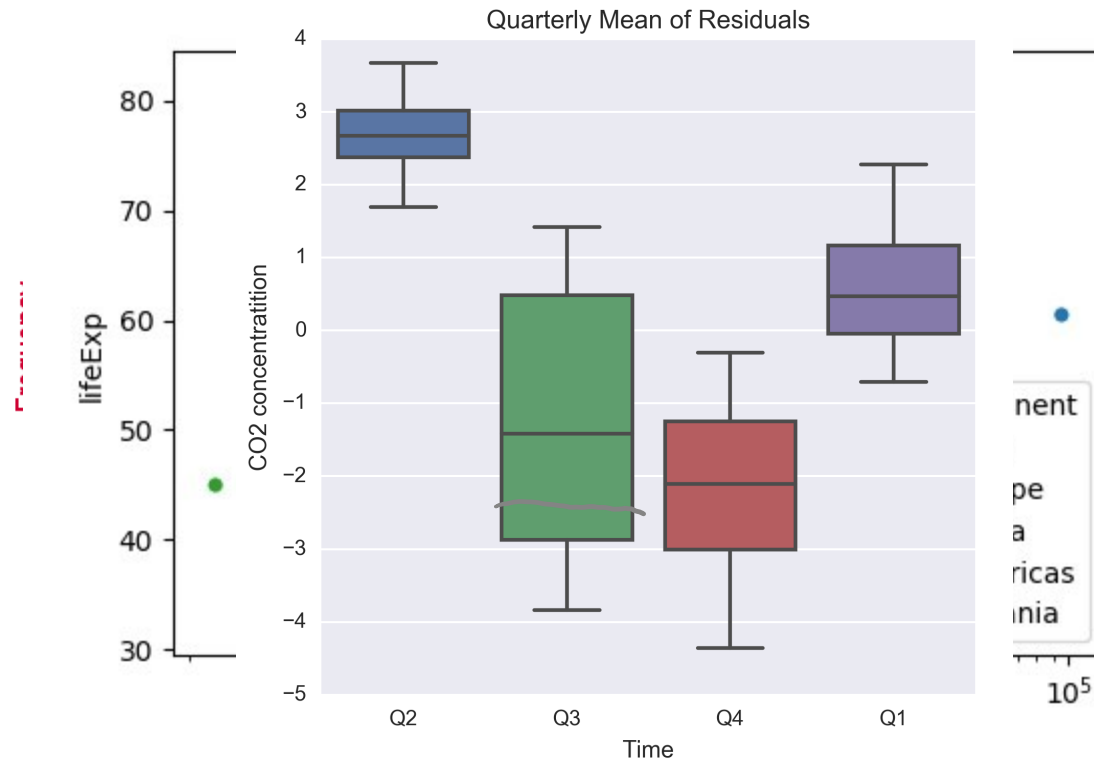
# Análisis exploratorio para datos cualitativos

1. Distribuciones de frecuencias
2. Gráficos de barras
3. Gráficos circulares



# Análisis exploratorio para datos cuantitativos

1. Distribuciones de frecuencias
2. Histogramas
3. Diagramas de dispersión
4. Diagrama de caja y brazos



# Análisis exploratorio bivariado

---

1. Medidas de asociación lineal
2. Correlación de Pearson