**TIPOS DE DATOS:**

Categóricos: Genero, categoría de película, método de pago

* Ordinal -> Existe un orden entre categorías
* Nominal -> No existe un orden entre categorías

Numéricos: Edad, altura, temperatura

* Discretos -> Números enteros
* Continuos -> Números decimales

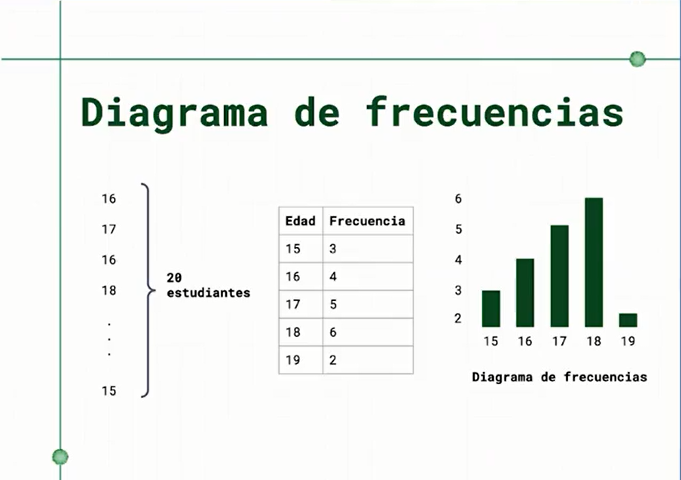
**MEDIDAS DE TENDENCIA CENTRAL**

Media -> Promedio

Mediana -> Dato central

Moda -> Dato que más se repite

**TABLA DE FRECUENCIAS (MODA)**



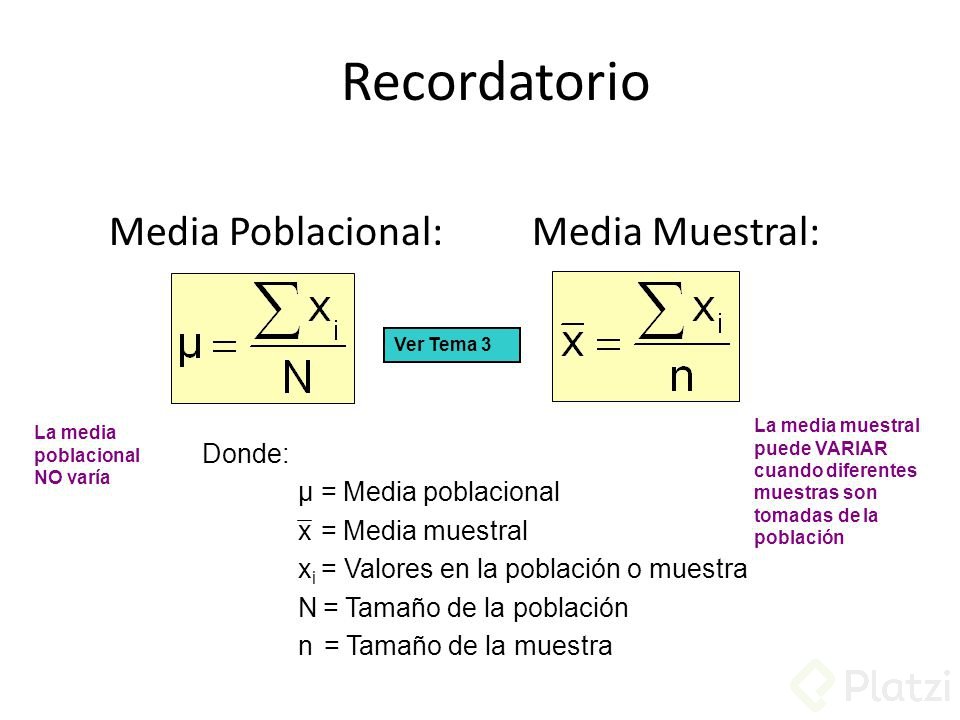
Cuando usar cual

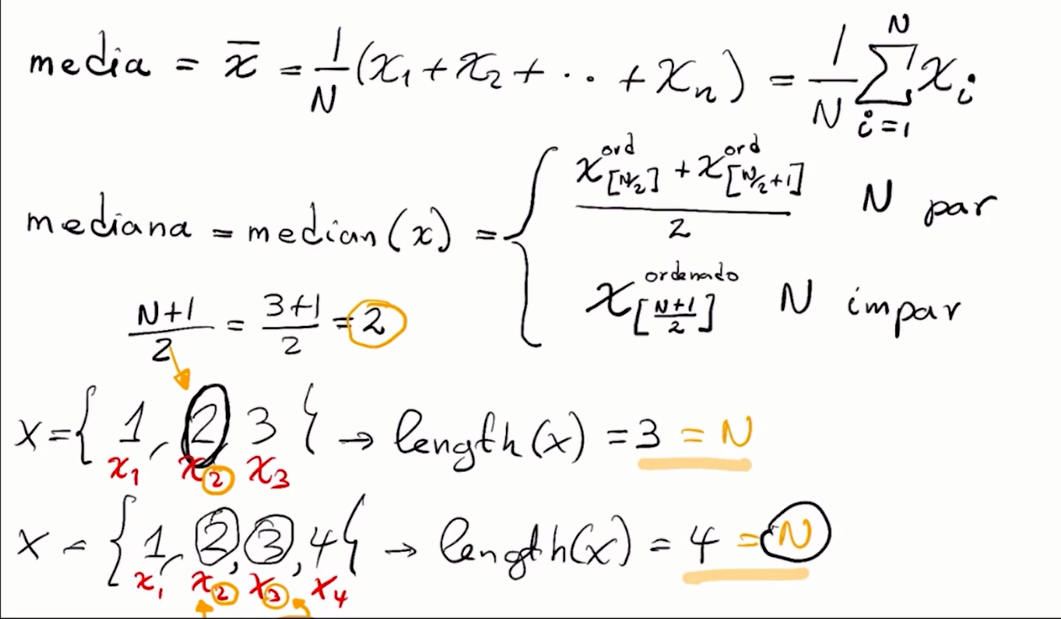
* La media es susceptible a valores atípicos
* La moda no aplica para datos numéricos continuos

Como dato curioso e interesante:

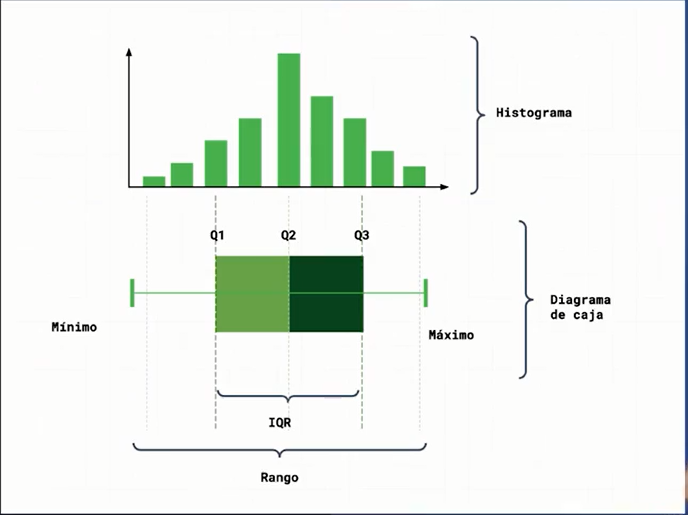
No solo existen esas 3 medidas de tendencias central, hay otras más poco conocidas, pero las abordaremos ahora mismo:

* Media ponderada**:** es una medida de tendencia central, que es apropiada cuando en un conjunto de datos cada uno de ellos tiene una importancia relativa (o peso) respecto de los demás datos. Se obtiene multiplicando cada uno de los datos por su ponderación (peso) para luego sumarlos, obteniendo así una suma ponderada; después se divide esta entre la suma de los pesos, dando como resultado la media ponderada.
* Media armónica**:** La media armónica es igual al número de elementos de un grupo de cifras entre la suma de los inversos de cada una de estas cifras.  
  En otras palabras, la media armónica es una medida estadística recíproca a la media aritmética, que es la suma de un conjunto de valores entre el número de observaciones.
* Media geométrica**:** es una cantidad arbitraria de números (por decir n números) es la raíz n-ésima del producto de todos los números; es recomendada para datos de progresión geométrica, para promediar razones, interés compuesto y números índice.



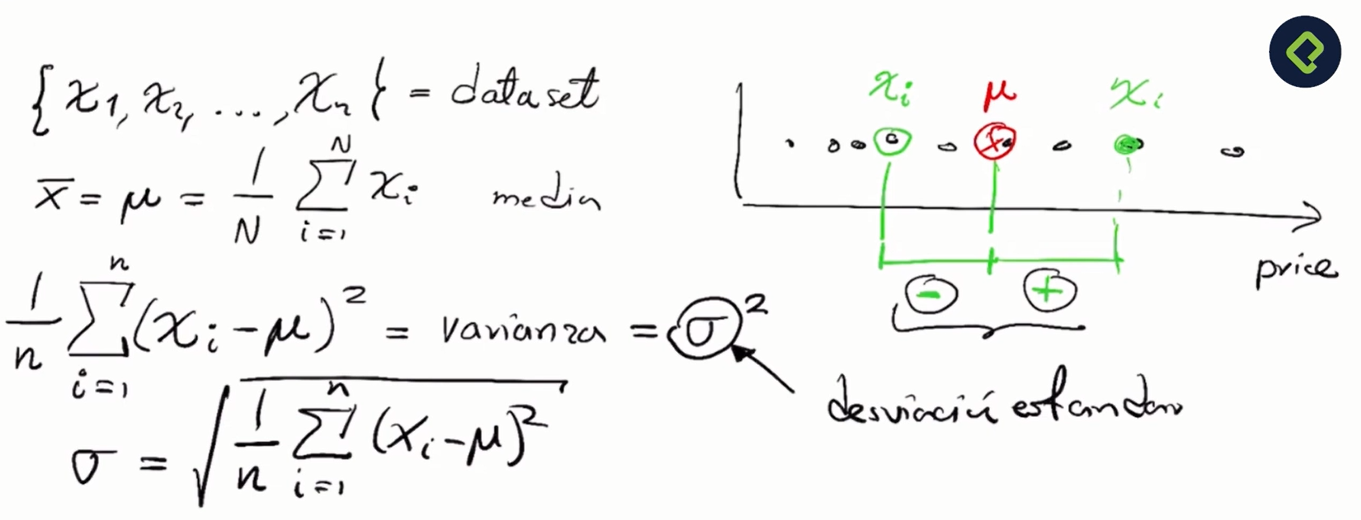


**MEDIDAS DE DISPERSIÓN**

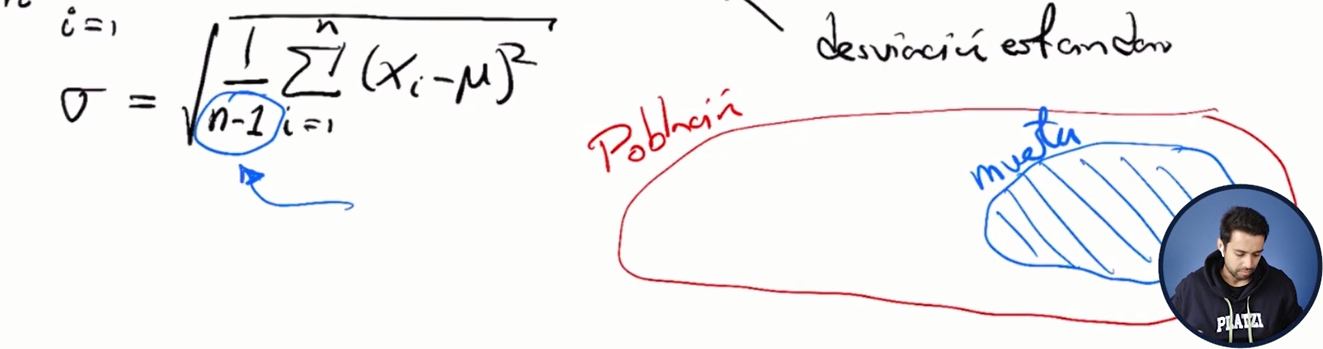


* **Rango:** El Rango es el intervalo entre el valor máximo y el valor mínimo.
* **Cuartiles:** Los cuartiles son valores que dividen una muestra de datos en cuatro partes iguales.
  + **1er cuartil (Q1):** 25% de los datos es menor que o igual a este valor.
  + **2do cuartil (Q2):** La mediana. 50% de los datos es menor que o igual a este valor.
  + **3er cuartil (Q3):** 75% de los datos es menor que o igual a este valor.
  + **Rango intercuartil:** La distancia entre el primer 1er cuartil y el 3er cuartil (Q3-Q1); de esta manera, abarca el 50% central de los datos.
* **Diagrama de caja o box plot:** representa gráficamente una serie de datos numéricos a través de sus cuartiles. De esta manera, el diagrama de caja muestra a simple vista la mediana y los cuartiles de los datos. También puede representar los valores atípicos de estos.

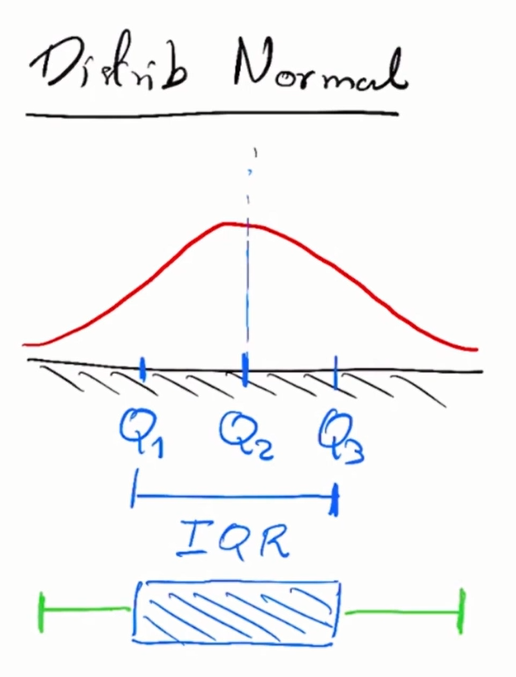
**DESVIACION ESTANDAR**

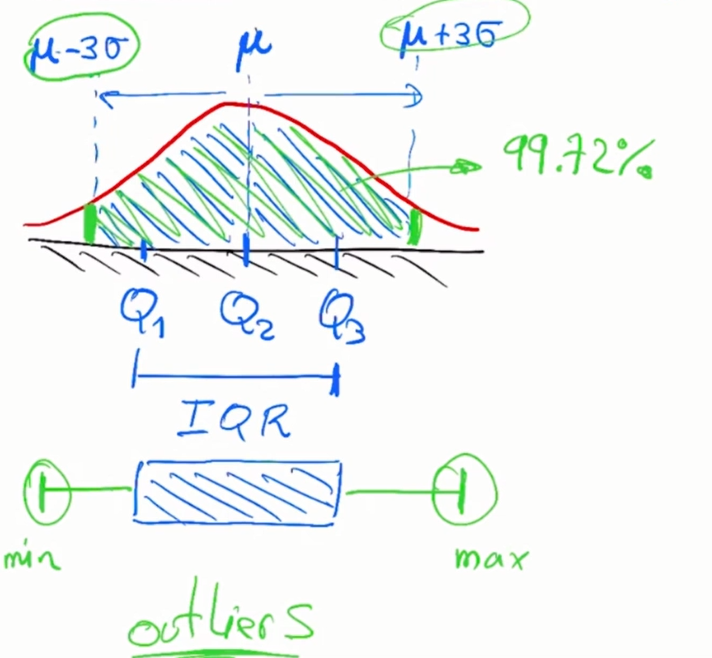


Cuando se toma toda la población va la formula de arriba, cuando se trabaja solo sobre una muestra, va la formula de abajo



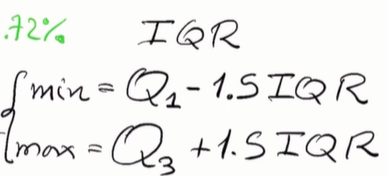
CAMPANA DE GAUS

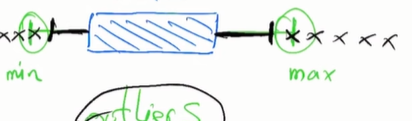




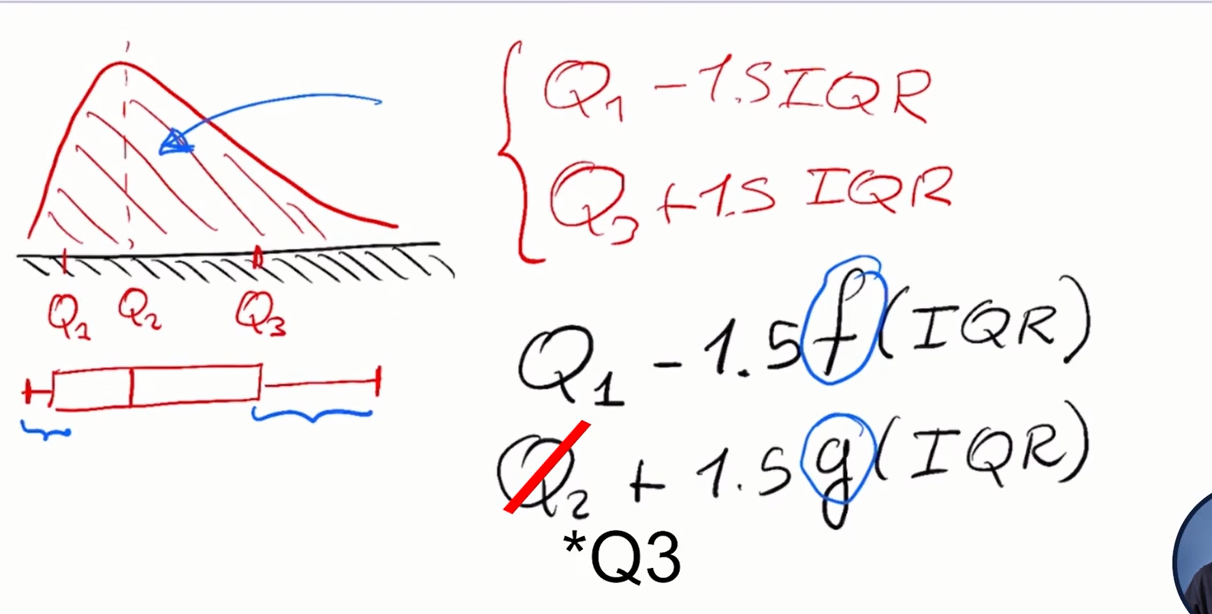
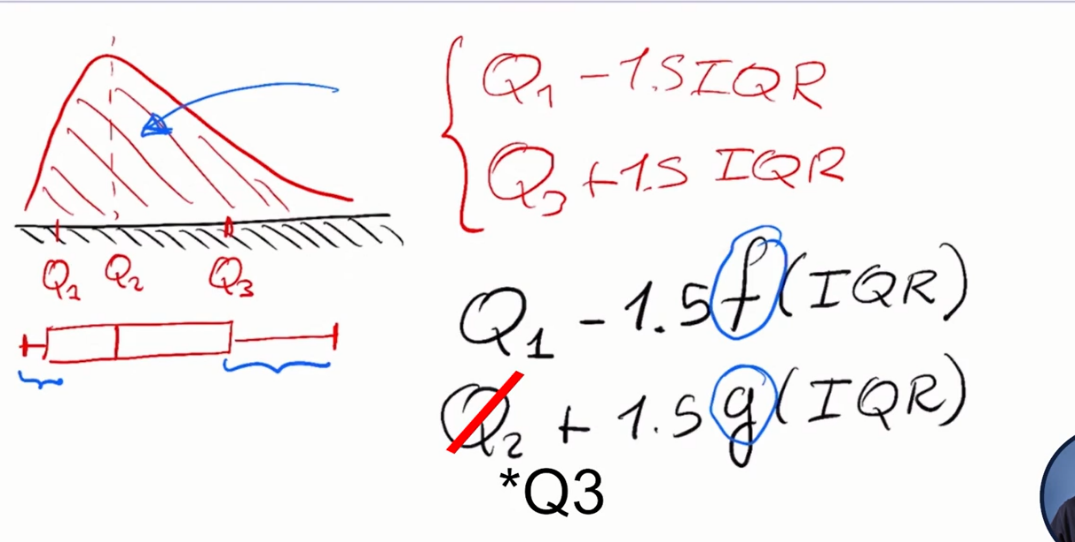
Los outliers son los casos atípicos/anomalos en la muestra (muy por encima o muy por debajo), miú significa mediana y la o signigica desviación estándar

**FORMULA PARA QUITAR DATOS ANOMALOS**





**DISTRIBUCION NO SIMETRICA**



**ASIMETRÍA EN DISTRIBUCIONES**

Como vimos en la clase, el hecho de que nuestra distribución tenga una tendencia a la derecha o a izquierda nos representa un problema, ya que no a acorde con una distribución y eso puede afectar a nuestros análisis si no tomamos en cuenta ese sesgo. No siempre hay que confiar en nuestra intuición o lo que vemos a simple vista, hay métodos como:

* Primer coeficiente de asimetría de Pearson (asimetría de modo)
* Segundo coeficiente de asimetría de Pearson (asimetría mediana)
* Coeficiente de Groeneveld y Meeden
* Coeficiente de Fisher

Por mencionar algunos.  
.

**Y por último, no hay que olvidar la curtosis:**

Una curtosis grande implica una mayor concentración de valores de la variable tanto muy cerca de la media de la distribución (pico) como muy lejos de ella (colas), al tiempo que existe una relativamente menor frecuencia de valores intermedios. Esto explica una forma de la distribución de frecuencias/probabilidad con colas más gruesas, con un centro más apuntado y una menor proporción de valores intermedios entre el pico y colas.  
Una mayor curtosis no implica una mayor varianza, ni viceversa.