

Las medidas de tendencia central son medidas que resumen, en un número, el valor más "típico" en un conjunto de datos "univariado" (una variable a la vez). Las medidas de tendencia central más comunes son:

- **Media.** - Es simplemente el promedio. Es decir, es la razón entre la suma de los valores y el número total de valores:

Población:

$$\mu = \frac{\sum_{i=1}^N x_i}{N}$$

Ideal.
Desconocido.
Imposible
de
Calcular.
Objetivo
Final.

Muestra:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

$\sum_{i=1}^n x_i$?
Notación de Suma

x_i = valores

N = número total de valores.

μ = media ("mu")

$$\bar{x} \approx \mu$$

x_i = valores

n = número total de valores.

\bar{x} = media

Se puede calcular.
Estimado para μ .

- **Mediana.** - Es el punto medio geométrico de un conjunto de datos. Para hallarla, se ordenan los datos en orden creciente (o decreciente) y se busca el valor que esté exactamente en medio. Si se tiene una cantidad impar de datos, la mediana es uno de los datos. Si se tiene una cantidad par, se suman los dos datos de en medio y se divide el valor entre 2. Note que SIEMPRE hay la misma cantidad de valores arriba y abajo de la mediana.

- **Moda.** - Es el valor que más se repite en el conjunto de datos. Note que puede haber más de una moda en un conjunto de datos.

- Ejemplo:

Suponga que tenemos los siguientes puntajes en partidas de un videojuego:

100, 130, 140, 160, 170

a) Partida	Puntaje
1	100
2	170
3	130
4	140
5	140

\Rightarrow Media = 140
Mediana = 140
Moda = TODO = 100, 170, 130, ...

$\frac{100+170+130+140+140}{5}$

3	130
4	140
5	160

Moda = $\text{TODO} = 100, 120, 130, \dots$

b) Partida Puntaje

1	110
2	160
3	150
4	110

\Rightarrow Media = 132.5
Mediana = 130
Moda = 110

- Ejercicio.- Hallar la media, mediana y moda de los siguientes datos **Cen clase: Google. Ver más abajo fórmulas de Google).**

41, 39, 52, 17, 23, 48, 39, 29, 27, 56, 23, 31, 35, 41.

¿Cuál de las 3 medidas es mejor? Todo depende del problema, pero por ejemplo si el conjunto presenta un **outlier** ("valor fuera de serie") y contamos con pocos datos, la mediana es, en general, más confiable. Si no hay outliers y tenemos muchos datos, en general, la media es preferible.

- Ejercicio.- Suponga que se analizan los ingresos mensuales de 5 familias:

Familia Ingreso mensual

1	60,000
2	50,000
3	80,000
4	70,000
5	10,000,000 (outlier)

Halle la media y mediana y analice cuál de dichos valores representa mejor al conjunto de datos.

A menudo es necesario hacer algunos ajustes a los datos, como por ejemplo, cambiar de unidades. Surge entonces la pregunta, ¿cómo cambian las medidas de tendencia central al realizar estos cambios?

- Sumar una constante a todos los datos: Esto afecta a las medidas de tendencia central sumándoles la misma constante (demostración en clase).
- Multiplicar todos los datos por una constante: De la misma manera, esto afecta a las medidas de tendencia central multiplicándolas por la misma constante (demostración en clase).
- Ejercicio.- Hallar la media, mediana y moda de 0, 2, 4, 4, 5 y 6

$$\frac{x_1+k + x_2+k + \dots + x_n+k}{n} = \frac{x_1+x_2+\dots+x_n + nk}{n}$$

$$= \boxed{\frac{x_1+x_2+\dots+x_n}{n}} + \cancel{\frac{nk}{n}}$$

$$= \underline{\bar{x} + k}$$