# TEMA 4

# PARÁMETROS Y ESTIMADORES

Medidas de posición y de dispersión.
Cálculo, características y propiedades.

# Población y Muestra

<ins>Población:</ins> Es un conjunto de elementos acotados en un tiempo y un espacio determinados con alguna característica común observable o medible.

<ins>Tamaño poblacional:</ins> número de elementos de la población y lo denotaremos con **N**.

<ins>Muestra</ins>: Es un subconjunto de elementos de la población, representativo del conjunto total.

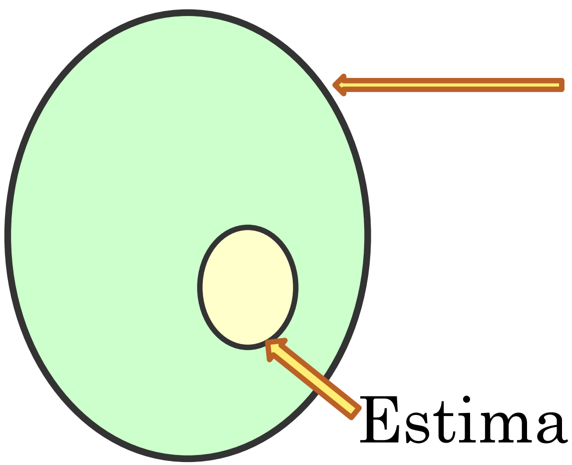
<ins>Tamaño muestral</ins>: número de elementos de la muestra y lo denotaremos con **n.**

# PARÁMETROS Y ESTIMADORES

Un **parámetro** es un número que describe la **población,** cuando se trabajo con el conjunto de datos de la población completa.

Cuando el conjunto de datos proviene de una muestra, el número obtenido es el **estimador** o **estadístico** que se utiliza como **una estimación del parámetro.**

**La diferencia** entre el parámetro y el estadístico es el **error de estimación.**



Parámetros: **N; μ; σ; σ²; π**

Estimadores o Estadísticos **n; $\overline{x}$; s; s²; p**

# Veamos un ejemplo

Tenemos valores de puntajes al arrojar un dado,

x: 1, 2, 3, 4, 5, 6. N=6 (esto sería la población) cuyo promedio es $\boldsymbol{\mu} = 3{,}5$

Luego se arroja un dado 4 veces, n=4 (muestra1) x: 3, 5, 2, 2
$\bar{x}1 = 3$

Si volvemos a arrojar otra vez 4 veces (muestra2), x: 3, 4, 4, 5
$\bar{x}2 = 4$

Entonces $\bar{x}1$ y $\bar{x}2$ son estimadores del parámetro $\mu=3{,}5$

# MEDIDAS DE POSICIÓN Y DE DISPERSIÓN

Existen muchas medidas que resumen la información, en este curso veremos las siguientes:

**Medidas de Posición**

- Media aritmética o promedio
- Mediana
- Moda o modo
- Cuartil
- Decil
- Percentil

**Medidas de Dispersión**

- Rango
- Variancia
- Desvío estándar
- Coeficiente de variación

Nota: Algunos autores denominan a la Media aritmética, una Medida de Tendencia Central.

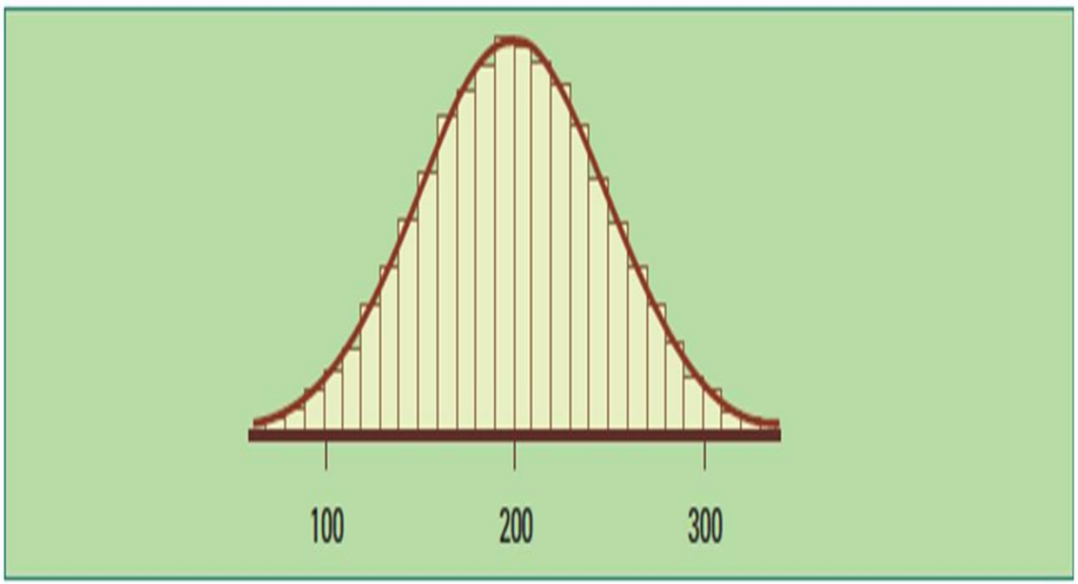
# MEDIDAS DE POSICIÓN

Las medidas de posición son funciones de los datos de una población o muestra que miden su distribución de frecuencias.

Recordemos que el Histograma es un gráfico que representa las frecuencias con la que se presentan los valores de una variable. Dichos valores se ubican en el eje “x”.

A partir del histograma se puede superponer una función con forma de campana denominada Distribución Normal.

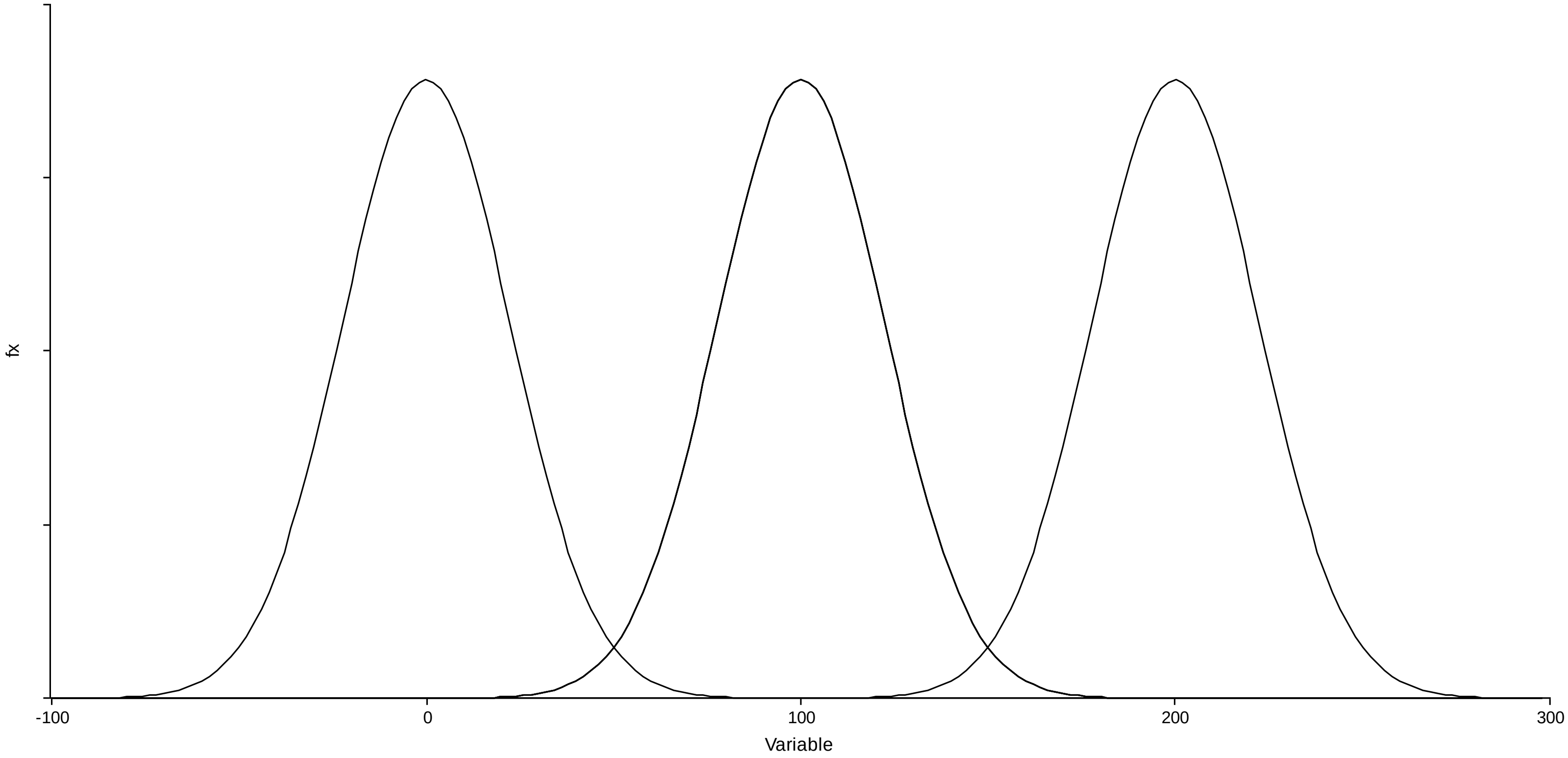
Distribución Normal




En muchas situaciones el valor central es el que tiene la frecuencia mas alta.

# Seguimos con Medidas de posición

Supongamos que tenemos 3 funciones que representan a tres poblaciones de datos, cada una con un rango de variación distinto y con un valor central diferente.




El valor central de cada población informa sobre la POSICION que ocupa un conjunto de datos en el campo de variación de la variable (rango de variación).

# Media aritmética

Si x1, x2,….., xn constituyen una muestra aleatoria de tamaño *n*, luego la **media muestral o promedio** se define como:

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

EJEMPLO: si tenemos precios de un producto en pesos, en 60 puntos de venta.

| | | | | |
|---|---|---|---|---|
| 1290,8 | 928,4 | 958,1 | 1040,2 | 1138,5 |
| 1133,5 | 1172,5 | 1331,2 | 1159,3 | 1065,6 |
| 1027,2 | 936,7 | 1049,4 | 977,4 | 1062,6 |
| 953,6 | 861,9 | 1048,6 | 1220,4 | 929,8 |
| 1013,5 | 940,7 | 1398,7 | 780,7 | 823,9 |
| 1088,4 | 819,5 | 938,6 | 1171,8 | 1006,7 |
| 910,8 | 1030,3 | 867,3 | 1100,9 | 1291,6 |
| 1175,4 | 909,9 | 1016,2 | 872,4 | 962,5 |
| 1083,7 | 1017,4 | 777,8 | 1122,3 | 1080,4 |
| 1070,6 | 1026,5 | 1119,2 | 997,5 | 1205,4 |
| 853,8 | 1095,2 | 1087,3 | 1063,4 | 1076,2 |
| 1160,1 | 1190,4 | 947,5 | 704,2 | 886,6 |

Media ($\bar{X}$)= $(\sum_{i=1}^{60} x_i)/60$=
( 1290.8+1133.5+1027.2+…+886.6)/60
=1032.85

El precio promedio del producto
es $\bar{X}$=1032,85

La Media también se la conoce como Esperanza Matemática

# Mediana

Es la magnitud del término de la serie que divide a la misma en dos grupos del mismo número de observaciones.

## n impar

$$Md = X_{\frac{n+1}{2}}$$

**La mediana deja la misma cantidad de datos a cada lado.**

Para hallar la mediana del conjunto de datos (4, 15, 8, 31, 17)

- Primero los ordenamos de menor a mayor (4, 8, 15, 17, 31).
- Luego, la mediana es el valor central (15).

## n par

$$Md = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}$$

4 8 15 17 promediamos el 8 y el 15 $\frac{8+15}{2} = 11{,}6$

# Mediana

En el ejemplo de precios de un producto en diferentes puntos de venta, n=60 es par, entonces, la mediana es el promedio de los 2 valores centrales,

$$Md = \frac{X_{30} + X_{31}}{2} = \frac{1030{,}3 + 1040{,}2}{2} = 1035{,}25$$

## Para datos agrupados (en tablas de Frecuencia)

$$Md = L + \frac{\frac{n}{2} - F_{i-1}}{f_i} . a$$

Donde:

L = límite superior de la clase anterior a la que presenta la primera frecuencia acumulada superior al 50% de las observaciones.

n = número de observaciones.

$F_{i-1}$ = frecuencia acumulada hasta la clase cuyo límite superior es L.

$f_i$ = frecuencia absoluta de la primer clase cuya frecuencia acumulada supera el 50% de las observaciones.

a = amplitud del intervalo de clase.

| Intervalos de Clase | Clases (x) | Frecuencias Absolutas (fi) | Frecuencias Relativas (fx) | Frecuencias Acumuladas Absolutas (Fi) | Frecuencias Acumuladas Relativas (Fx) |
|---|---|---|---|---|---|
| 700,00 – 799,99 | 750 | 3 | 0,05 | 3 | 0,05 |
| 800,00 – 899,99 | 850 | 7 | 0,12 | 10 | 0,17 |
| 900,00 – 999,99 | 950 | 13 | 0,22 | 23 | 0,38 |
| 1000,00 – 1099,99 | 1050 | 20 | 0,33 | 43 | 0,72 |
| 1100,00 – 1199,99 | 1150 | 11 | 0,18 | 54 | 0,9 |
| 1200,00 – 1299,99 | 1250 | 4 | 0,07 | 58 | 0,97 |
| 1300,00 – 1399,99 | 1350 | 2 | 0,03 | 60 | 1 |
| Total | | 60 | 1,00 | | |

$$Md = L + \frac{\frac{n}{2} - F_{i-1}}{f_i}.a$$

$$Md = 999,99 + \frac{\frac{60}{2} - 23}{20}.100 = 1034,99$$

# Moda o Modo

Es el valor de la variable que tiene mayor frecuencia, generalmente se calcula para tablas de frecuencia:

$$Mo = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} a$$

$$Mo = 1000 + \frac{7}{7+9} * 100 = 1043,75$$

| Intervalos de Clase | Clases (x) | Frecuencias Absolutas (fi) |
|---|---|---|
| 700,00 – 799,99 | 750 | 3 |
| 800,00 – 899,99 | 850 | 7 |
| 900,00 – 999,99 | 950 | 13 |
| 1000,00 – 1099,99 | 1050 | 20 |
| 1100,00 – 1199,99 | 1150 | 11 |
| 1200,00 – 1299,99 | 1250 | 4 |
| 1300,00 – 1399,99 | 1350 | 2 |
| Total | | 60 |

# Moda o Modo cuando la variable es cualitativa

*Ejemplo de una encuesta a norteamericanos sobre sus preferencias de auto.*

| Origen | FA | FR |
|---|---|---|
| American | 128 | 0,38 |
| Europeo | 45 | 0,13 |
| Japones | 166 | 0,49 |
| Total | 339 | 1 |

Variable: Origen
Moda: Japones

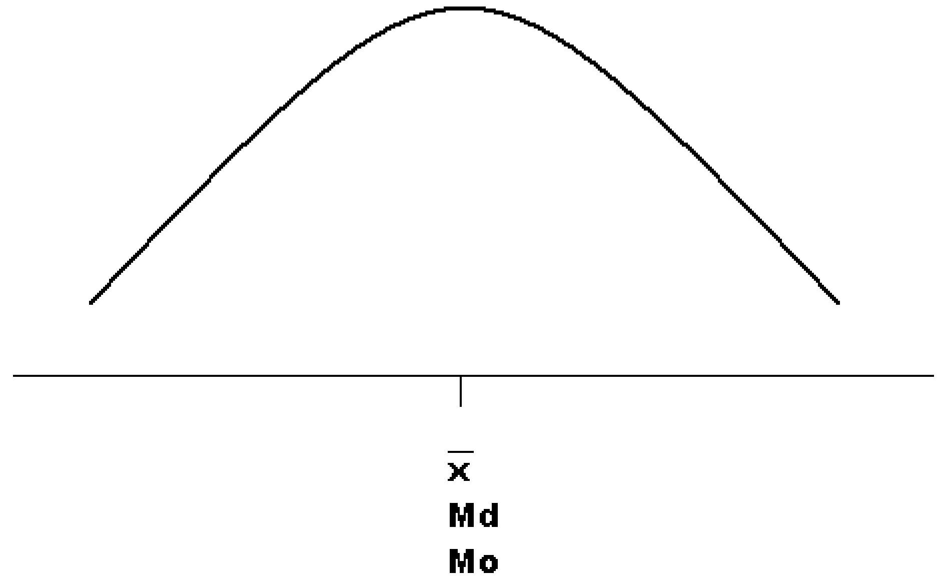
| Tipo | FA | FR |
|---|---|---|
| Familiar | 176 | 0,52 |
| Sport | 109 | 0,32 |
| Trabajo | 54 | 0,16 |
| Total | 339 | 1 |

Variable: Tipo
Moda: Familiar

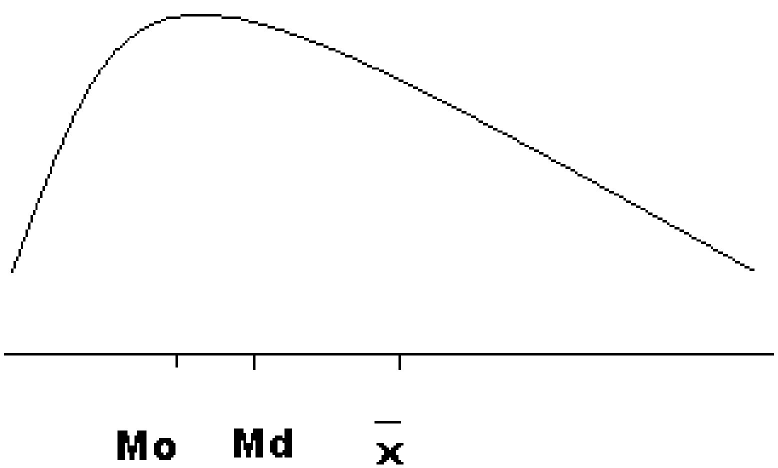
| Tamaño | FA | FR |
|---|---|---|
| Chico | 153 | 0,45 |
| Grande | 43 | 0,13 |
| Medio | 142 | 0,42 |
| Total | 338 | 1 |

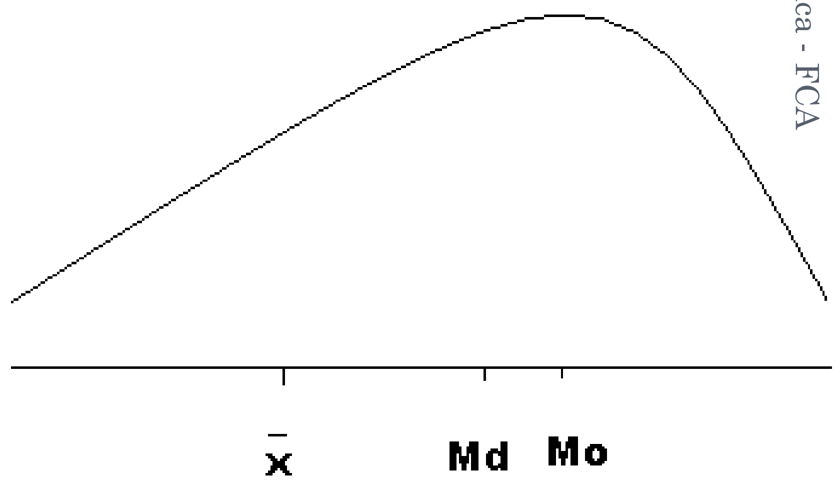
Variable: Tamaño
Moda: Chico

**Distribución simétrica**




**Distribución asimétrica**



# MEDIDAS RESUMEN

- Una medida resumen es un número. Se obtiene a partir de una muestra y sirven para describirla o caracterizarla.
- Las medidas resumen permiten tener una idea rápida de cómo son los datos.
- La media por ejemplo, nos permite sintetizar los datos con un único valor.
- Las medidas que hemos visto y que veremos son útiles para describir los datos.

# MEDIDAS DE POSICIÓN NO CENTRALES

## CUANTILES

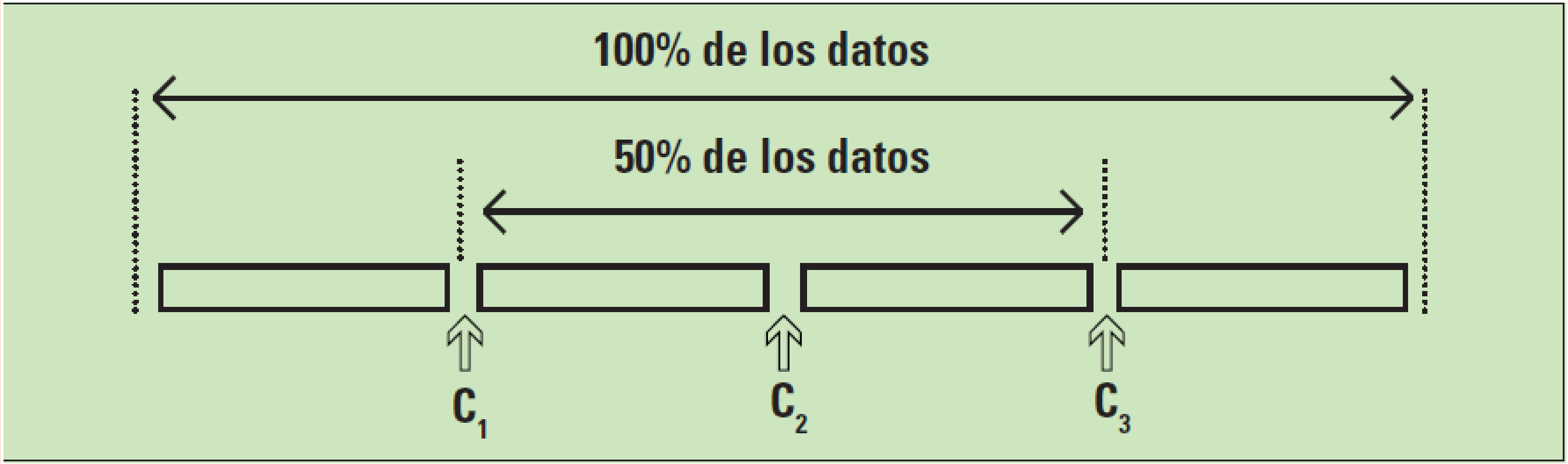
Si x1, x2,....., xn constituyen una muestra de tamaño *n* entonces el **cuantil p**, es el valor tal que el número de valores menores o iguales a él constituyen la proporción p del número total de observaciones en la muestra.

La mediana sería es el cuantil 0.50 de su distribución de frecuencias muestral, ya que el 50% de los datos es menor o igual que la mediana.

# MEDIDAS DE POSICIÓN NO CENTRALES

- CUARTILES): dividen al total de datos en 4 grupos.
- DECILES : dividen al total de datos en 10 grupos.
- PERCENTILES: dividen al total de datos en 100 grupos.

# Cuartiles




Los cuartiles son

$C_1$: cuartil 1 (25% de los datos $< C_1$)

$C_2$: cuartil 2 ( 50% de los datos $< C_2$)

$C_3$: cuartil 3( 75% de los datos $< C_3$)

# La posición de los cuartiles se obtiene:

**n impar**

$Posición$ $C_1$: 1/4 (n+1)
$Posición$ $C_2$: 2/4 (n+1)
$Posición$ $C_3$: 3/4 (n+1)

**n par**

Promedio de las posiciones
$C_1$: (n/4) y (n/4)+1)
$C_2$: (n/2) y (n/2)+1)
$C_3$: 3n/4 y (3n/4)+1

El cuartil 2 , $C_2$ es la mediana.

*S*e pueden calcular los cuartiles C1 y C3, como las medianas de la mitad inferior de la muestra y la mitad superior.

Los **deciles** son

$C_1$: decil 1 (10 % de los datos $<C_1$)

$C_2$: decil 2 ( 20% de los datos $<C_2$)

$C_3$: decil 3( 30% de los datos $<C_3$)

……………………………………………

$C_9$:decil 9 (95% de los datos $<C_9$)

Los **percentiles** son

$C_1$: percentil 1 (1% de los datos $<C_1$)

$C_2$: percentil 2 ( 2% de los datos $<C_2$)

……………………………………………

$C_{10}$: percentil 10 ( 10% de los datos $<C_{10}$)

……………………………………………

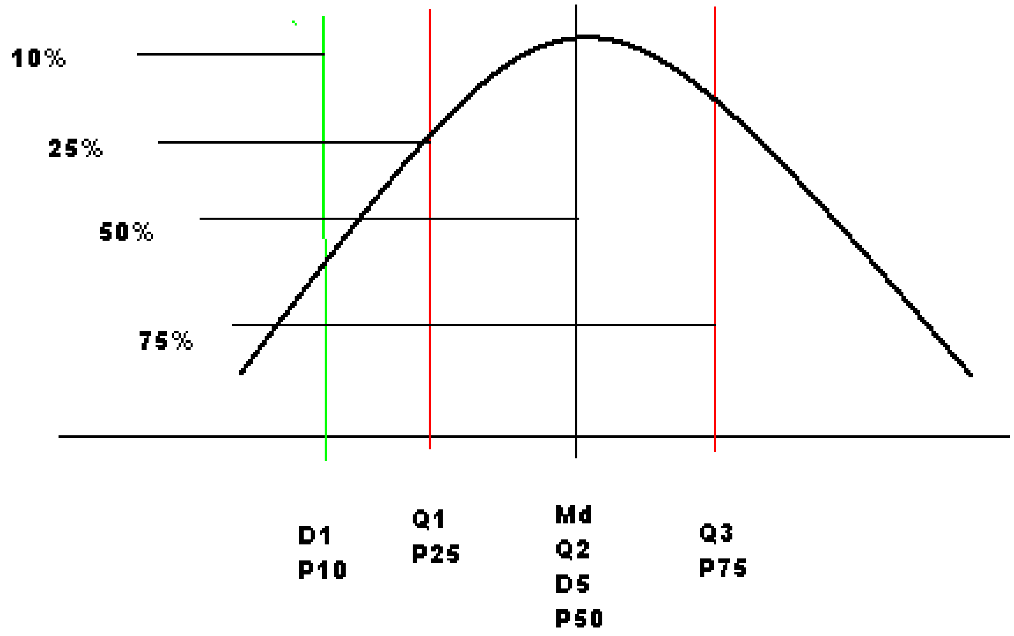
$C_{30}$: percentil 30( 30% de los datos $<C_{30}$)

……………………………………………

$C_{99}$: percentil 99( 99% de los datos $<C_{99}$)

# CUANTILES

- CUARTILES (Q)
- DECILES (D)
- PERCENTILES(P



# Gráfico de caja (Box-plot).

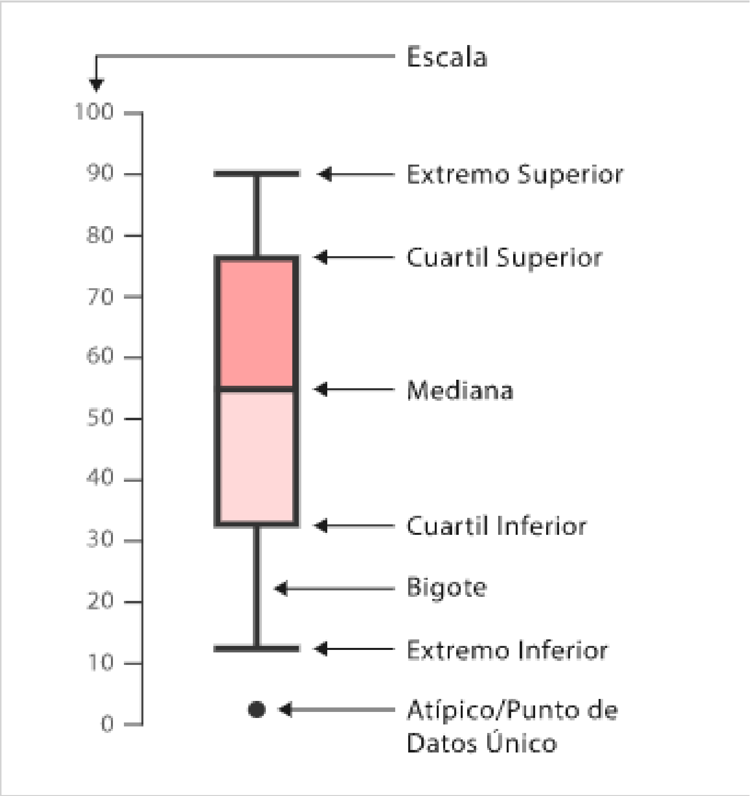
Es un gráfico muy utilizado en estadística ya que con cinco medidas resumen se realiza el gráfico.

Esas medidas son : el mínimo, el cuartil 1, la mediana, el cuartil 3 y el máximo.

*Algunos programas estadísticos basados en diferentes autores en lugar de utilizar el mínimo y máximo valor utilizan percentiles (P5 y P95) o (P1 y P99) de esta manera pueden representar valores atípicos.*

Se observa una caja la cual representa el 50% de los datos se encuentran entre cuartiles 1 y 3 .

# Gráfico de caja (Box-plot).




Un diagrama de cajas y bigotes es una manera conveniente de mostrar visualmente grupos de datos numéricos a través de sus cuartiles.

Las líneas que se extienden paralelas a las cajas se conocen como «bigotes», y se usan para indicar variabilidad fuera de los cuartiles superior e inferior. Los valores atípicos se representan a veces como puntos individuales que están en línea con los bigotes. Los diagramas de cajas y bigotes se pueden dibujar vertical u horizontalmente.

# Ejemplo del uso de los percentiles

PERCENTILES DE LA ALTURA (m) DE MUJERES Y VARONES DE 16 AÑOS. TABLA 19.1

| Percentil | **3** | 10 | **25** | 50 | 75 | 90 |
|---|---|---|---|---|---|---|
| Mujer | **1,49** | 1,53 | **1,56** | 1,60 | 1,64 | 1,68 |
| Varón | 1,56 | 1,60 | 1,65 | 1,70 | 1,74 | 1,79 |

***Fuente.*** *Guías para la Evaluación del Crecimiento. Sociedad Argentina de Pediatría. 2001.*

El cuadro representa valores que resumen la altura medida en metros de mujeres y varones de 16 años. Veamos como interpretamos los valores remarcados:

Percentil 3=1.49 significa que sólo el 3% de las mujeres miren menos que 1.49 a los 16 años, y que el 97 % mide mas que 1.49 metros.

Percentil 25=1.56 significa que el 25% de las mujeres miden menos que 1.56 a los 16 años, y que el 75 % mide mas que 1.56 metros.

# Veamos el siguiente ejemplo que nos conduce a la necesidad de otras medidas

Se presenta 3 grupos de datos, intentaremos describirlos con su promedio

| | | | | | | | | Media |
|---|---|---|---|---|---|---|---|---|
| GRUPO 1 | 8 | 8 | 9 | 10 | 11 | 12 | 12 | $\overline{X_1}=10$ |
| GRUPO 2 | 5 | 6 | 8 | 10 | 12 | 14 | 15 | $\overline{X_2}=10$ |
| GRUPO 3 | 1 | 2 | 5 | 10 | 15 | 18 | 19 | $\overline{X_3}=10$ |

Vemos que la media es la misma, sin embargo cada grupo tiene distintos valores, una medida central parecería insuficiente para describir a los datos.

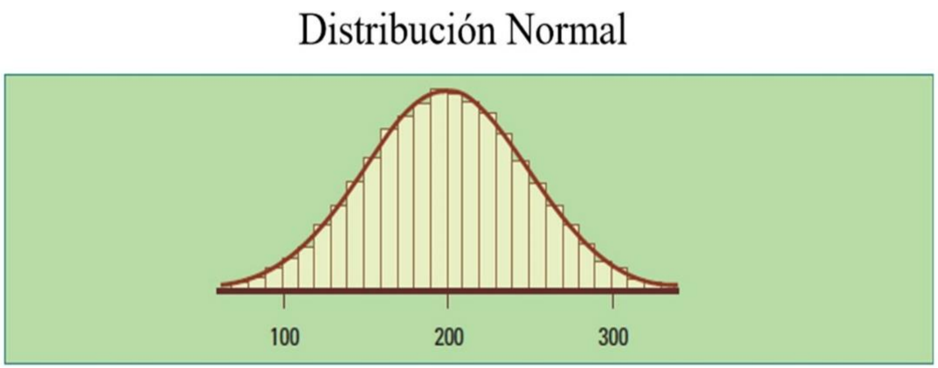
# Si calculamos otra medida como el Rango = Max-Min

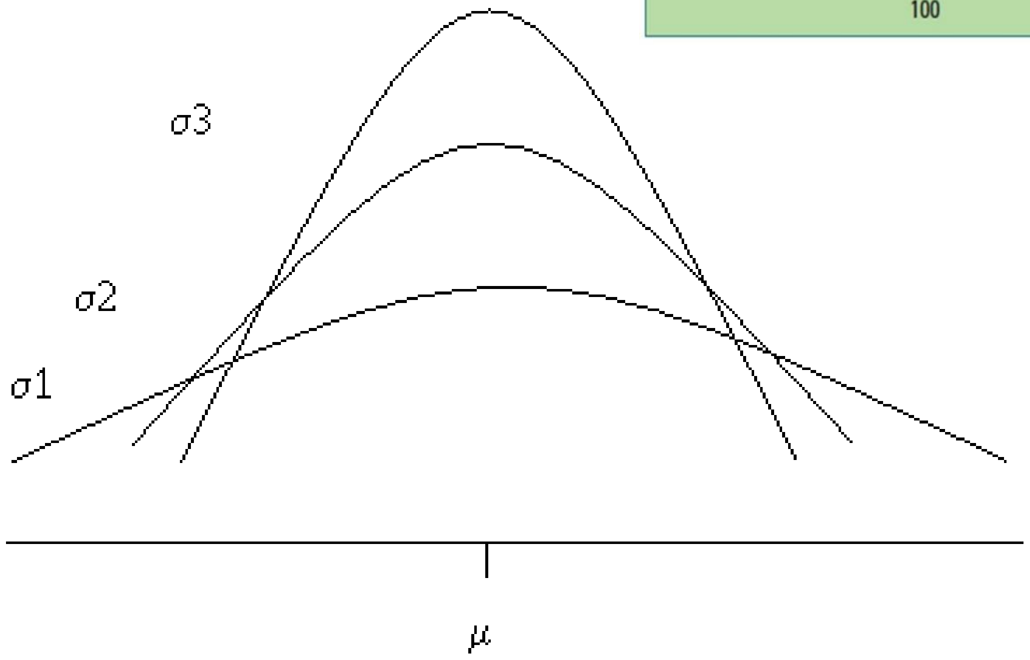
| | | | | | | | | Media | Rango |
|---|---|---|---|---|---|---|---|---|---|
| GRUPO 1 | 8 | 8 | 9 | 10 | 11 | 12 | 12 | $\overline{X_1}=10$ | $R_1 = 4$ |
| GRUPO 2 | 5 | 6 | 8 | 10 | 12 | 14 | 15 | $\overline{X_2}=10$ | $R_2 = 10$ |
| GRUPO 3 | 1 | 2 | 5 | 10 | 15 | 18 | 19 | $\overline{X_3}=10$ | $R_3 = 18$ |

Introducimos el concepto de Rango como una medida indicadora de cuan variable son los datos ya que comparamos el máximo y el mínimo valor.

# MEDIDAS DE DISPERSION

# MEDIDAS DE DISPERSIÓN

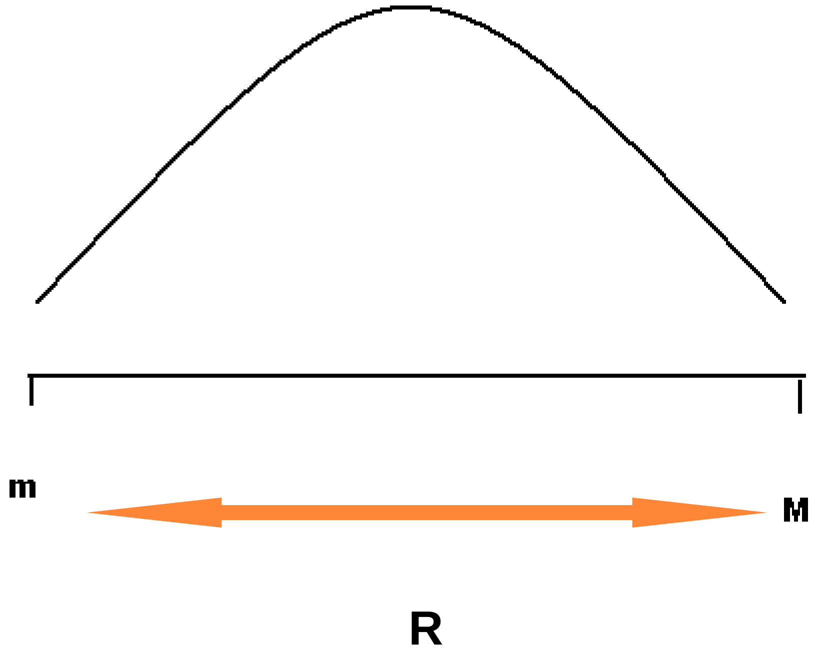



**Informan sobre la VARIABILIDAD que tiene un conjunto de datos.**

# Rango:

R= Max-min




En el ejemplo de precio del producto, el Rango= 1398,70 – 704,20=694,5

# OTRAS MEDIDAS DE DISPERSIÓN

El rango tiene la ventaja que se calcula muy fácil, de manera intuitiva nos permite observar la variabilidad de los datos, sin embargo utiliza sólo 2 datos para su cálculo mostrando la mayor o máxima variación, si hubiera un valor atípico, estaría sobre estimando la variabilidad.

Se presenta a continuación una medida de dispersión basada en la media y que utiliza todos los datos.
El **desvío estándar** representa una distancia típica de cualquier punto del conjunto de datos a su centro (la media).
Se interpreta como la variación promedio que tienen los datos.
Para calcular el desvío estándar primero debemos calcular la **Varianza** cuya interpretación se verá en el último módulo referido a Diseños Experimentales.

Variancia: Es el promedio aritmético de los desvíos cuadráticos de los términos de la serie con respecto a la media aritmética.

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

**De manera abreviada:**

$$S^2 = \frac{\sum_{i=1}^{n} x^2_i - \left(\sum_{i=1}^{n} x_i\right)^2}{n-1}$$

Esta última es simplemente otra expresión de la varianza

Si hubiéramos tenido información de sólo 12 puntos de venta del precio del Producto, hubiéramos tenido otra Media y otra Varianza, ya que es sería una muestra diferente.

| |
|---|
| 1290.8 |
| 1133.5 |
| 1027.2 |
| 953.6 |
| 1013.5 |
| 1088.4 |
| 910.8 |
| 1175.4 |
| 1083.7 |
| 1070.6 |
| 853.8 |
| 1160.1 |

Media =1063.45

Varianza= $S^2$

$S^2 = ((1290.8-1063.45)^2+(1133.5-1063.45)^2+..........+(1160.1-1063.45)^2)/12$

=14789.45

DESVÍO ESTÁNDAR: ES LA RAÍZ CUADRADA DE LA VARIANZA

$$\sqrt{S^2}$$

Desvío Estándar= $\sqrt{S^2}$= √(14789.45)= 121.61

Esto significa que la diferencia o variación promedio del precio del producto es de $ 121.61 en los diferentes puntos de venta.

# Coeficiente de Variación

Es una medida es adimensional y permite comparar la variabilidad de características medidas en diferentes escalas.

$$CV = \frac{S}{\bar{x}}.100$$

Por ejemplo, si se comparan dos índices para medir nivel de ataque de pulgones y ambos están basados en técnicas completamente diferentes, que dan puntajes cuyas unidades de medida son distintas, se dirá que el índice que tenga menor coeficiente de variación es el menos variable.

# Resumiendo

Los parámetros son valores calculados con todos los datos de la población y los estimadores son valores calculados con los datos de la muestra.

Cualquier conjunto de datos tiene 2 propiedades importantes un Valor Central y la Dispersión alrededor de ese valor.