



# Clase 6 : Medidas de tendencia Central

# ¿Qué son las medidas de tendencia central?

Las medidas de tendencia central son valores que representan el centro de un conjunto de datos. Las medidas más comunes son la **media**, la **mediana** y la **moda**.



# Media aritmética

La media aritmética es la suma de todos los valores de un conjunto de datos dividida por el número de valores. Es sensible a los **valores atípicos/Extremos/outliers**.



$$\bar{X} = \frac{x_1 + x_2 + \dots + x_N}{N}$$

**X**  
Variable

**x<sub>1</sub>, x<sub>2</sub>, ..., x<sub>N</sub>**  
Datos

Se llama Media aritmética de un lote de datos al promedio de los valores del lote:

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{\sum_{i=1}^N x_i}{N}$$



# Mediana

La mediana es el valor que ocupa la posición central en un conjunto de datos ordenado. Es menos sensible a los **valores atípicos** que la media aritmética.

Se denomina **Mediana (  $M_e$  )** al número real tal que a lo sumo el 50% de los datos son menores que él y a lo sumo el 50% son mayores.

# Mediana



Si tenemos conjunto de datos con un número impar de elementos, entonces la mediana es simplemente el valor central

Si tenemos un conjunto de datos con un número par de elementos, la mediana se calcula tomando la media aritmética de los dos valores centrales después de ordenar el conjunto.

Sea el lote de datos:  $X_1, X_2, \dots, X_N$ . Para indicar el orden en el lote usaremos la notación:

$$X_{(1)}, X_{(2)}, \dots, X_{(N)}$$

N impar:	N par:
$M_e = X_{\left(\frac{N+1}{2}\right)}$	$M_e = \frac{X_{\left(\frac{N}{2}\right)} + X_{\left(\frac{N}{2}+1\right)}}{2}$

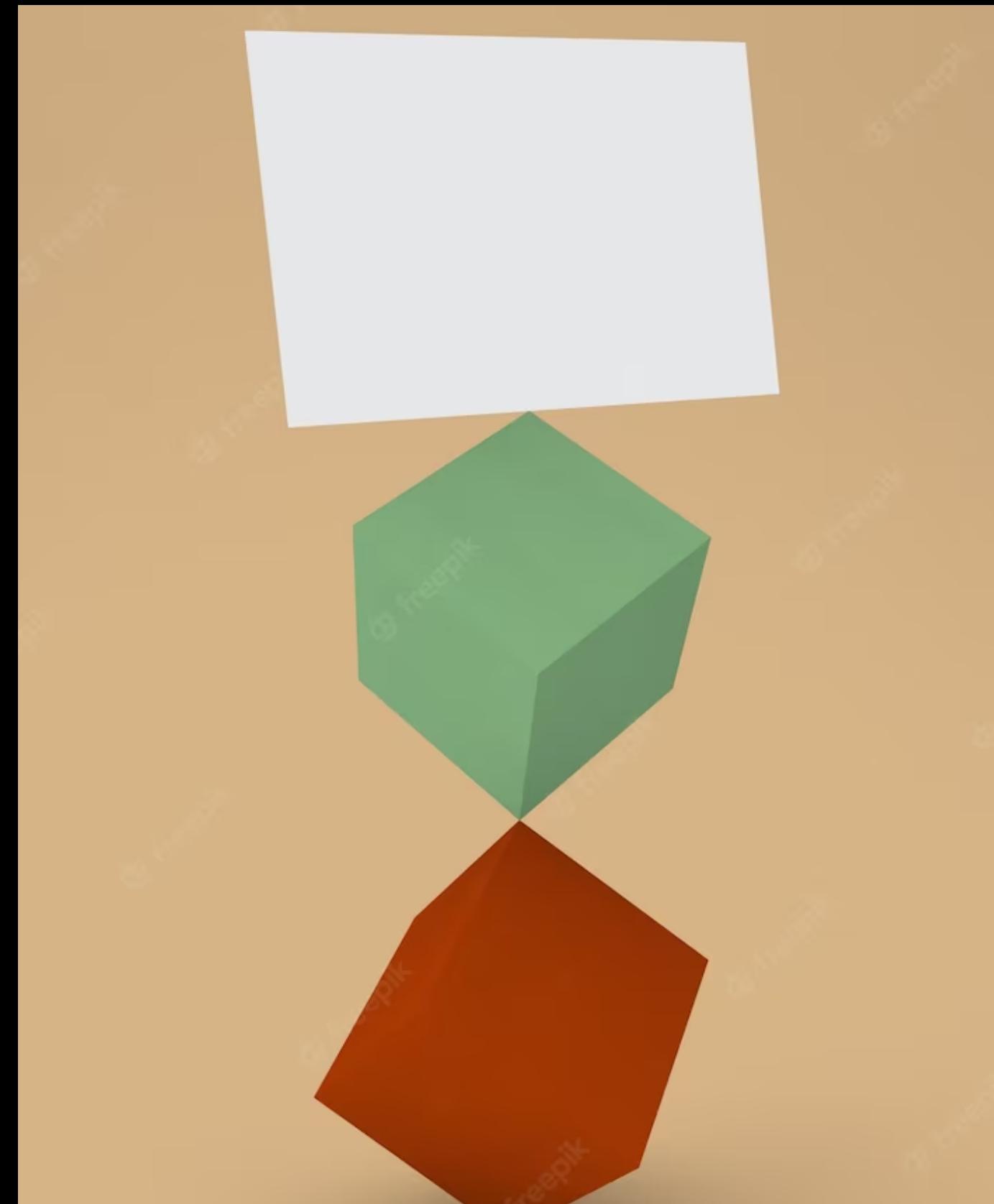


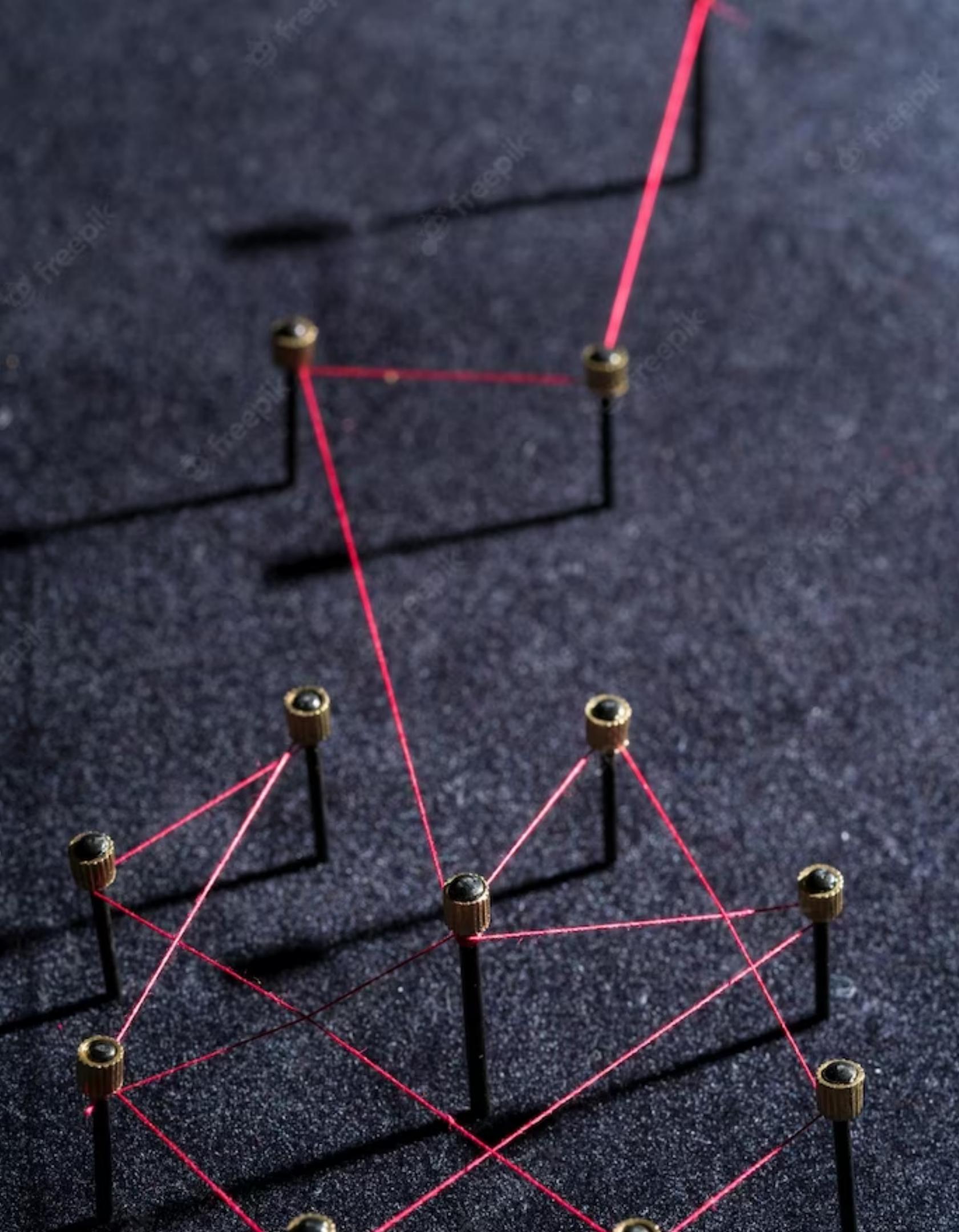
# Moda

La moda es el valor que se repite con más frecuencia en un conjunto de datos. Puede haber más de una moda o ninguna. Es útil para datos **categóricos**.

## Usos de las medidas de tendencia central

Las medidas de tendencia central son útiles para resumir y describir un conjunto de datos. También se pueden utilizar para **comparar** conjuntos de datos.





# Medidas de tendencia central

Las medidas de tendencia central, como la media y la mediana, nos permiten entender el centro de la distribución de los datos. Sin embargo, estas medidas no nos dicen nada sobre la variabilidad de los datos. Para ello, necesitamos medidas de dispersión.



# Explorando la variabilidad



# Introducción

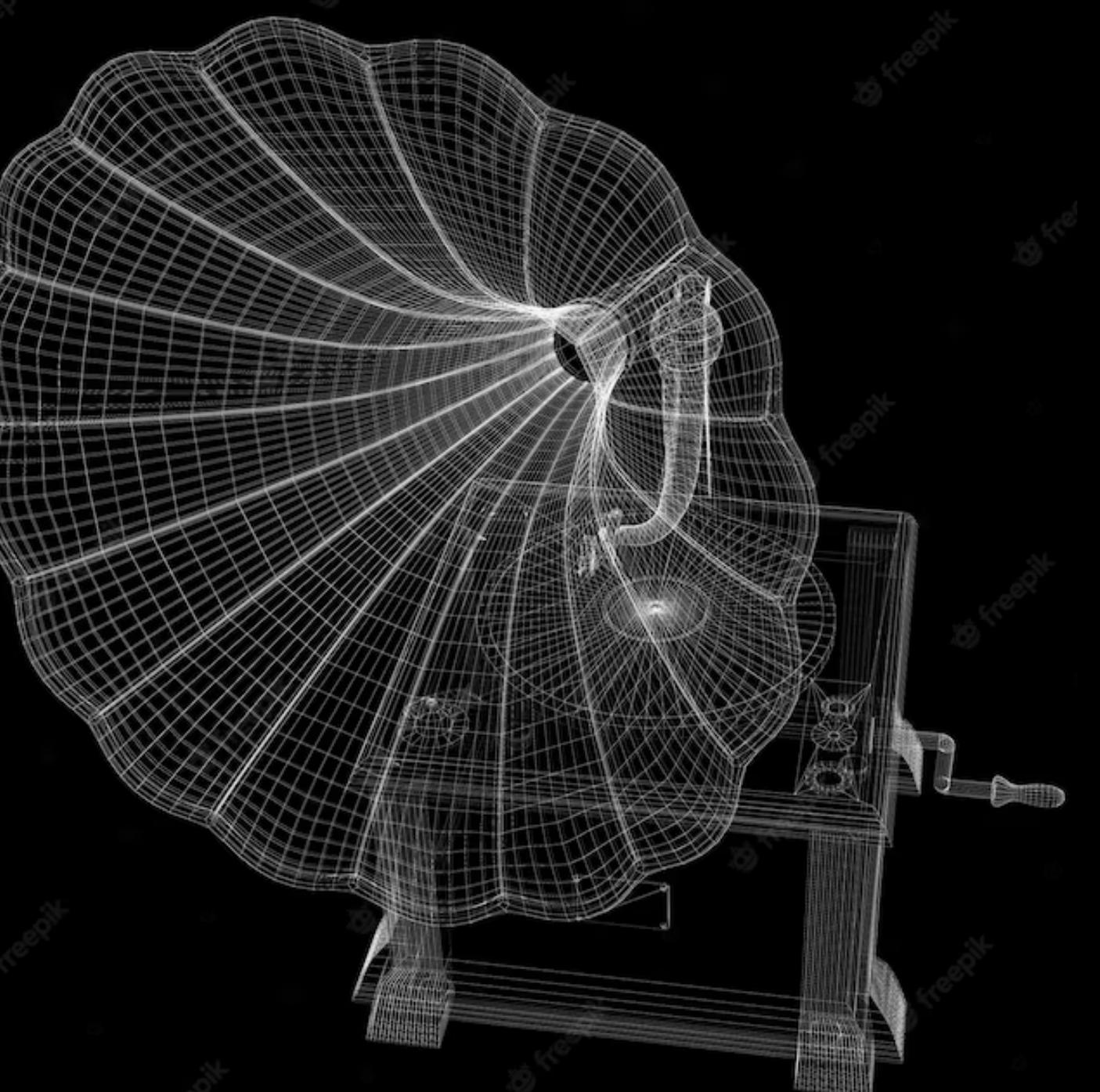
En el análisis de datos, la variabilidad es una medida fundamental que nos permite entender la distribución de los datos y tomar decisiones basadas en éstos. Exploraremos las distintas medidas de variabilidad

# Varianza

La **varianza** es una medida de dispersión que indica qué tan lejos están los valores del conjunto de datos de su media. Es la media de los cuadrados de las desviaciones de los valores respecto a la media. Una varianza alta indica que los valores están muy dispersos.

$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}$ . Para su cálculo suele utilizarse la siguiente fórmula equivalente:  $\sigma^2 = \frac{\sum_{i=1}^N x_i^2}{N} - \bar{x}^2$ . En el caso que los datos estén presentados en una tabla de frecuencias, la fórmula más adecuada para el cálculo de la varianza es:

$\sigma^2 = \frac{\sum_{j=1}^m f_j (x_j - \bar{x})^2}{N}$ , donde  $m$  es la cantidad de datos diferentes (si es una distribución discreta) o la cantidad de intervalos (si es una distribución en intervalos de clase, en cuyo caso se usa la marca de clase).



# Desviación estándar

La desviación estándar es una medida de dispersión que nos indica cuánto se alejan los datos de la media. Una desviación estándar alta indica que los datos están muy dispersos, mientras que una desviación estándar baja indica que los datos están muy juntos.

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}}$$



# Rango intercuartil

El rango intercuartil es otra medida de dispersión que se calcula como la diferencia entre el tercer y primer cuartil. Esta medida es útil para detectar valores atípicos en la distribución de los datos.

