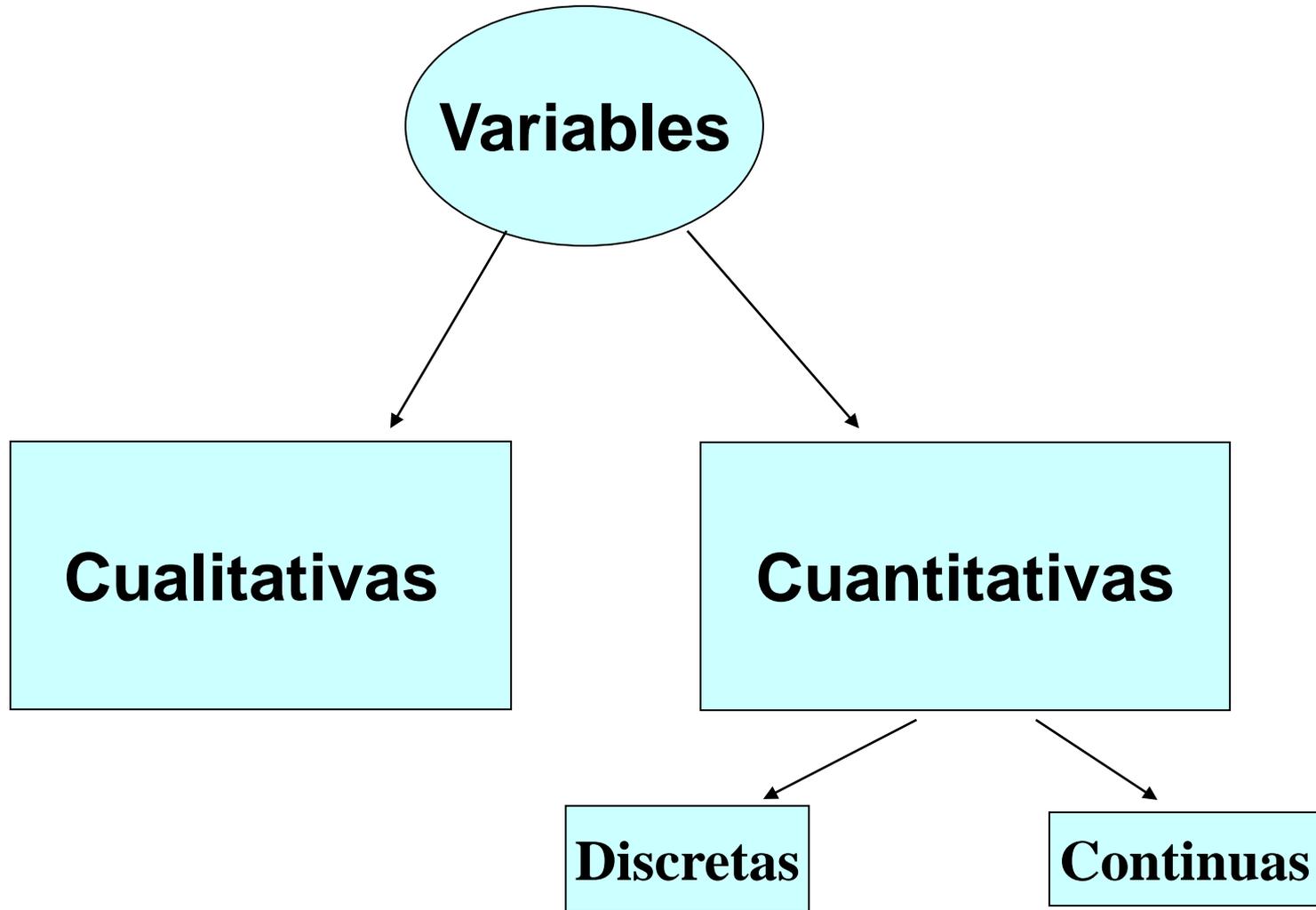


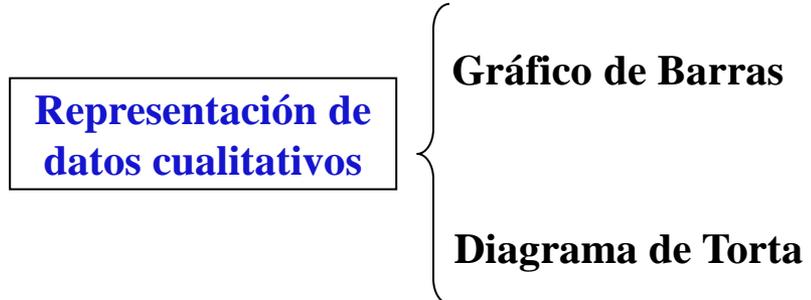
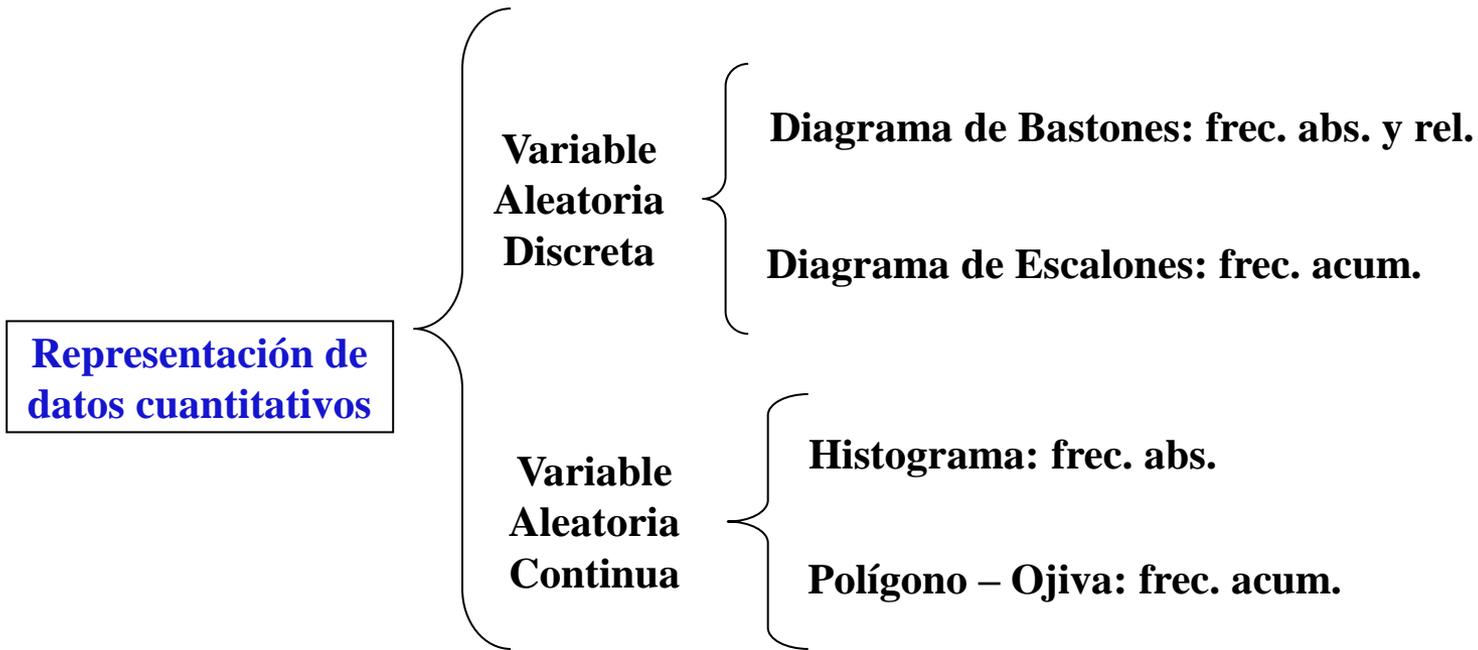
# Estadística Descriptiva

MEDIDAS DE POSICIÓN Y DISPERSIÓN

# Clasificación de variables por el tipo de resultado



# Resumen



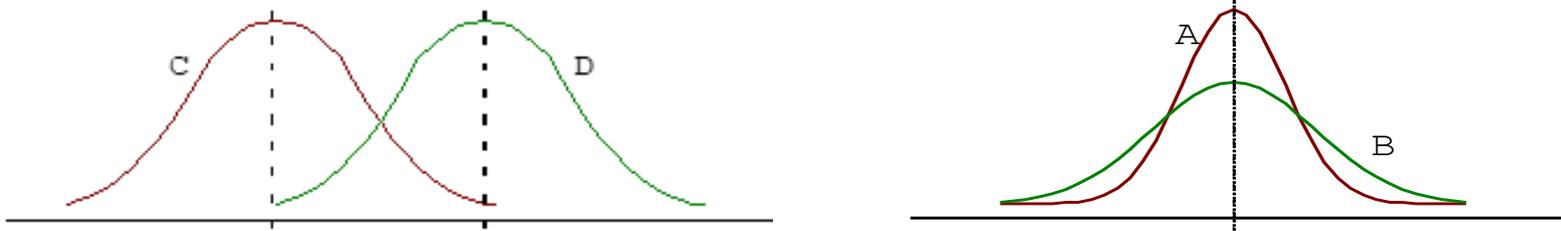
# Estadística Descriptiva

## Medidas resumen

Sintetizan con un valor las características de la distribución de la variable, en el conjunto de datos observado.

El objetivo es ser representativas de ellos.

# Medidas Resumen



Las podemos clasificar en medidas de:

**Posición:** Identifican la **ubicación** de la distribución de la variable a lo largo de una escala.

Caracterizan al conjunto de valores de la variable mediante algún **valor típico** considerado como el mejor representante global (que generalmente presenta una tendencia a situarse en valores centrales).

**Dispersión:** **Grado de** homogeneidad o heterogeneidad presente en el conjunto de objetos estadísticos.

# Medidas Resumen

## MEDIDAS DE POSICIÓN

Son valores que intentan representar al conjunto de datos y tienden a ubicarse *hacia el centro* de la distribución.

Frecuentemente se los llama PROMEDIOS.

- Modo ( $M_o$ )
- Mediana ( $M_n$ )

Media Aritmética  $\left\{ \begin{array}{l} (\mu) \text{ si los datos provienen de una población} \\ (\bar{X}) \text{ si los datos provienen de una muestra} \end{array} \right.$

*Medidas que representan Posición (pero no centro).*

- Cuartiles ( $Q_1, Q_2, Q_3$ )
- Percentiles ( $P_1, P_2, \dots, P_{99}$ )

Se tienen los siguientes valores correspondientes a 12 personas que asistieron a un curso de capacitación para la elaboración de procedimientos escritos, observando al finalizar el curso las siguientes variables: calificación obtenida, y tiempo en responder la evaluación. Luego de 7 días se contaron la cantidad de errores cometidos en la elaboración de procedimientos .

¿Qué tipo de variables identifica?

	Cualitativa	Cuantitativa Discreta	Cuantitativa Continua
Persona	Calificación (V)	Número de errores (Y)	Tiempo de respuesta (X)
1	Sobresaliente	0	3,4
2	Sobresaliente	1	5,8
3	Desaprobado	3	4,3
4	Sobresaliente	2	6,9
5	Aprobado	1	8,6
6	Sobresaliente	1	9,3
7	Aprobado	2	1,7
8	Aprobado	4	5,4
9	Sobresaliente	0	6,3
10	Desaprobado	3	7
11	Aprobado	2	8,2
12	Sobresaliente	1	5,1

## Medidas de Posición

Con **Escalas Nominales**: La única medida descriptiva que puede considerarse en el caso de variables categóricas es la CATEGORIA MODAL.

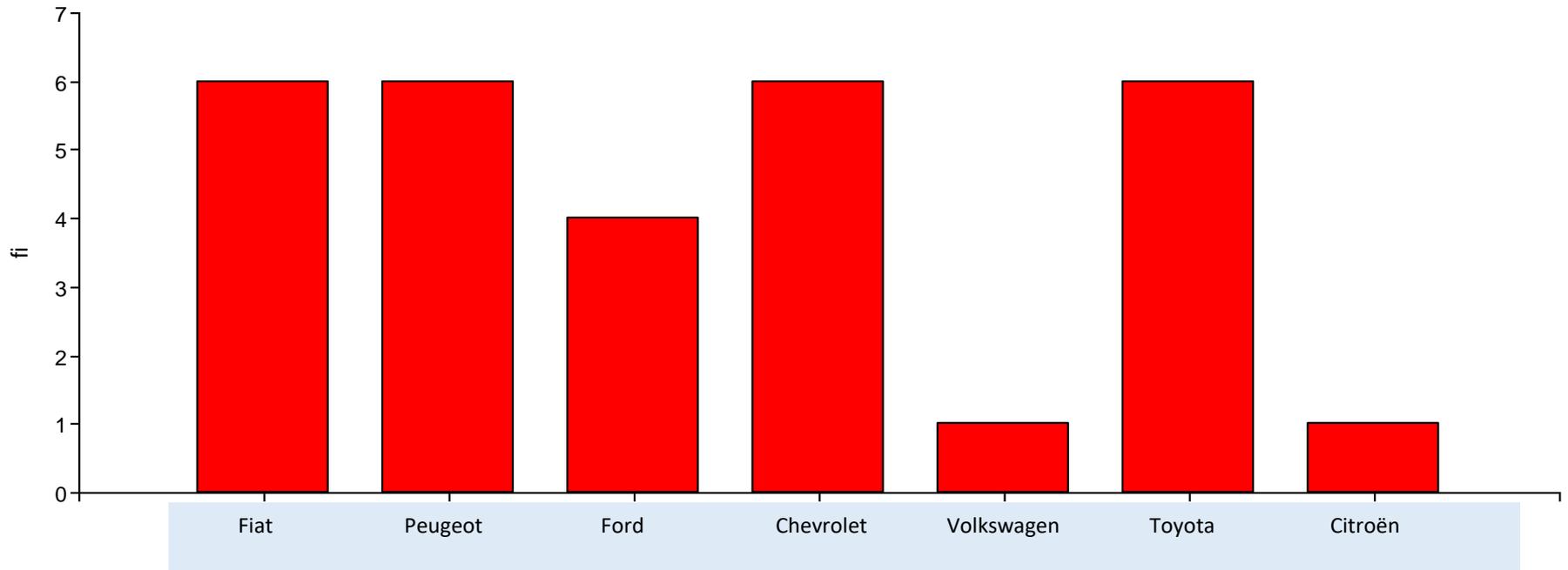
**MODO**: se define como el **valor** de la variable que tiene asociada la **mayor frecuencia** de ocurrencia.

Puede no ser único y tiene las mismas unidades que la variable.

En un ejemplo para la variable

**X**: marca de autos

$x_i$	$f_i$	$h_i$ (%)
Fiat	6	20,00
Peugeot	6	20,00
Ford	4	13,33
Chevrolet	6	20,00
Volkswagen	1	3,33
Toyota	6	20,00
Citroën	1	3,33
Total	30	100



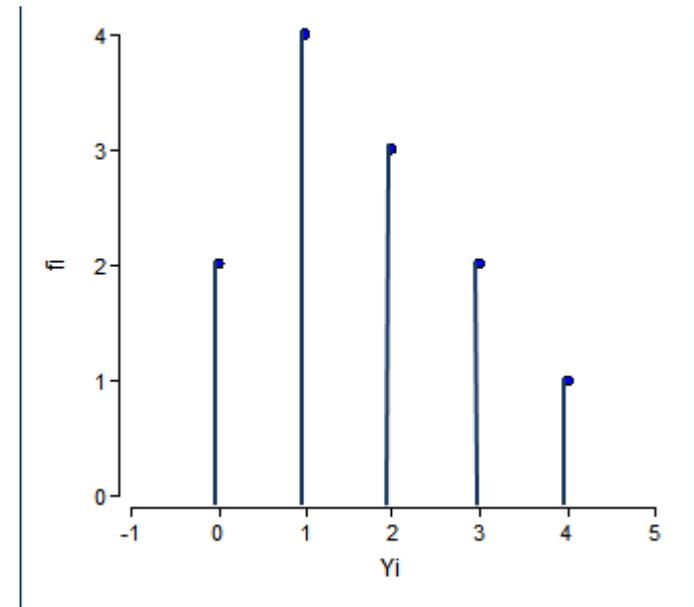
## MODO v.a. Discreta

## Medidas de Posición

**Y:** número de errores al elaborar procedimientos escritos a los 7 días de realizar el curso

$Y_i$	$f_i$	$F_i$
0	2	2
1	4	6
2	3	9
3	2	11
4	1	12
	12	

Modo  $\rightarrow$  1 error en siete días



# MODO v.a. Continua

# Medidas de Posición

**X:** tiempo en responder la evaluación

$X_i$	$X_i'$	$f_i$	$F_i$
[1,5 – 3,5)	2,5	2	2
[3,5 – 5,5)	4,5	3	5
[5,5 – 7,5)	6,5	4	9
[7,5 – 9,5)	8,5	3	12
		12	

Ubicar el Intervalo Modal, y luego aplicar la fórmula

$$M_o = L_i + c \left[ \frac{\Delta_1}{\Delta_1 + \Delta_2} \right]$$

$L_i$  = Límite inferior del intervalo modal.

$f_{M_o}$  = Frecuencia absoluta del intervalo modal.

$f_{M_o-1}$  = Frecuencia absoluta del intervalo anterior al modal.

$f_{M_o+1}$  = Frecuencia absoluta del intervalo posterior al modal.

$C$  = amplitud del intervalo

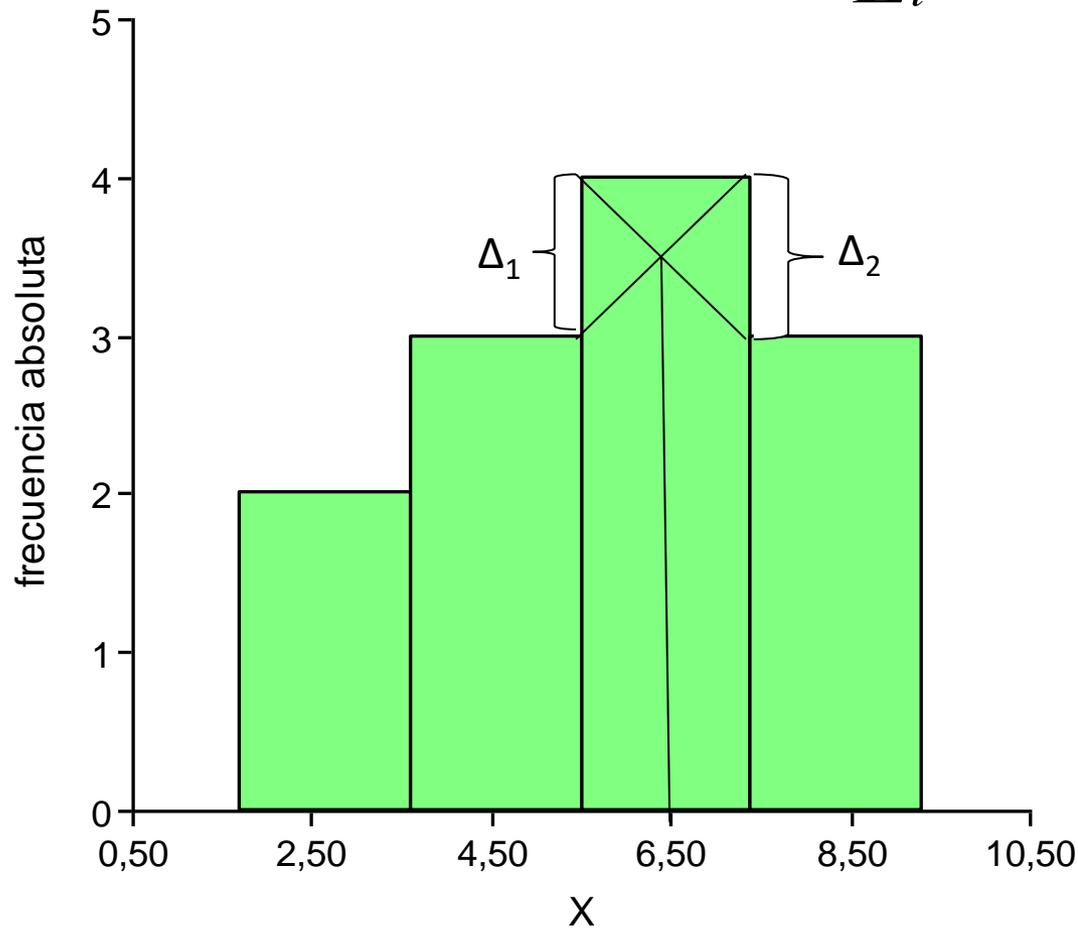
$$M_o = L_i + \left[ \frac{f_{M_o} - f_{M_o-1}}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})} \right] \cdot c$$

$$M_o = 5,5 + 2 \cdot \left[ \frac{4 - 3}{(4-3) + (4-3)} \right] = 6,5$$

**MODO** v.a. Continua

## Medidas de Posición

$$Mo = L_i + c \left[ \frac{\Delta_1}{\Delta_1 + \Delta_2} \right]$$



# MEDIANA - Medidas de Posición

**Mediana:** es un valor dentro del rango de la variable tal que el 50% de las observaciones es inferior o igual a este valor y el 50% de las observaciones es superior o igual al mismo. Se designa con **Mna**, y tiene las mismas unidades de medida que la variable.

Observación: así definida la mediana podría no ser única, sin embargo existe una convención para determinar el valor.

*La Mediana depende de la posición que ocupan los datos en la ordenación y no del valor numérico de los mismos.*

## **VENTAJAS:**

Fácil de calcular.

Es de sencilla interpretación.

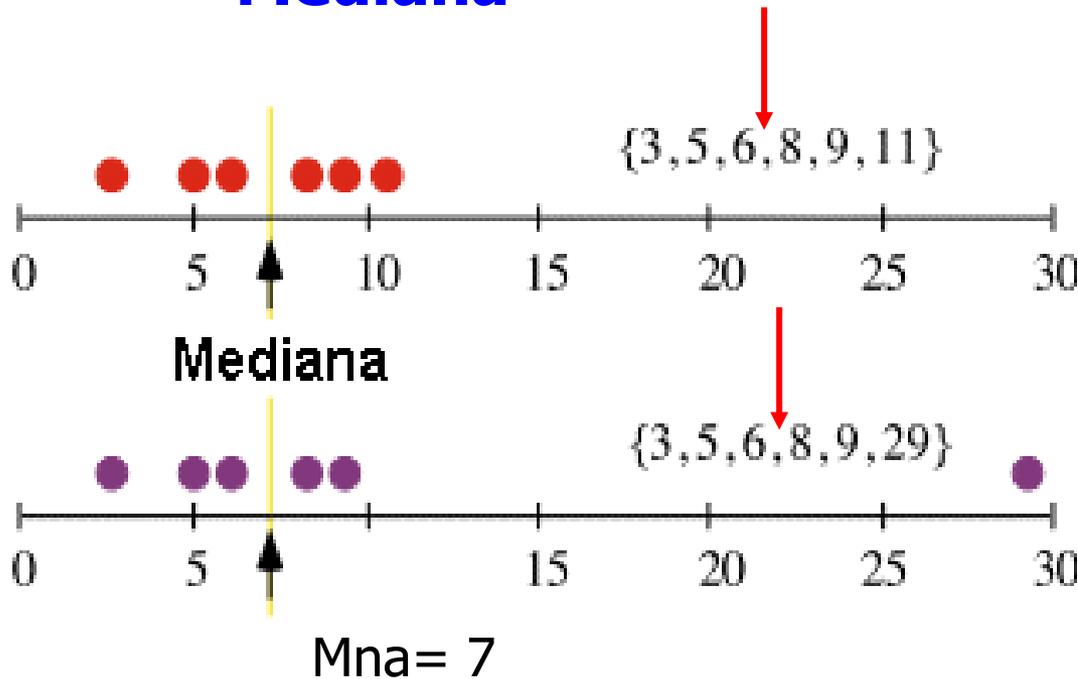
Es única, con la convención, para cada grupo de datos.

NO está influenciada por los valores extremos, lo que la hace más representativa en casos de asimetría.

## **DESVANTAJA:**

*No tiene en cuenta a los valores de la variable, sino solo a su posición.*

# Mediana



Si el tamaño de la muestra es manejable, se ordenan las observaciones de menor a mayor y la mediana será, el valor central (cuando  $n$  es impar), o el promedio de los dos valores centrales (cuando  $n$  es par).

$$Mna^0 = \frac{n + 1}{2}$$

Se puede observar que:

- 1) La Mediana no se ve afectada por la presencia de valores alejados en el conjunto de datos.
- 2) La Mediana no hace uso de toda la información disponible.
- 3) Es conveniente cuando los datos son asimétricos.

Cuando tenemos gran cantidad de datos se usan las Frec. Acumuladas y la posición de la mediana.

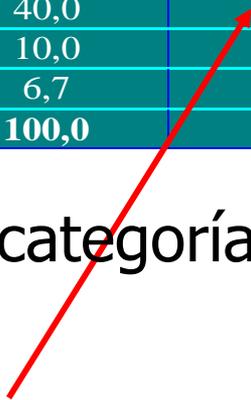
## Medidas de Posición

Con **Escalas Ordinales** además de la Categoría Modal se puede obtener la CATEGORIA MEDIANA.

**CATEGORÍA MEDIANA:** es aquella que se sitúa en el centro de la serie al ordenar los datos, es decir en la posición dada por  $[(n+1)/2]$



<i>Satisfacción del Cliente</i>	<i>Nº de clientes</i>	<i>Porcentajes</i>	<i>Frecuencias Acumuladas</i>	<i>Porcentajes Acumulados</i>
<i>Completamente Satisfecho</i>	13	43,3	13	43,3
<i>Satisfecho</i>	12	40,0	25	83,3
<i>Insatisfecho</i>	3	10,0	28	93,3
<i>Completamente Insatisfecho</i>	2	6,7	30	100
<i>Total</i>	30	100,0		



En este ejemplo, la categoría mediana es **SATISFECHO**.

$$Mna^0 = \frac{n + 1}{2} = (30+1)/2 = 15,5$$

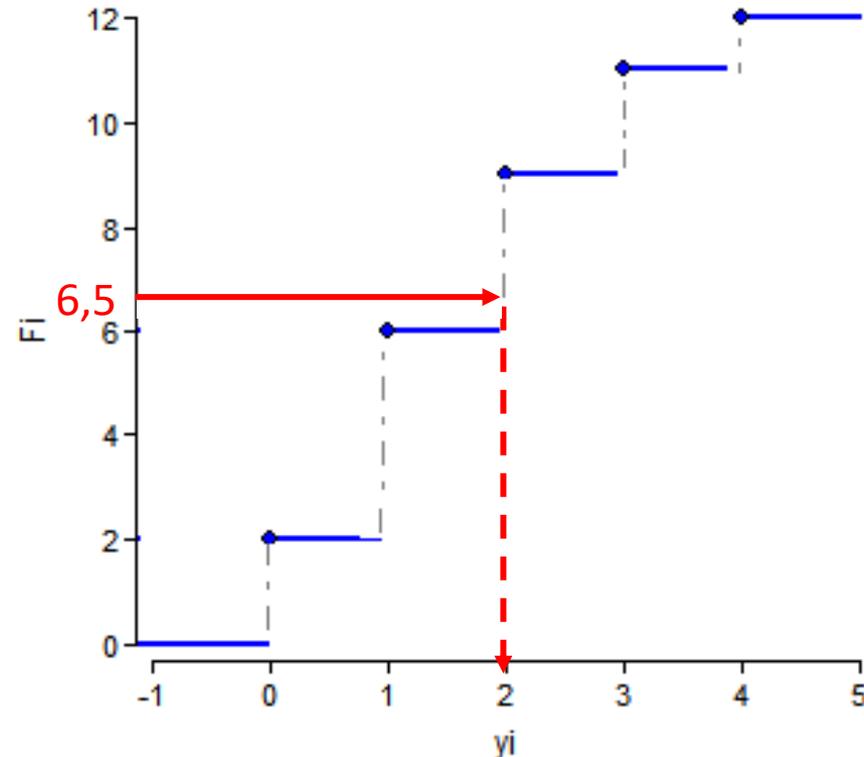
# Ejemplo V. A. Cuantitativa Discreta

**MEDIANA (Mna):** La Mediana divide al conjunto ordenado en dos partes con igual porcentaje de datos.

**Y:** número de errores al elaborar procedimientos escritos a los 7 días de realizar el curso

$Y_i$	$f_i$	$F_i$
0	2	2
1	4	6
2	3	9
3	2	11
4	1	12
	12	

$$Mna^0 = \frac{n + 1}{2} = (12+1)/2 = 6,5$$



Mna: 2 errores en siete días

# MEDIANA

# Medidas de Posición

Aquí encontraremos el intervalo mediana. (V. A. Cuantitativa Continua)

En ejemplo es  $[5,5; 7,5)$  y luego por la fórmula  $Mna^0 = \frac{n + 1}{2}$

$X_i$	$X_i'$	$f_i$	$F_i$
$[1,5 - 3,5)$	2,5	2	2
$[3,5 - 5,5)$	4,5	3	5
$[5,5 - 7,5)$	6,5	4	9
$[7,5 - 9,5)$	8,5	3	12
		12	

$L_i$  = Límite inferior del intervalo mediana.

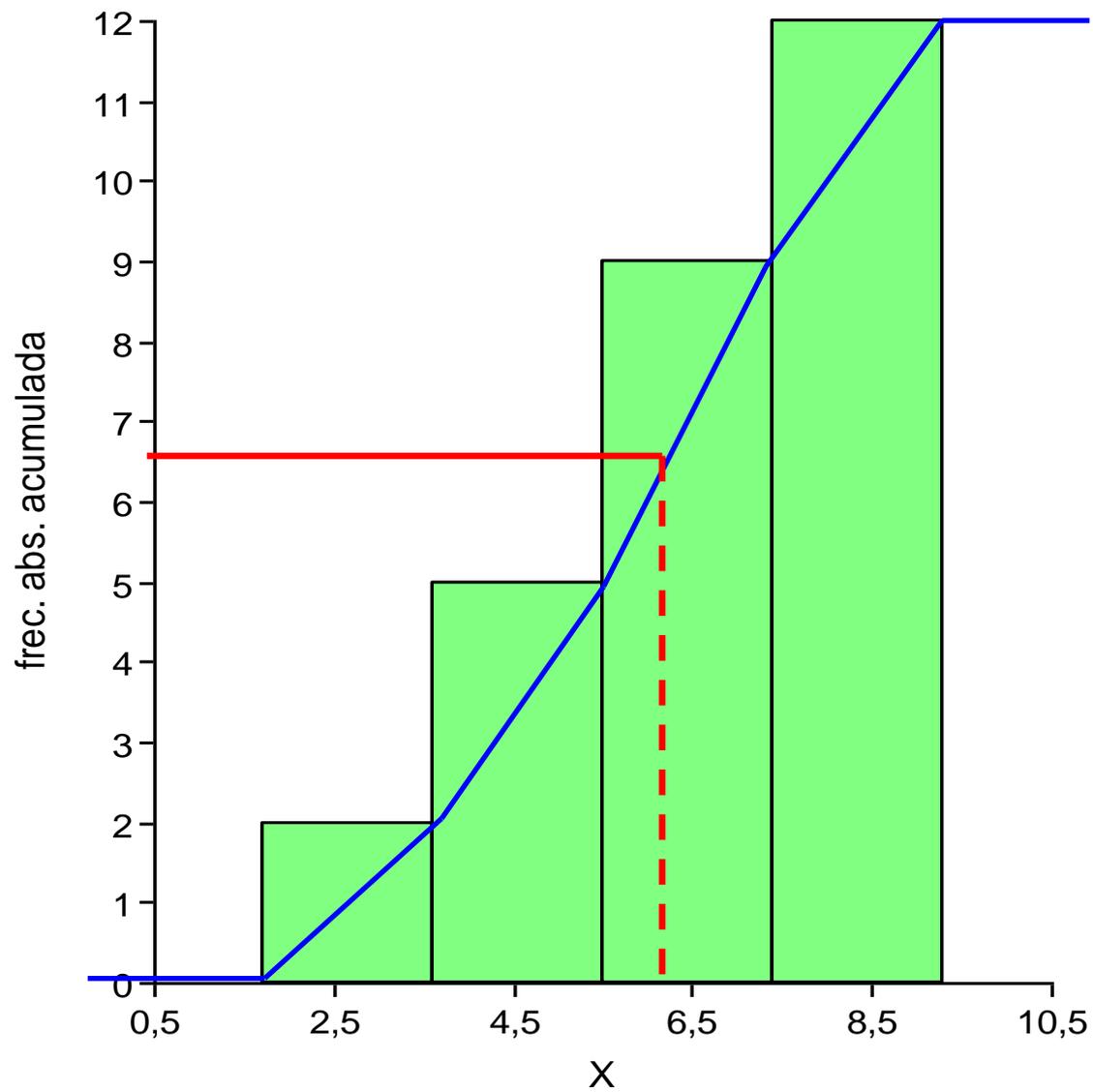
$F_i$  = Frecuencia acumulada hasta el intervalo mediana.

$F_{i-1}$  = Frecuencia acumulada hasta el intervalo anterior a la mediana.

$c$  = amplitud del intervalo.

$$Mna = L_i + \left[ \frac{\left(\frac{n+1}{2}\right) - F_{i-1}}{(F_i - F_{i-1})} \right] \cdot c$$

$$Mna = 5,5 + 2 \cdot \left[ \frac{(12+1) - 5}{9 - 5} \right] = 6,25$$



# MEDIA ARITMÉTICA – Medidas de Posición

**Media aritmética:** es lo que frecuentemente se llama **promedio** de los datos. Se representa con  $\bar{X}$ . En su cálculo intervienen todos los valores observados de la variable y su fórmula es:

$$\text{Media Aritmética Muestral } \bar{x} = \frac{\sum x_i}{n} ; \bar{x} = \frac{\sum x_i f_i}{n} ; \bar{x} = \frac{\sum x'_i f_i}{n}$$

Media Aritmética Poblacional =  $\mu$

Es un valor numérico comprendido entre  $x_{\text{Máx}}$  y  $x_{\text{mín}}$ , pudiendo o no coincidir con uno de los valores observados, es única y tiene las mismas unidades de medida que la variable.

## **VENTAJAS:**

- Fácil de calcular.
- Es única en cada grupo de datos.
- Es de sencilla interpretación.
- Utiliza todos los valores disponibles para su cálculo

# MEDIA ARITMÉTICA – Medidas de Posición

## DESVENTAJAS:

- **No** puede computarse en el caso de **intervalos abiertos**. Por ejemplo en el caso de ingresos a veces se deja un último intervalo con 5000\$ o más, eso es un intervalo abierto.
- **Muy influenciada por valores extremos**, por ende, puede perder representatividad en casos de presencia de un pequeño conjunto de valores muy distintos a los del grupo general.

Ejemplo:

Conjunto 1: 2 – 4 – 5 – 6 – 7 – 10 – 11

$$\bar{x} = 6,428$$

Conjunto 2: 2 – 4 – 5 – 6 – 7 – 10 – 90

$$\bar{x} = 17,714$$

# MEDIA ARITMÉTICA – Medidas de Posición

## PROPIEDADES

a) La sumatoria de los desvíos (diferencia entre los valores observados y la media aritmética) es igual a cero.

Para datos no agrupados. 
$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

y para datos agrupados 
$$\sum_{i=1}^m (x'_i - \bar{x}) f_i = 0$$

b) La suma de cuadrados de los desvíos es mínima si se calculan estos respecto de la media aritmética correspondiente a ese conjunto de datos.

$$\sum_{i=1}^n (x_i - \bar{x})^2 \leq \sum_{i=1}^n (x_i - a)^2 \quad \text{para cualquier } a \in \mathbb{R}$$

# MEDIA ARITMÉTICA – Medidas de Posición

## PROPIEDADES

c) El promedio de un grupo de medias aritméticas, cada una de ellas ponderada por la cantidad de observaciones que le dio origen, coincide en el promedio de las observaciones individuales. Analicemos el siguiente caso. Si se tienen los siguientes grupos de datos::

$$\begin{aligned}x_{1i} &= 3 - 4 - 8 - 3 - 7 \\x_{2i} &= 10 - 12 - 11 - 13 - 14 \\x_{3i} &= 7 - 3 - 9 - 8 - 3 - 1 - 4\end{aligned}$$

y se quieren calcular las medias aritméticas, entonces se obtienen

$$\bar{x}_1 = 5 \quad n_1 = 5 \quad ; \quad \bar{x}_2 = 12 \quad n_2 = 5 \quad ; \quad \bar{x}_3 = 5 \quad n_3 = 7$$

$$\bar{x} = \frac{\bar{x}_1 \cdot n_1 + \bar{x}_2 \cdot n_2 + \bar{x}_3 \cdot n_3}{n_1 + n_2 + n_3} = \frac{25 + 60 + 35}{17} = 7,06 \quad \text{Coincide con} \quad \bar{x} \text{ grande} = \frac{\sum_{i=1}^{17} x_i}{17} = 7,06$$

d) Si un conjunto de observaciones toma el mismo valor para todas ellas, es decir  $x_i = k$  para todo  $i = 1, \dots, n$ , la media aritmética es igual al valor de la constante.

$$\bar{x} = \frac{\sum k}{n} = \frac{nk}{n} = k$$

# MEDIA ARITMÉTICA – Medidas de Posición

## PROPIEDADES

La media aritmética de una constante es la misma constante.

$$\bar{x} = \frac{\sum k}{n} = \frac{nk}{k} = k$$

e) La media aritmética de una variable por una constante es igual a la media aritmética de la variable multiplicada por la constante.

$$\text{si } y_i = k * x_i \Rightarrow \bar{y} = k * \bar{x}$$

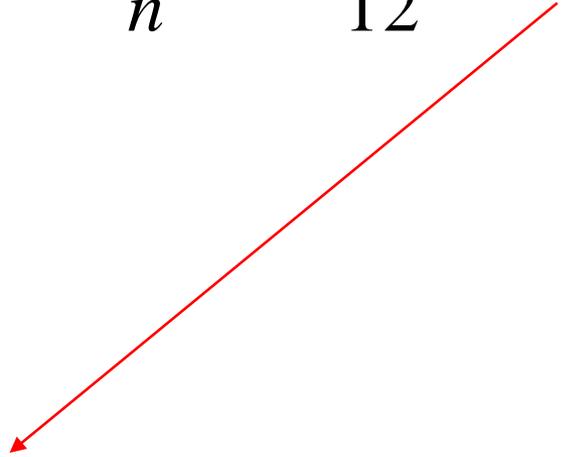
f) La media aritmética de una variable mas (o menos) una constante es igual a la media aritmética de la variable mas (o menos) la constante..

$$\text{si } y_i = x_i \pm k \Rightarrow \bar{y} = \bar{x} \pm k$$

**Media aritmética**

**Y:** número de errores cometidos a los 7 días de finalizado el curso.

$y_i$	$f_i$	$F_i$	$y_i \cdot f_i$
0	2	2	0
1	4	6	4
2	3	9	6
3	2	11	6
4	1	12	4
	12		20

$$\bar{y} = \frac{\sum_{i=1}^n y_i \cdot f_i}{n} = \frac{20}{12} = 1,6666$$


Media aritmética

aproximadamente 2 errores en siete días

(Cuantitativa Discreta)

# MEDIA ARITMÉTICA – Medidas de Posición

V. A. Cuantitativa Continua

**MEDIA ARITMÉTICA** para datos agrupados en intervalos

$x_i$	$x'_i$	$f_i$	$x'_i \cdot f_i$
[1,5 – 3,5)	2,5	2	5
[3,5 – 5,5)	4,5	3	13,5
[5,5 – 7,5)	6,5	4	26
[7,5 – 9,5)	8,5	3	25,5
		12	70

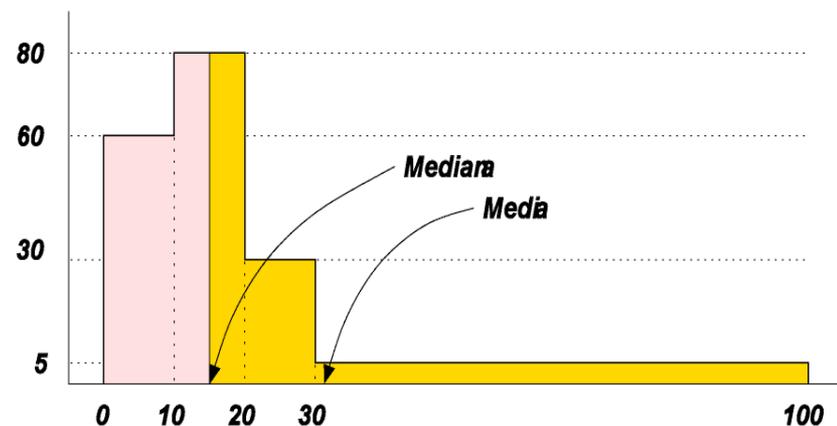
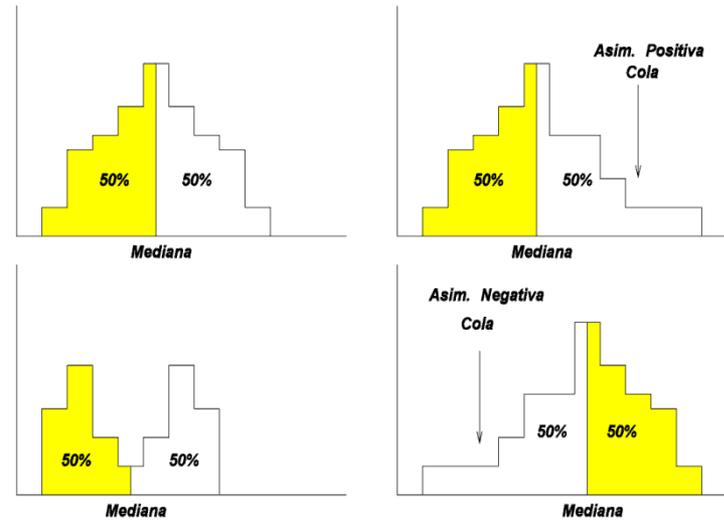
$$\bar{x} = \frac{\sum_{i=1}^m x'_i \cdot f_i}{n}$$

$$\bar{X} = 70/12 = 5,83333$$

# Medidas de Posición

## Asimetría

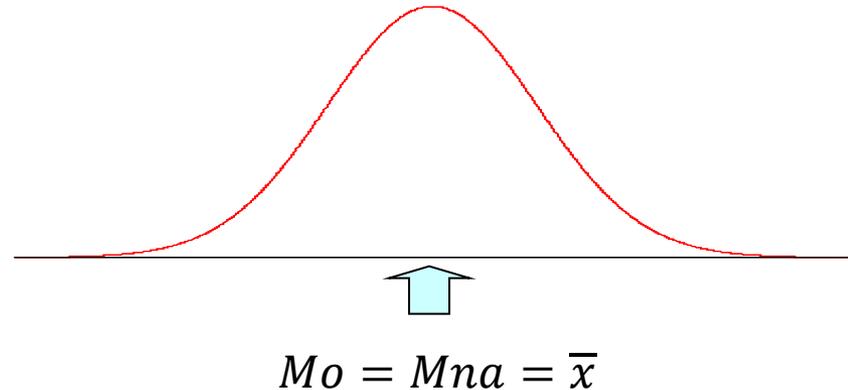
- Una distribución es simétrica si la mitad izquierda de su distribución es la imagen especular de su mitad derecha.
- En las distribuciones simétricas media y mediana coinciden. Si sólo hay una moda también coincide.
- La asimetría es positiva o negativa en función de a qué lado se encuentra la cola de la distribución.
- La media tiende a desplazarse hacia los valores extremos (colas).
- Las discrepancias entre las medidas de centralización son indicación de asimetría.



# Medidas de Posición

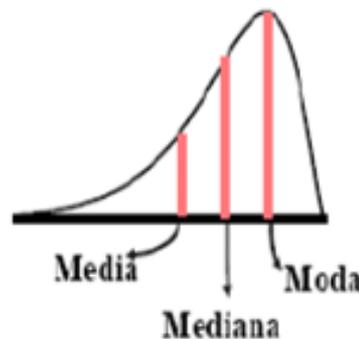
## Posiciones Relativas

- Si la distribución es simétrica y unimodal, las tres medidas ocuparán la abscisa correspondiente al máximo central:

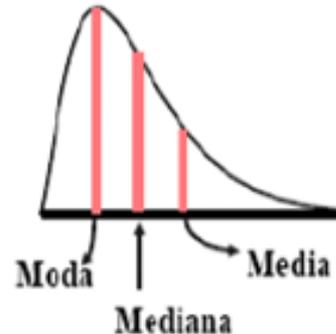


- Si la distribución es asimétrica, por efecto de la presencia de valores alejados:

Asimétrica negativa



Asimétrica positiva



# ELECCION DE LA MEDIDA DE POSICIÓN CENTRAL MAS REPRESENTATIVA

Depende de la naturaleza de la variable y de la forma de la distribución:

## **Variable Cualitativa**

- Escala Nominal: la CATEGORIA MODAL
- Escala Ordinal: la CATEGORIA MODAL o la CATEGORIA MEDIANA

## **Variable Cuantitativa**

- Distribución de frecuencias Simétrica: la MEDIA ARITMÉTICA
- Distribución de frecuencias Asimétrica: la MEDIANA

# Medidas de Posición

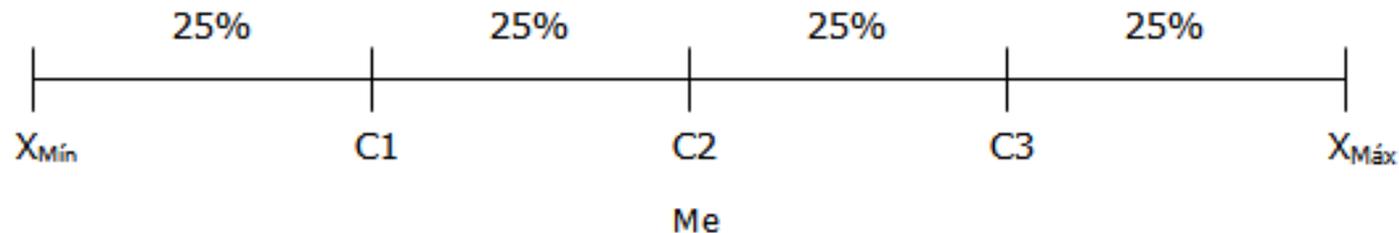
## CUARTILES (Q1, Q2, Q3):

Son tres valores de la variable (observables o no) que dividen al conjunto de datos, una vez que han sido ordenados, en cuatro partes del mismo tamaño.

$C_1=Q_1$  separa el 25% más bajo.

$C_2=Q_2$  es la Mediana (separa el 50%).

$C_3=Q_3$  separa el 25% más alto.



El cuartil inferior ( $C_1$ ) es el valor (observado o no) que ocupa la posición  $(n + 3)/4$  en el grupo ordenado, dejando a su izquierda el 25% de las observaciones y a su derecha el 75%.

El cuartil superior ( $C_3$ ) deja a su izquierda el 75% de las observaciones y a su derecha el 25%.

# Medidas de Posición

## CUARTILES (Q1, Q2, Q3):

Ejemplo: Dados los siguientes 13 datos ordenados en una tabla de frecuencias, calcular los cuartiles inferior y superior.

$x_i$	14	18	24	26	35	39	43	45	76	82	88	92	98
-------	----	----	----	----	----	----	----	----	----	----	----	----	----

$$Q_1^o = \frac{n+3}{4} = \frac{16}{4} = 4 \Rightarrow Q_1 = 26$$

$$Mna^o = \frac{n+1}{2} = \frac{14}{2} = 7 \Rightarrow Mna = 43$$

$Q_3$  se cuentan los mismos lugares que  $Q_1^o$  de derecha a izquierda y se lee la observación

$$Q_3 = 82$$

## Medidas de Posición

### **PERCENTILES ( $P_1, P_2, \dots, P_{99}$ ):**

Son noventa y nueve valores de la variable que dividen al conjunto de datos, una vez que han sido ordenados, en cien partes del mismo tamaño.

$P_5$  separa el 5% más bajo.

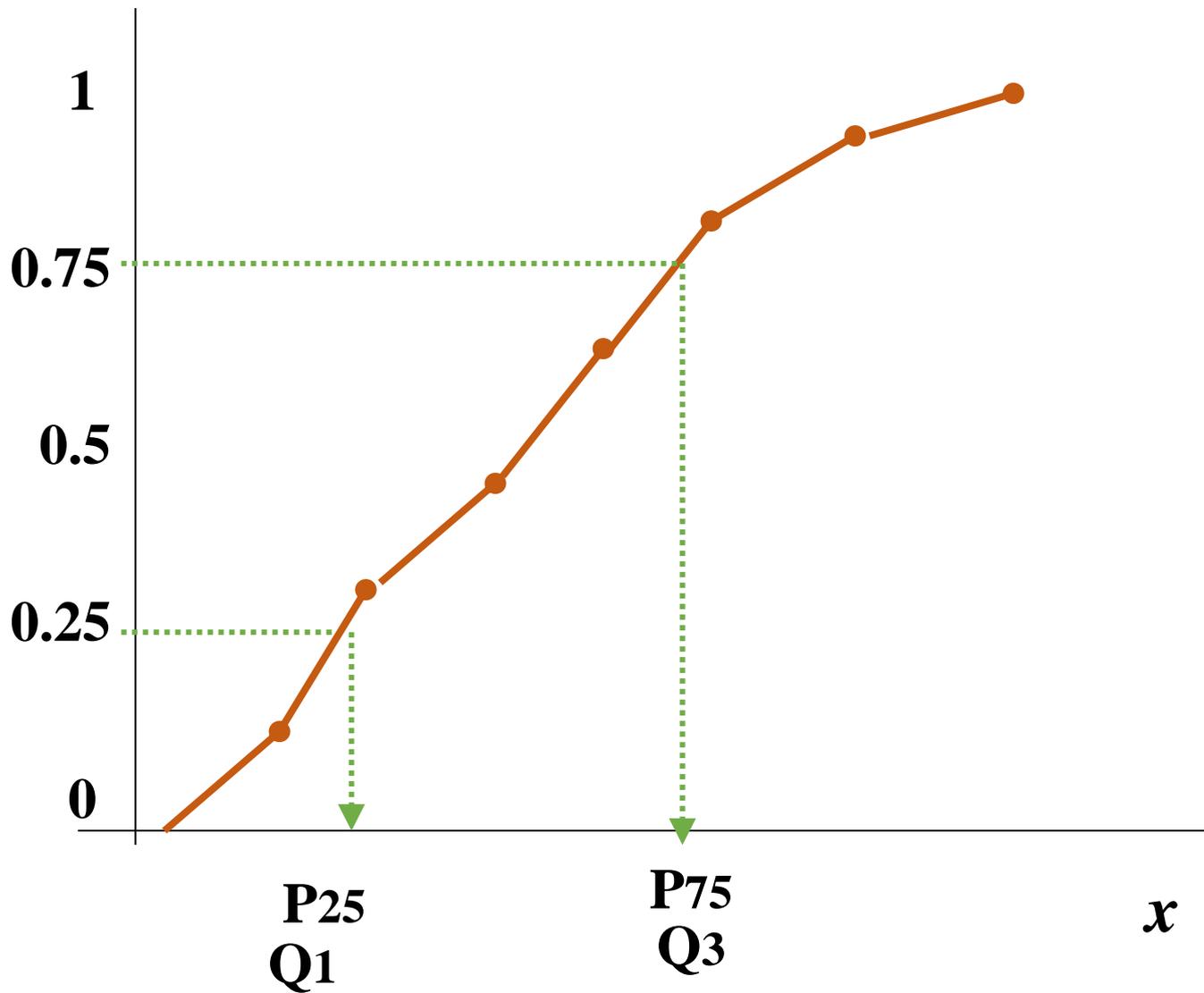
$P_{25}=Q_1$  separa el 25% más bajo.

$P_{50}=Q_2=Me$  separa el 50%.

$P_{75}=Q_3$  separa el 25% más alto.

$P_{90}$  separa el 10% más alto.

*Frecuencia Acumulada Relativa (H)*



# MEDIDAS DE DISPERSIÓN

## *Variable Numérica*

### **MEDIDAS DE DISPERSIÓN**

Los valores medios no logran captar ciertas características de un conjunto de datos. Por ejemplo:

Conjunto 1: 55 – 60 – 65

Conjunto 2: 10 – 60 – 110

Ambos tienen la misma media (60) pero en el primero las observaciones son más homogéneas que las del segundo.

Las medidas de dispersión intentan captar el grado de variabilidad de los datos.

## *Variable Numérica*

### **MEDIDAS DE DISPERSIÓN**

- Rango (R), recorrido o amplitud
- Varianza [ $S^2$  ó  $\sigma^2$  ]
- Desviación Estándar [ $S$  ó  $\sigma$ ]
- Coeficiente de Variación (C.V.)

# Medidas de Dispersión

## **RANGO (R):**

Es la diferencia entre los valores máximo y mínimo del conjunto de datos:

$$R = X_{\text{máx}} - X_{\text{mín}}$$

El Rango del conjunto:

32 – 16 – 18 – 27 – 10 – 25 – 19 – 49 – 28

Es  $R = 49 - 10 = 39$

- El Rango es una medida pobre
- Se emplea frecuentemente en Control de Calidad, con muestras pequeñas.
- Es sensible a valores extremos

# Medidas de Dispersión

## **VARIANZA [ $S^2$ ó $\sigma^2$ ]:**

Mide el grado de dispersión de los datos con referencia a la media aritmética.

Es el promedio (media) de los cuadrados de los desvíos de cada observación respecto de la media aritmética:

$$\text{Varianza Poblacional: } \sigma^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2$$

$$\text{Varianza Muestral: } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{x})^2$$

- Es sensible a valores extremos (alejados de la media).
- Sus unidades son el cuadrado de las de la variable.

# Medidas de Dispersión

## VARIANZA [ $S^2$ ]

Si el siguiente conjunto es una muestra de datos:

32 – 16 – 18 – 27 – 10 – 25 – 19 – 49 – 28

No es práctico calcular la media y luego

$$s^2 = \frac{1}{9-1} [(32 - 24,88)^2 + \dots + (28 - 24,88)^2] = 128,6$$

Expresión de trabajo

$$s^2 = \frac{1}{n-1} \left[ \sum x_i^2 - \frac{(\sum x_i)^2}{n} \right]$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^m f_i (x'_i - \bar{x})^2 = \frac{1}{n-1} \left[ \sum x'_i{}^2 f_i - \frac{(\sum x'_i f_i)^2}{n} \right]$$

### DESVENTAJA:

- Se expresa en unidades correspondientes a la variable, pero al cuadrado

## Propiedades:

**a-** La varianza es una cantidad no negativa  $\text{Var}(x) = 0$ , puesto que se trata de una sumatoria de números positivos porque son cuadrados.  $s_x^2 \geq 0$

**b-** La varianza de una constante es cero, es decir que vale exactamente cero sólo cuando todas las  $x_i$  son iguales, ya que la media coincide con ese valor común y las desviaciones son todas nulas (así como sus cuadrados), resultando  $s_x^2 = 0$ .

Simbólicamente:

$$\text{si } X = k \Rightarrow s^2 = \frac{1}{n-1} \sum_{i=1}^n (k - k)^2 = 0$$

**c-** Si los valores de la variable se multiplican (o dividen) por una constante, cambia la dispersión puesto que al multiplicar el producto depende del valor de la variable a considerar. Los desvíos se verán afectados y como estos se elevan al cuadrado, aparecerá la constante al cuadrado

Simbólicamente:

$$\text{si } y_i = k * x_i \Rightarrow V(Y) = V(k * X) = \frac{1}{n-1} \sum_{i=1}^n (k * x_i - k\bar{x})^2 = k^2 * s_x^2 = k * V(X)$$

## Propiedades:

**d-** Si a una variable se le suma una constante la varianza no cambia, no se altera.

$$\mathbf{Var(x + c) = Var(x)}$$

El efecto de sumar (o restar) a cada valor de la variable una constante es que toda la distribución se traslada hacia la derecha (o izquierda si la constante se resta) una distancia igual a la constante pero sin cambiar la dispersión.

Puede también pensarse en el sentido de que como la varianza de una suma o diferencia de variables es suma de variabilidades (las varianzas **siempre** se suman, se agregan) será

$$\mathbf{Var(x + c) = Var(x) + Var c = Var(x) + 0 = Var(x)}$$

# Medidas de Dispersión

## DESVIACIÓN ESTÁNDAR [S - $\sigma$ ]:

La varianza presenta el inconveniente de quedar expresada en unidades elevadas al cuadrado, que en ocasiones pierde significado.

La Desviación Estándar [S(X)] es la Raíz Cuadrada de la Varianza:

$$S = \sqrt{S^2(X)}$$

En el ejemplo anterior:

$$S = \sqrt{128,6} = 11,3$$

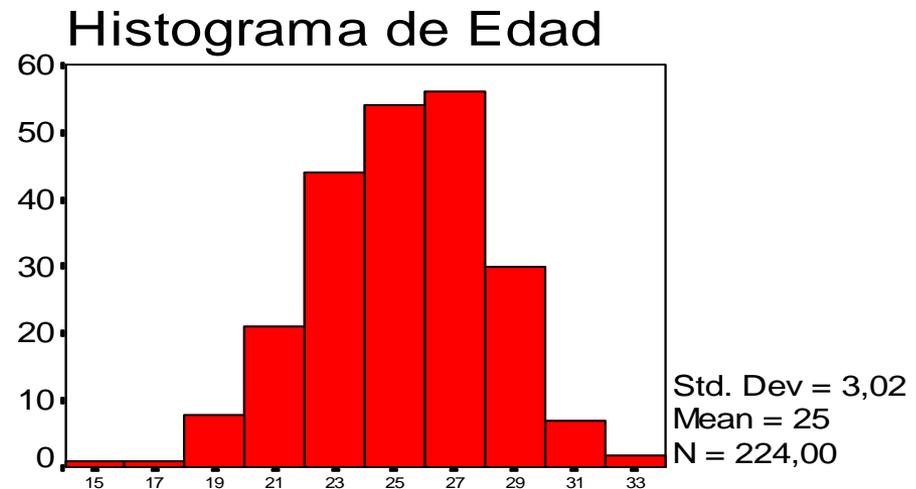
# Medidas de Dispersión

## DESVIACION ESTANDAR:

**Ejemplo:** Se ha registrado la edad en una muestra de individuos:

<i>Edades</i>	<i>Nº de personas</i>
[14 – 16)	1
[16 – 18)	1
[18 – 20)	8
[20 – 22)	21
[22 – 24)	44
[24 – 26)	54
[26 – 28)	56
[28 – 30)	30
[30 – 32)	7
[32 – 34)	2

$$S(X) = 3,03$$



# DESVIACION ESTANDAR:

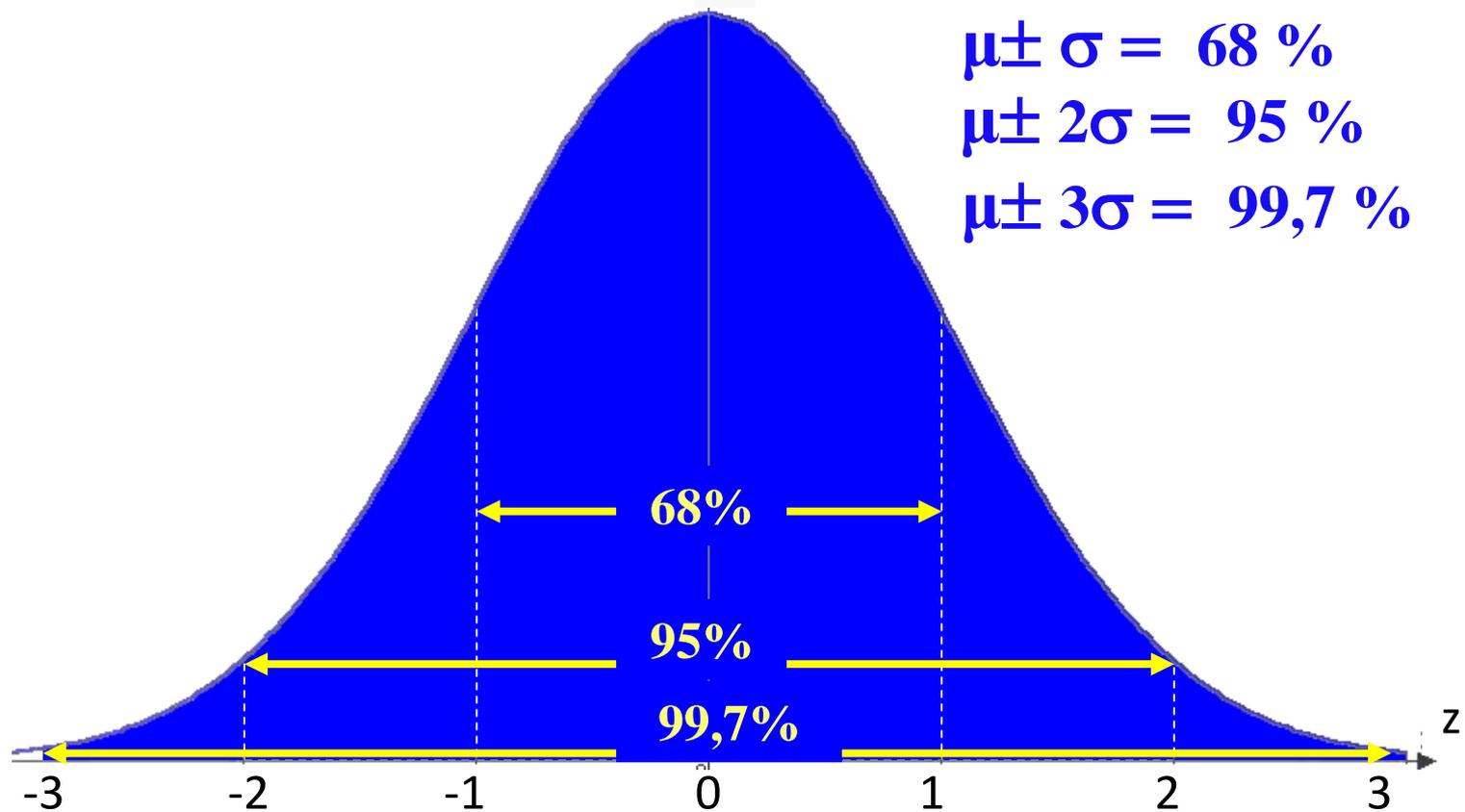
Con este desvío estándar se soluciona el inconveniente de interpretar en términos del problema, puesto que está expresado en las **mismas unidades de medida que la variable original**.

Para tener un significado cuantitativo de la magnitud de esta medida de dispersión será necesario referirse a un tipo de distribución particular que se denomina población o distribución normal. Son más altas en la parte central y bajan con bastante rapidez en sus extremos.

Puede demostrarse que entre el promedio  $\mu$  y una vez la desviación estándar en más y en menos, es decir entre  $(\mu - \sigma)$  y  $(\mu + \sigma)$ , se encuentra alrededor del 68% de la población; entre  $\mu - 2\sigma$  y  $\mu + 2\sigma$  se encuentra cerca del 95% de la población; y entre  $\mu - 3\sigma$  y  $\mu + 3\sigma$  casi la totalidad de la población (el 99,8%).

Cuando se tienen muestras, con "n" suficientemente grande, que hayan sido seleccionadas en forma aleatoria de una población normal o casi normal, el polígono de frecuencias que se obtiene tiene características similares a las de la distribución normal (más similares en tanto n sea mas grande y los intervalos considerados mas chicos a fin de que la línea del polígono se "suavice")

En cualquier distribución normal las probabilidades delimitadas entre la media y



# Medidas de Dispersión

## COEFICIENTE DE VARIACION (CV)

Cuando se desea comparar variabilidades de dos conjuntos distintos, no siempre se puede emplear la Desviación Estándar (¿por qué?).

El Coeficiente de Variación expresa a la Desviación Estándar como un porcentaje de la Media Aritmética:

$$C.V. (\%) = \frac{S(x)}{\bar{x}} \cdot 100$$

El C.V. es ADIMENSIONAL.

# Medidas de Dispersión

## COEFICIENTE DE VARIACION (CV)

### Ejemplo:

Un inversor debe decidirse por la Acción A o por la B de dos compañías de electrónica. ¿Cuál debe elegir si desea optar por la que tiene comportamiento más homogéneo?

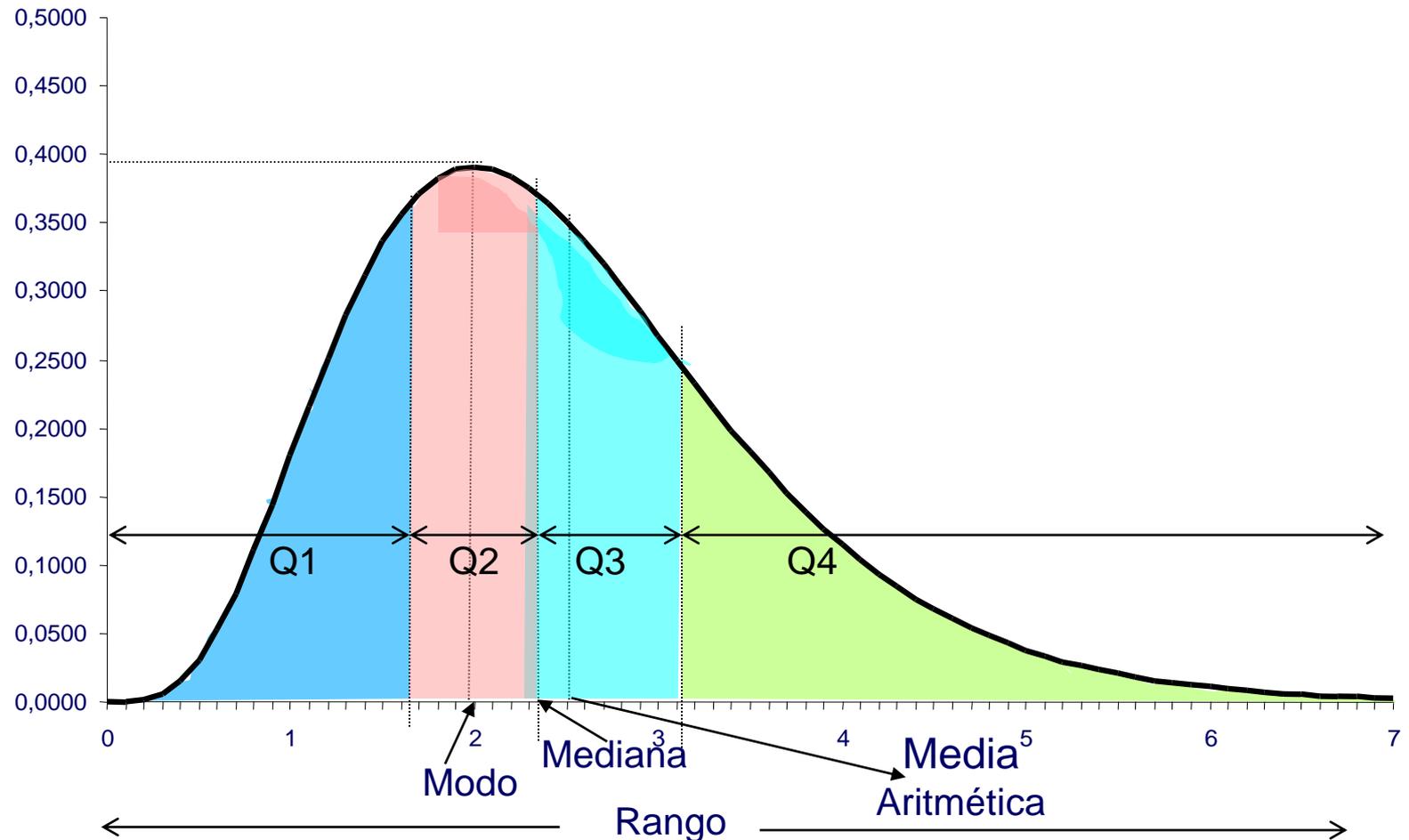
<i>Acción</i>	<i>Precio Promedio (\$)</i>	<i>Desvío Estándar (\$)</i>
<i>A</i>	15.000	500
<i>B</i>	5.000	300

$$CV_A = \frac{500}{15.000} \cdot 100 = 3,3 \%$$

$$CV_B = \frac{300}{5.000} \cdot 100 = 6,0 \%$$

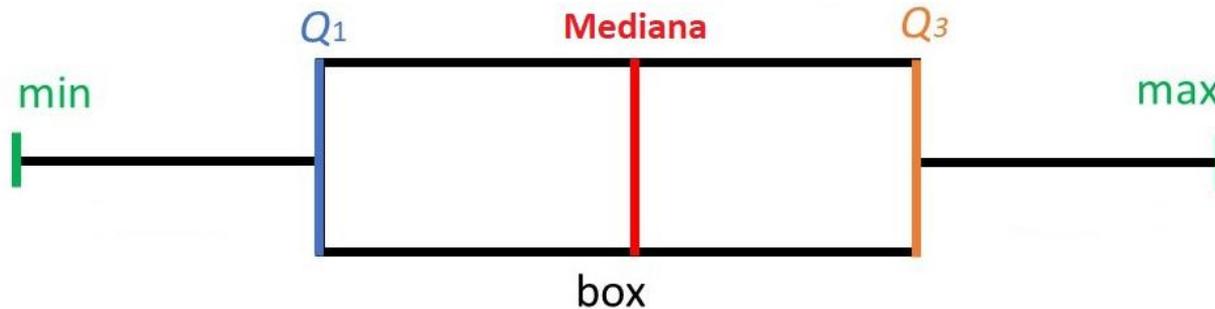
Debe elegir la Acción A.

# Medidas de Tendencia y Dispersión

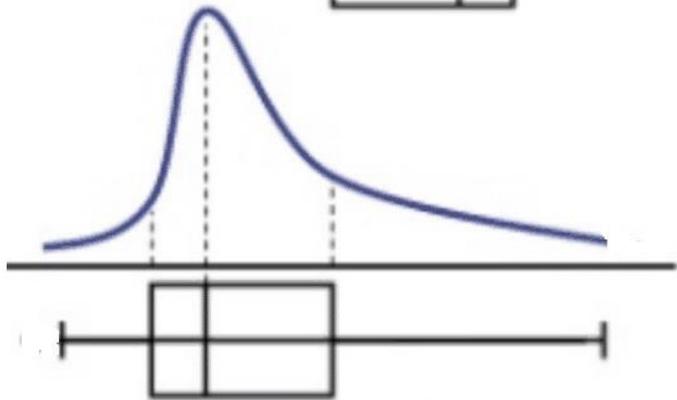
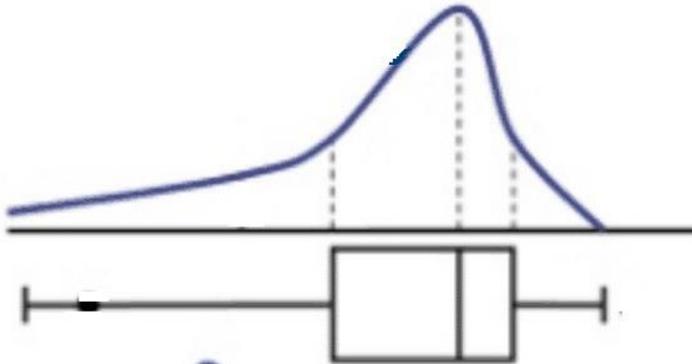
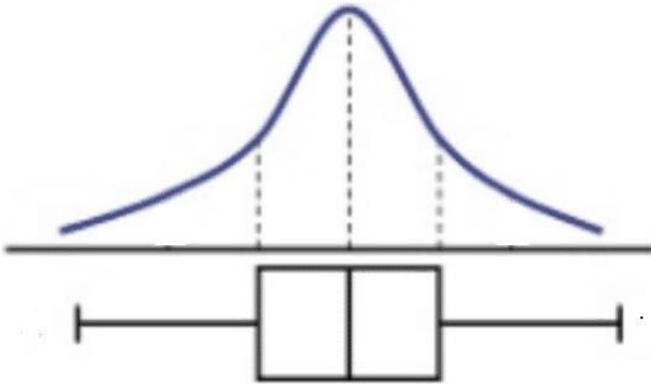


# Diagrama de Cajas

El profesor J.W.Tukey ha estudiado métodos estadísticos para que los "dueños" de los datos aprendan a conocer la información que de esos datos se desprende sin necesidad de tener gran base matemática. Propone el llamado Box-Plot o diagrama de caja.



## DIAGRAMA DE CAJAS Y SIMETRÍA

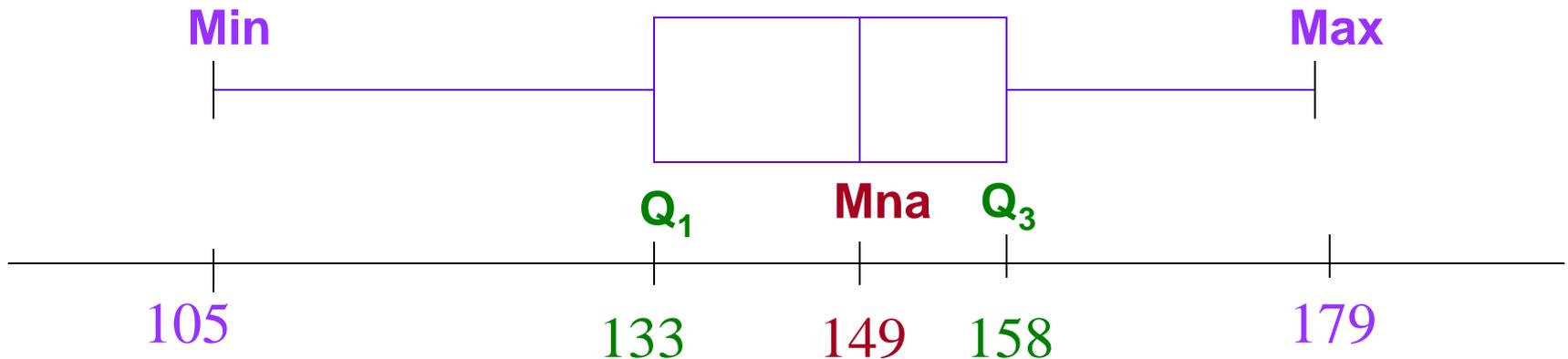


## Tiempo (en minutos) en realizar una tarea

105	110	112	112	118	119	120	120	120
125	126	127	128	130	132	133	134	135
138	138	138	138	141	142	144	145	146
148	148	148	149	150	150	150	151	151
153	153	154	154	154	154	155	156	156
158	160	160	160	163	164	164	165	166
168	168	170	172	172	176	179		

### Un resumen de esta serie en 61 valores

**Min = 105 ; Max = 179 ;  $Q_1 = 133$  ;  $Q_3 = 158$  ;  $Q_2 = Mn = 149$**



(“Box-and-Whisker” plot)

# ¿Qué hemos visto?

- Medidas de Tendencia central: Modo, Mediana y Media Características
- Otras medidas de posición: cuartiles y percentiles
- Medidas de dispersión
  - con unidades: rango,, varianza, desvío estándar
  - sin unidades: coeficiente de variación
  - Diagramas de cajas y simetría

# Gracias

