

# ESTADÍSTICA

Profesora:

María del Valle Valladolid Guijarro

Correo: [mvallado@est-econ.uc3m.es](mailto:mvallado@est-econ.uc3m.es)

Despacho: 7.3.J07 (Leganés)

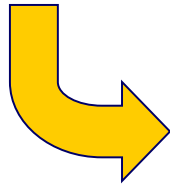
Horario de tutorías: Lunes de 17:00 a 18:00

# Tema 1: Estadística descriptiva

- 1. Introducción. El propósito de la estadística**
- 2. Descripción de datos mediante tablas**
- 3. Descripción de datos mediante gráficos**
- 4. Medidas características de una variable**

# Introducción. El propósito de la estadística

¿Qué es la estadística? ¿Por qué estudiamos estadística?



**Aprender de la observación**



A partir de un número reducido de datos,  
sacar conclusiones que sean generalizables

**Fenómeno  
real**

**Datos  
observados**

**Análisis  
estadístico**

**Aprendizaje  
sobre el fenómeno**

## Dos formas alternativas de acceder al conocimiento



A partir de teorías

---

- Leyes físicas
- Reglas matemáticas
- Propiedades de los materiales ideales



A partir de los modelos  
teóricos DEDUCIMOS la  
realidad

DEDUCIR= Sacar consecuencias de un principio, proposición o supuesto.



A partir de la observación

---

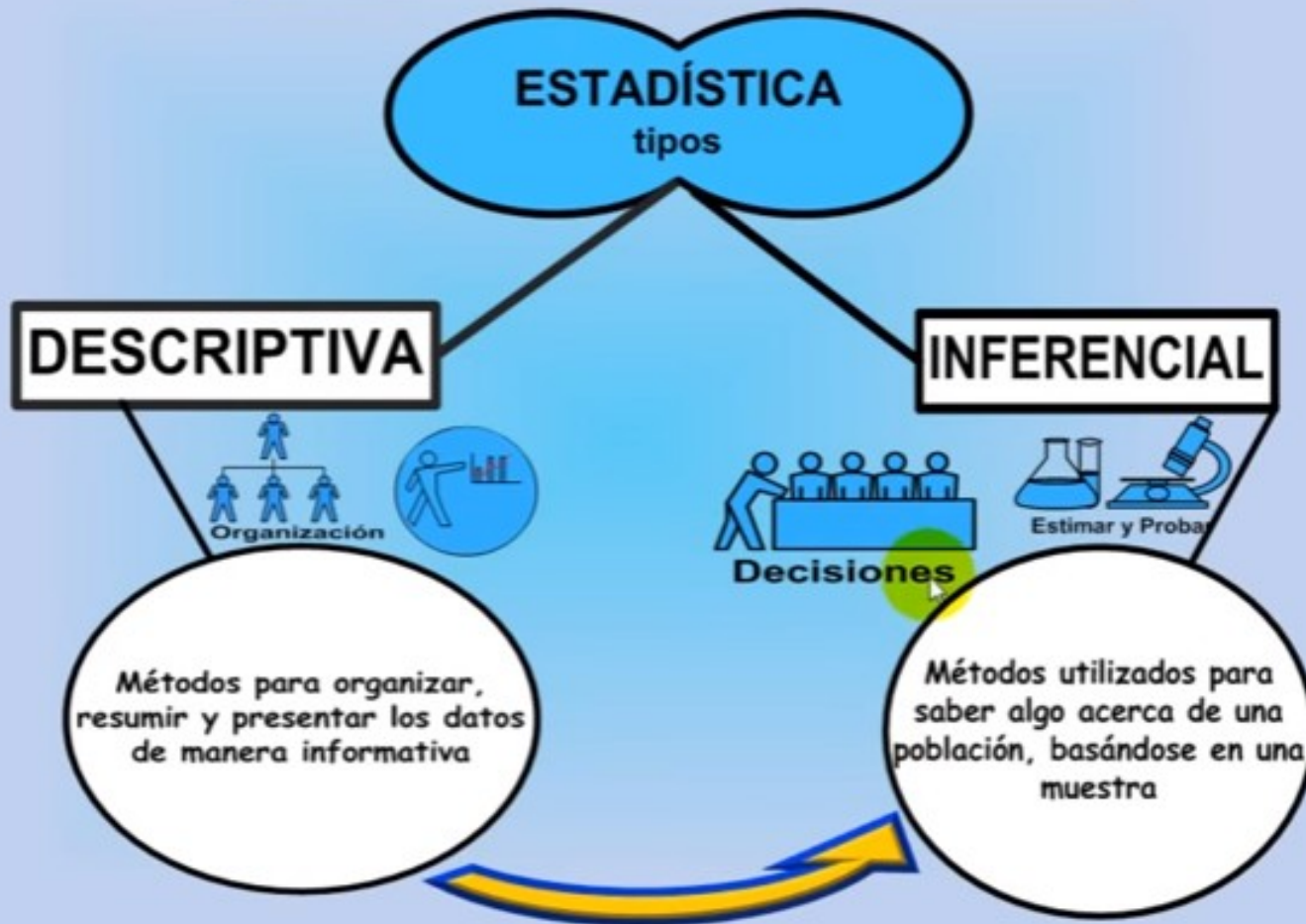
- Datos
- Estadística



A partir de los datos  
INDUCIMOS o INFERIMOS  
un modelo (empírico)

INDUCIR= Extraer, a partir de determinadas observaciones o experiencias particulares, el principio general que en ellas está implícito.

# Clasificación de la Estadística



## Población y muestra.

**Población** es el conjunto de individuos, con alguna característica común, sobre el que se hace un estudio estadístico.

La **muestra** es un subconjunto de la población, seleccionada de modo que ponga de manifiesto las características de la misma, de ahí que la propiedad más importante de las muestras es su representatividad.

El proceso seguido en la extracción de la muestra se llama **muestreo**



## Variables estadísticas

La característica a estudiar en una población es la **variable estadística**.

Las variables estadísticas pueden ser esencialmente de dos tipos **cualitativas y cuantitativas**.

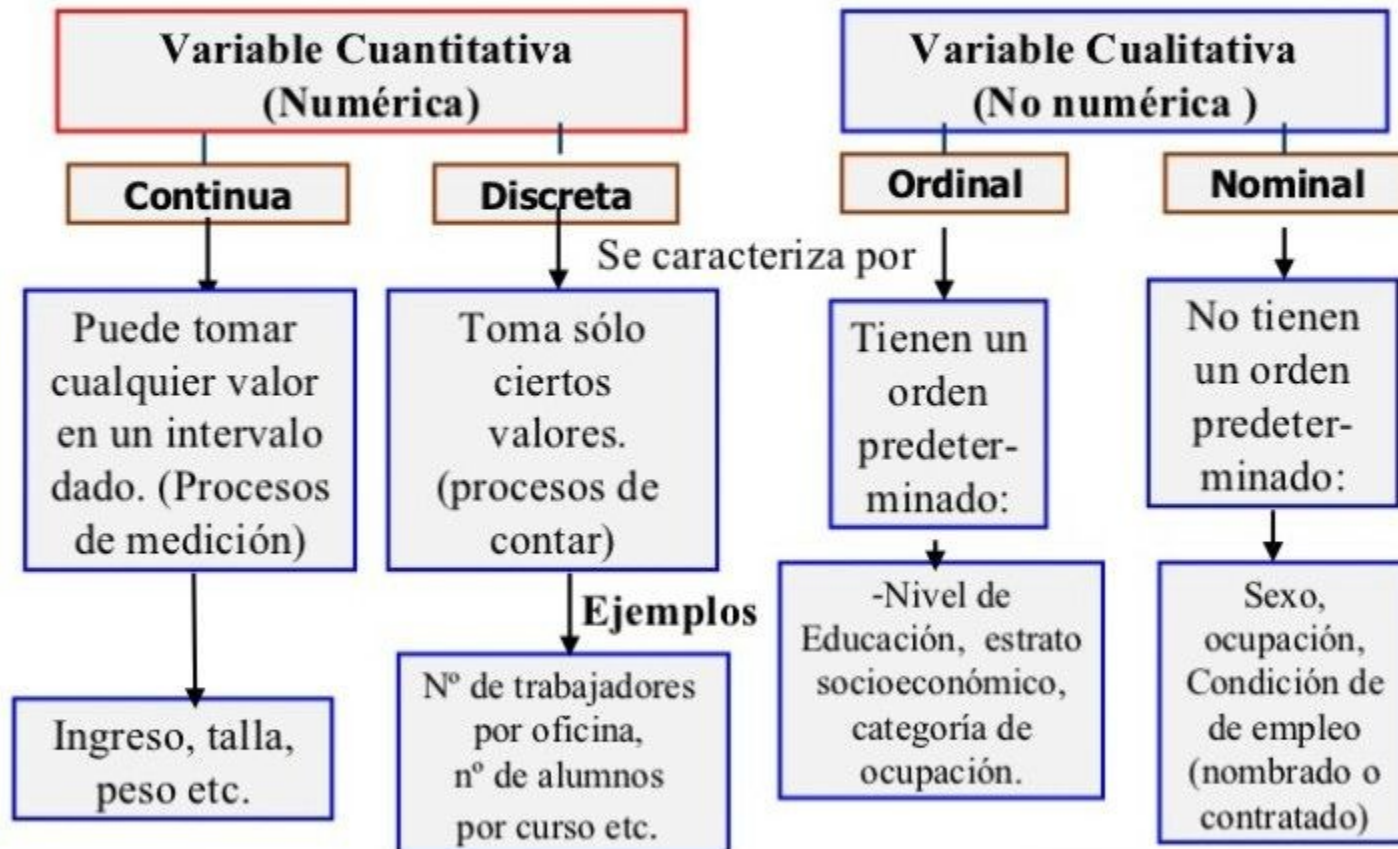
Las variables cualitativas son las que no aparecen en forma numérica sino como una categoría o atributo.

Las variables cuantitativas son las que pueden expresarse numéricamente, y a su vez pueden ser:

- ✓ **Cuantitativas discretas**, si sólo pueden tomar un número finito de valores.

- ✓ **Cuantitativas continuas** cuando pueden tomar cualquier valor de un intervalo.

# Clasificación de Variables





# Tema 1: Estadística descriptiva

1. Introducción. El propósito de la estadística
2. Descripción de datos mediante tablas
3. Descripción de datos mediante gráficos
4. Medidas características de una variable

# Descripción de datos mediante tablas

Objetivo: Resumir la información para facilitar su análisis

## Tablas univariantes

Muestran la frecuencia de cada valor observado

Ejemplo 1: Número de cilindros de 155 vehículos (fichero cardata.xls)

Class	Value	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
1	3	1	0,0065	1	0,0065
2	4	104	0,6710	105	0,6774
3	5	3	0,0194	108	0,6968
4	6	30	0,1935	138	0,8903
5	8	17	0,1097	155	1,0000

## Tablas univariantes

### Tipos de frecuencias

- **Frecuencias absolutas:** son el número de veces que se repite un número en un conjunto de datos.
- **Frecuencias absolutas acumuladas:** es la suma de las frecuencias absolutas.
- **Frecuencia relativa:** corresponde a las veces que se repite un número en un conjunto de datos respecto al total, pero se expresa en porcentajes (%).
- **Frecuencia relativa acumulada:** es la suma de las frecuencias relativas.

## Tablas univariantes

Ejemplo 2: Mes de nacimiento de 95 estudiantes de primer curso

Class	Value	Frequency	Relative Frequency
1	Enero	15	0,1579
2	Febrero	5	0,0526
3	Marzo	10	0,1053
4	Abril	9	0,0947
5	Mayo	10	0,1053
6	Junio	13	0,1368
7	Julio	9	0,0947
8	Agosto	7	0,0737
9	Septiembre	6	0,0632
10	Octubre	1	0,0105
11	Noviembre	3	0,0316
12	Diciembre	7	0,0737

# Descripción de datos mediante tablas

## Tablas univariantes

Si hay muchos valores distintos: se agrupan en intervalos –clases–

Ejemplo: Precio de 155 vehículos (fichero cardata.xls)

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		0,0		0	0,0000	0	0,0000
1	0,0	2000,0	1000,0	1	0,0065	1	0,0065
2	2000,0	4000,0	3000,0	70	0,4516	71	0,4581
3	4000,0	6000,0	5000,0	60	0,3871	131	0,8452
4	6000,0	8000,0	7000,0	14	0,0903	145	0,9355
5	8000,0	10000,0	9000,0	8	0,0516	153	0,9871
6	10000,0	12000,0	11000,0	0	0,0000	153	0,9871
7	12000,0	14000,0	13000,0	0	0,0000	153	0,9871
8	14000,0	16000,0	15000,0	2	0,0129	155	1,0000
9	16000,0	18000,0	17000,0	0	0,0000	155	1,0000
above	18000,0			0	0,0000	155	1,0000

Número de clases?  $\longrightarrow \sqrt{n}$

# Tema 1: Estadística descriptiva

1. Introducción. El propósito de la estadística
2. Descripción de datos mediante tablas
3. Descripción de datos mediante gráficos
4. Medidas características de una variable

3.1 Diagrama de barras

3.2 Histograma y polígono de frecuencias

3.3 Diagrama de tarta

3.4 Series temporales

## 3.1 Diagrama de barras

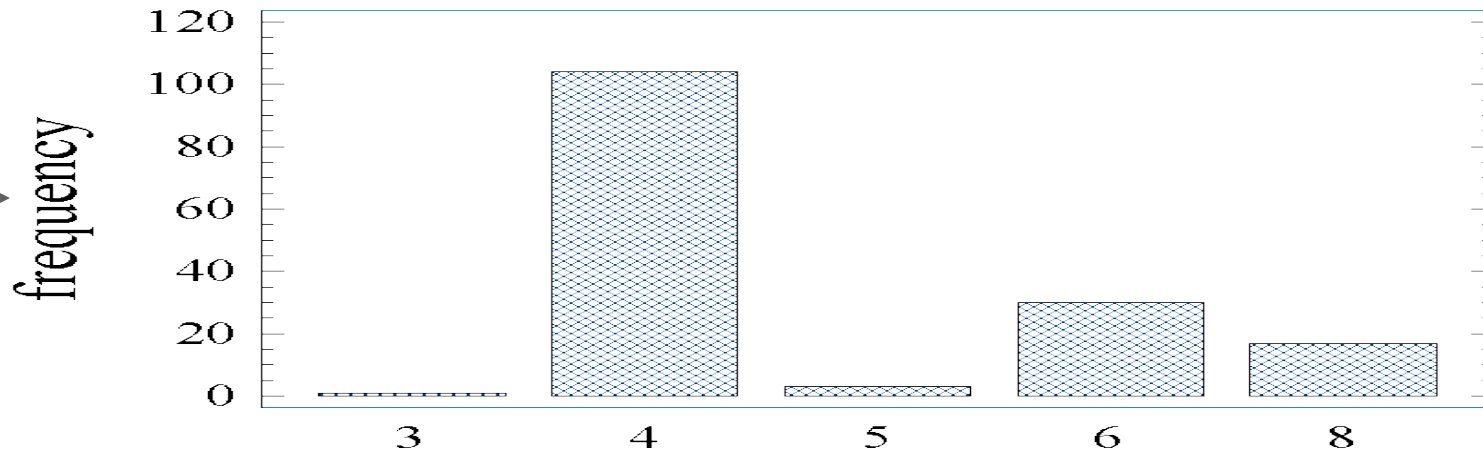
Es la representación gráfica de una tabla de frecuencias sin agrupar en intervalos

Frecuencia de cada valor observado

Ejemplo: Número de cilindros de 155 vehículos (fichero cardata.xls)

Class	Value	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
1	3	1	0,0065	1	0,0065
2	4	104	0,6710	105	0,6774
3	5	3	0,0194	108	0,6968
4	6	30	0,1935	138	0,8903
5	8	17	0,1097	155	1,0000

**Barchart for cylinders**





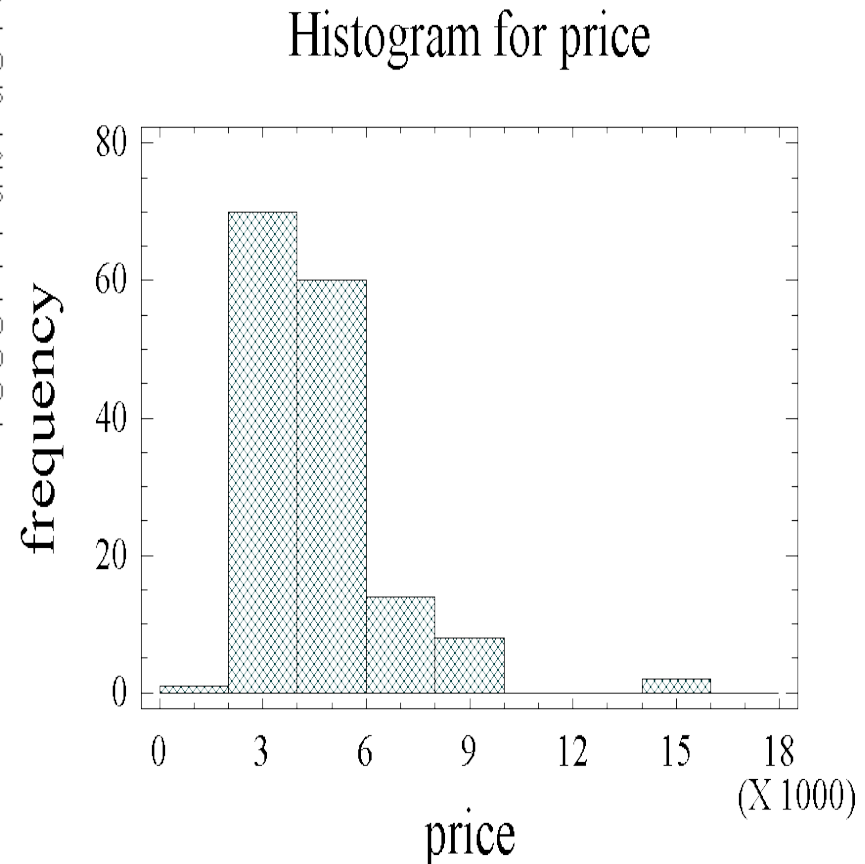
## 3.2 Histograma y polígono de frecuencias

Es la representación gráfica de una tabla de frecuencias con datos agrupados

Ejemplo: Precio de 155 vehículos (archivo cardata.xls)

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		0,0		0	0,0000	0	0,0000
1	0,0	2000,0	1000,0	1	0,0065	1	0,0065
2	2000,0	4000,0	3000,0	70	0,4516	71	0,4581
3	4000,0	6000,0	5000,0	60	0,3871	131	0,8452
4	6000,0	8000,0	7000,0	14	0,0903	145	0,9355
5	8000,0	10000,0	9000,0	8	0,0516	153	0,9871
6	10000,0	12000,0	11000,0	0	0,0000	153	0,9871
7	12000,0	14000,0	13000,0	0	0,0000	153	0,9871
8	14000,0	16000,0	15000,0	2	0,0129	155	1,0000
9	16000,0	18000,0	17000,0	0	0,0000	155	1,0000
above	18000,0			0	0,0000	155	1,0000

El histograma es una de las herramientas gráficas más útiles para resumir información



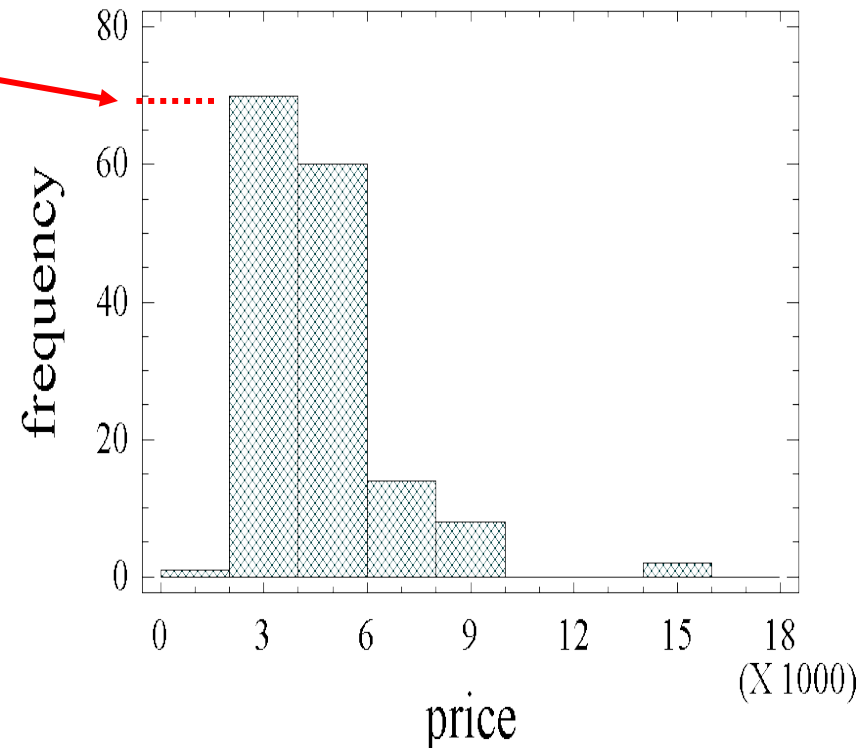
## 3.2 Histograma y polígono de frecuencias

Es la representación gráfica de una tabla de frecuencias con datos agrupados

Ejemplo: Precio de 155 vehículos (archivo cardata.xls)

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		0,0		0	0,0000	0	0,0000
1	0,0	2000,0	1000,0	1	0,0065	1	0,0065
2	2000,0	4000,0	3000,0	70	0,4516	71	0,4581
3	4000,0	6000,0	5000,0	60	0,3871	131	0,8452
4	6000,0	8000,0	7000,0	14	0,0903	145	0,9355
5	8000,0	10000,0	9000,0	8	0,0516	153	0,9871
6	10000,0	12000,0	11000,0	0	0,0000	153	0,9871
7	12000,0	14000,0	13000,0	0	0,0000	153	0,9871
8	14000,0	16000,0	15000,0	2	0,0129	155	1,0000
9	16000,0	18000,0	17000,0	0	0,0000	155	1,0000
above	18000,0			0	0,0000	155	1,0000

Histogram for price



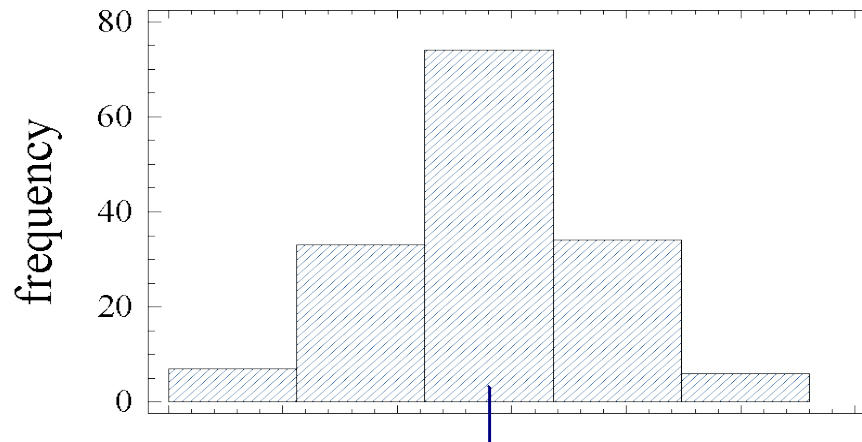
El histograma sirve para resumir la siguiente información:

- Concentraciones
- Asimetrías
- Huecos
- Atípicos

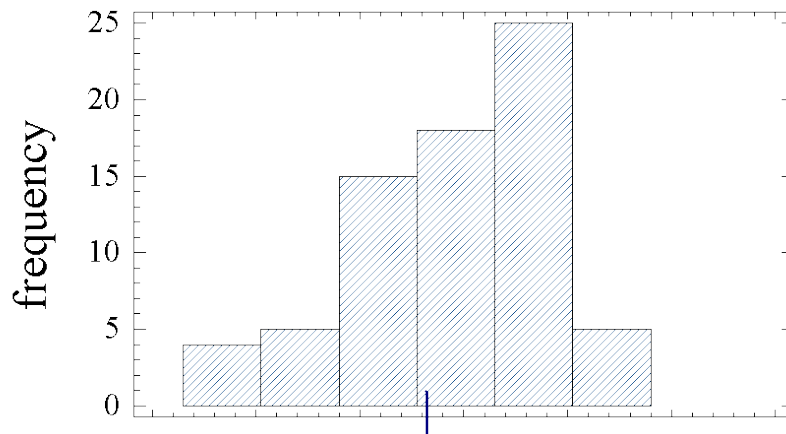
## 3.2 Histograma y polígono de frecuencias

Es la representación gráfica de una tabla de frecuencias con datos agrupados

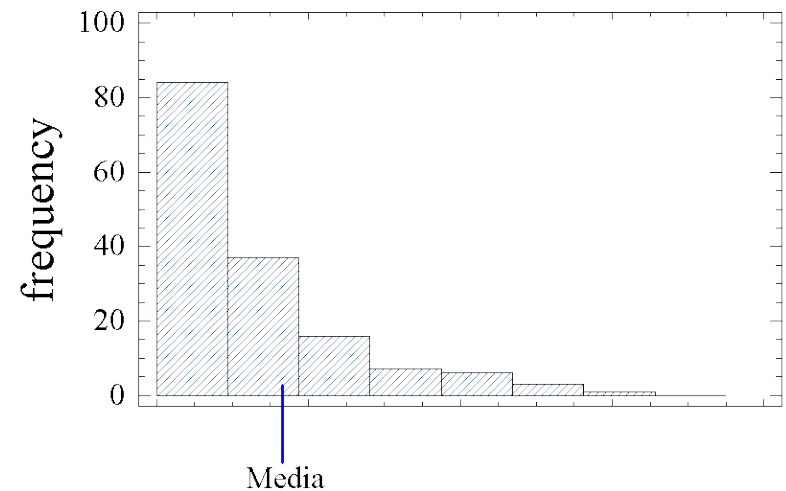
Distribución simétrica



Distribución asimétrica negativa

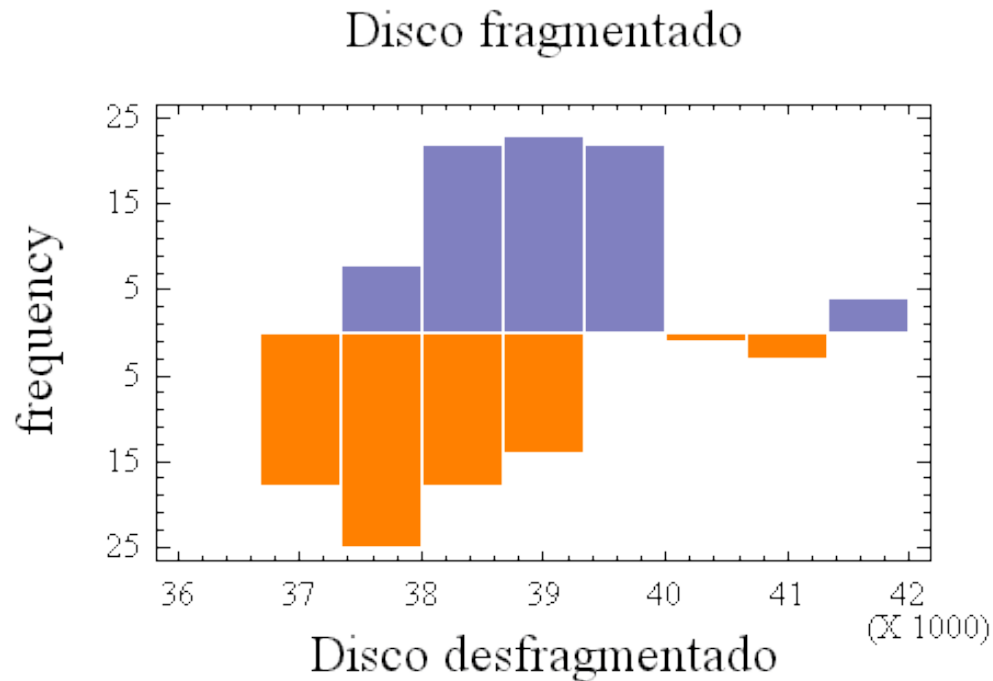


Distribución asimétrica positiva



### Ejemplo:

Tiempo que un ordenador tarda en escribir un fichero de 300 Mb en su disco duro. Se hacen dos experimentos; uno en el que el disco duro está desfragmentado, y otro en el que el disco duro tiene una fragmentación del 40%. Cada experimento se repite 79 veces.

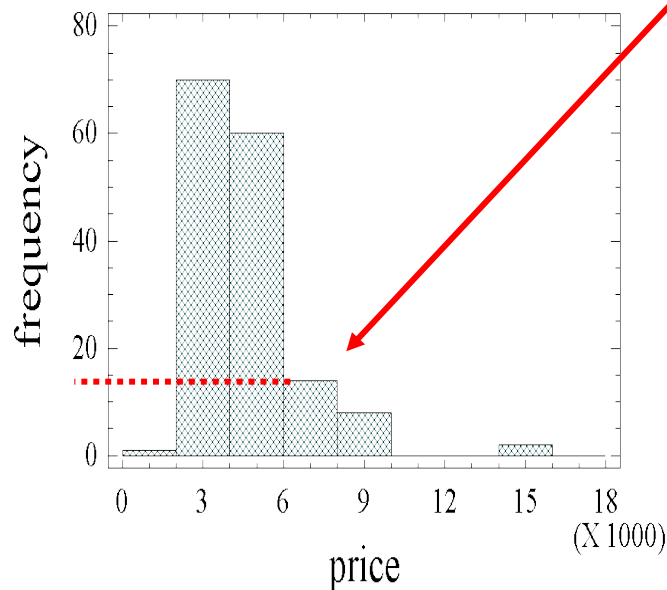


El histograma también puede hacerse con las frecuencias acumuladas.

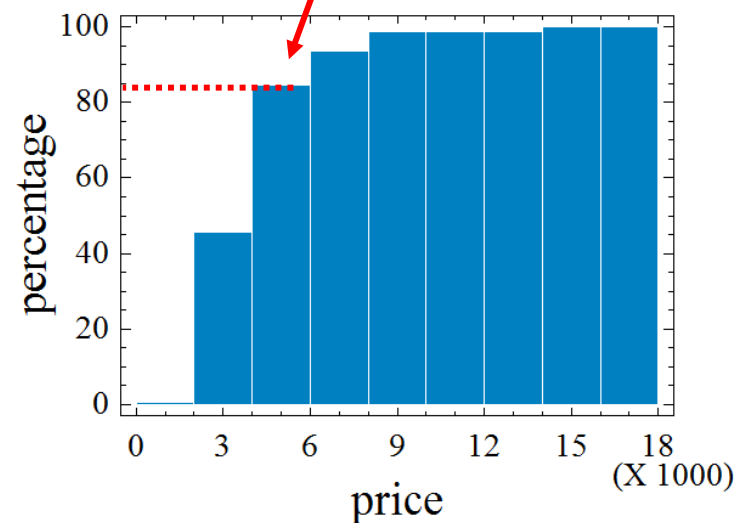
También puede expresarse en frecuencias relativas o absolutas

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		0,0		0	0,0000	0	0,0000
1	0,0	2000,0	1000,0	1	0,0065	1	0,0065
2	2000,0	4000,0	3000,0	70	0,4516	71	0,4581
3	4000,0	6000,0	5000,0	60	0,3871	131	0,8452
4	6000,0	8000,0	7000,0	14	0,0903	145	0,9355
5	8000,0	10000,0	9000,0	8	0,0516	153	0,9871
6	10000,0	12000,0	11000,0	0	0,0000	153	0,9871
7	12000,0	14000,0	13000,0	0	0,0000	153	0,9871
8	14000,0	16000,0	15000,0	2	0,0129	155	1,0000
9	16000,0	18000,0	17000,0	0	0,0000	155	1,0000
above	18000,0			0	0,0000	155	1,0000

Histogram for price



Histogram for price



## 3.2 Histograma y polígono de frecuencias

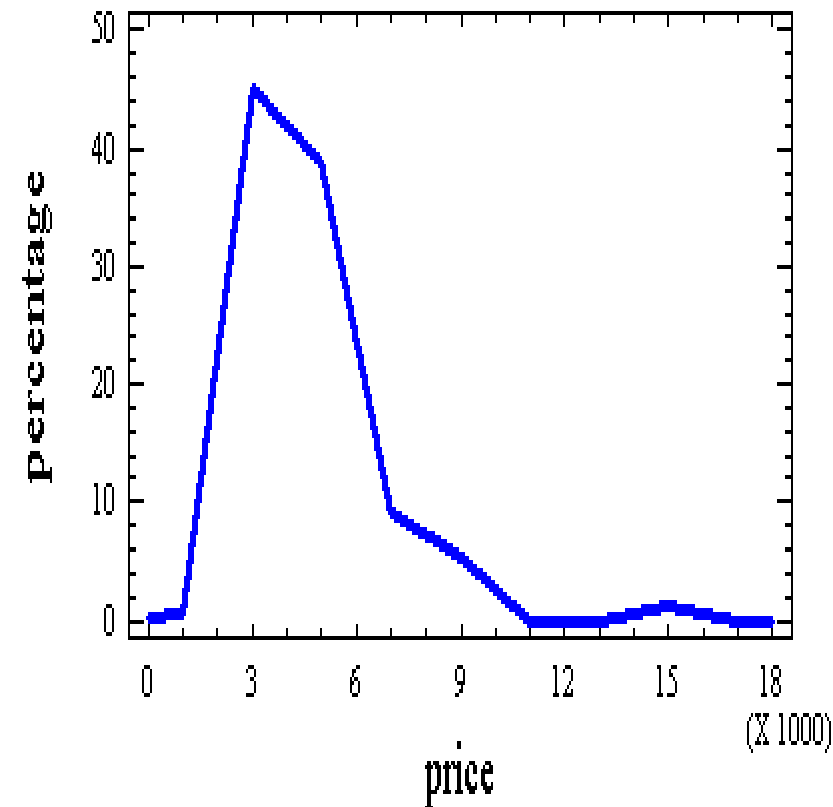
Es la representación gráfica de una tabla de frecuencias con datos agrupados

Ejemplo: Precio de 155 vehículos (archivo cardata.xls)

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		0,0		0	0,0000	0	0,0000
1	0,0	2000,0	1000,0	1	0,0065	1	0,0065
2	2000,0	4000,0	3000,0	70	0,4516	71	0,4581
3	4000,0	6000,0	5000,0	60	0,3871	131	0,8452
4	6000,0	8000,0	7000,0	14	0,0903	145	0,9355
5	8000,0	10000,0	9000,0	8	0,0516	153	0,9871
6	10000,0	12000,0	11000,0	0	0,0000	153	0,9871
7	12000,0	14000,0	13000,0	0	0,0000	153	0,9871
8	14000,0	16000,0	15000,0	2	0,0129	155	1,0000
9	16000,0	18000,0	17000,0	0	0,0000	155	1,0000
above	18000,0			0	0,0000	155	1,0000

El polígono de frecuencias se consigue uniendo con líneas los puntos centrales de la parte superior del histograma.

Histogram for price



## 3.2 Histograma y polígono de frecuencias

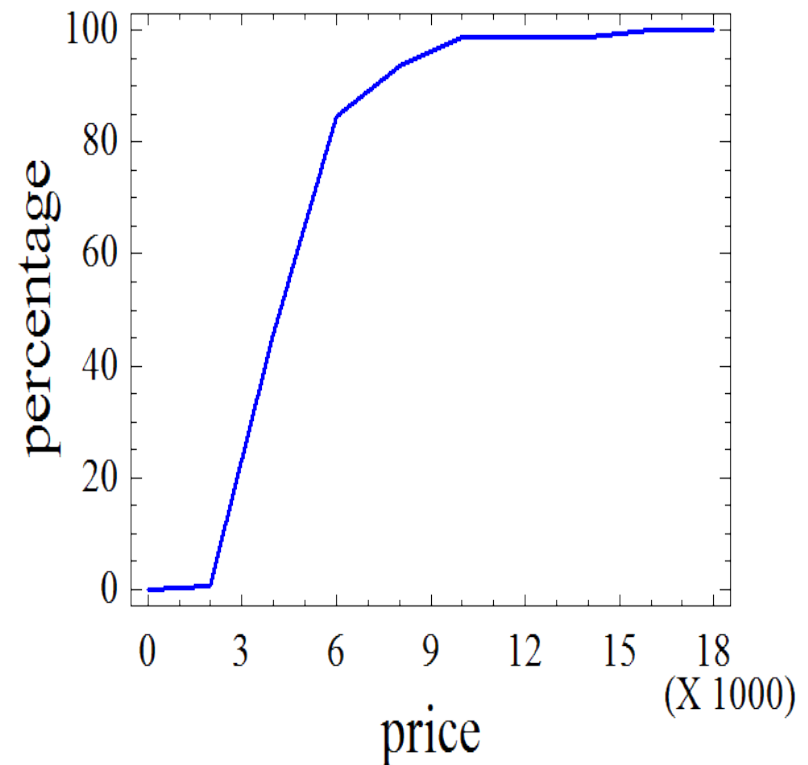
Es la representación gráfica de una tabla de frecuencias con datos agrupados

Ejemplo: Precio de 155 vehículos (fichero cardata.xls)

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		0,0		0	0,0000	0	0,0000
1	0,0	2000,0	1000,0	1	0,0065	1	0,0065
2	2000,0	4000,0	3000,0	70	0,4516	71	0,4581
3	4000,0	6000,0	5000,0	60	0,3871	131	0,8452
4	6000,0	8000,0	7000,0	14	0,0903	145	0,9355
5	8000,0	10000,0	9000,0	8	0,0516	153	0,9871
6	10000,0	12000,0	11000,0	0	0,0000	153	0,9871
7	12000,0	14000,0	13000,0	0	0,0000	153	0,9871
8	14000,0	16000,0	15000,0	2	0,0129	155	1,0000
9	16000,0	18000,0	17000,0	0	0,0000	155	1,0000
above	18000,0			0	0,0000	155	1,0000

Puede realizarse también con las frecuencias acumuladas

Histogram for price



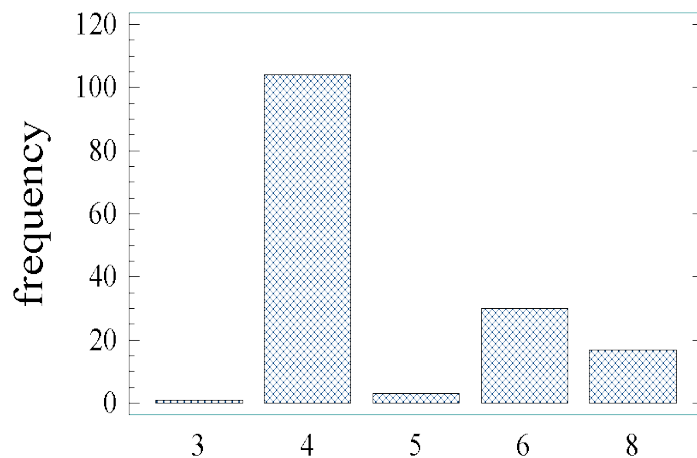
### 3.3 Diagrama de tarta

Es un círculo dividido en porciones proporcionales a las frecuencias relativas

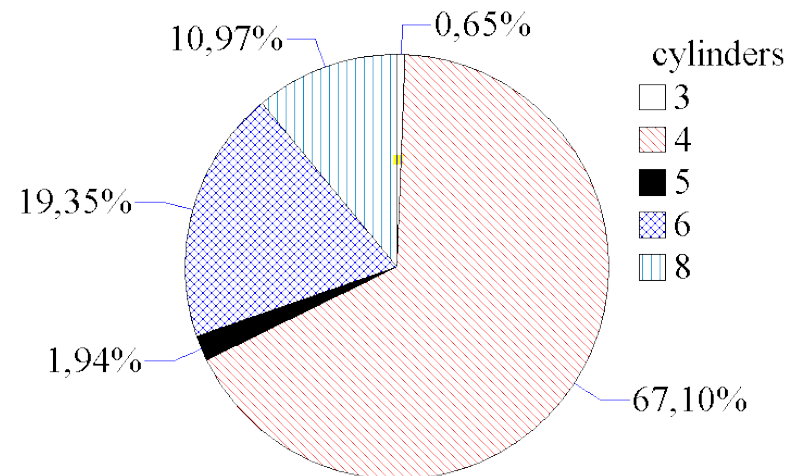
Ejemplo: Número de cilindros de 155 vehículos (fichero cardata.xls)

Class	Value	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
1	3	1	0,0065	1	0,0065
2	4	104	0,6710	105	0,6774
3	5	3	0,0194	108	0,6968
4	6	30	0,1935	138	0,8903
5	8	17	0,1097	155	1,0000

Barchart for cylinders



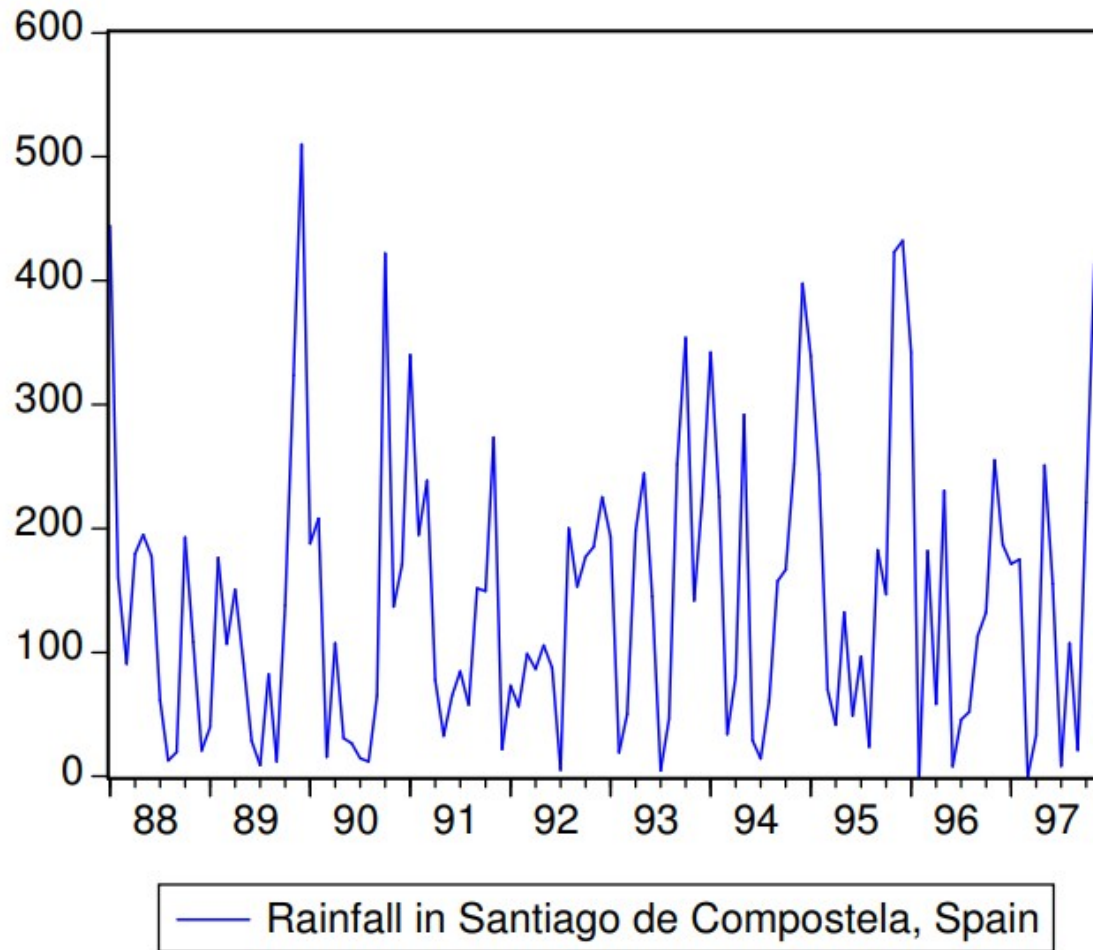
Piechart for cylinders





### 3.4 Series temporales

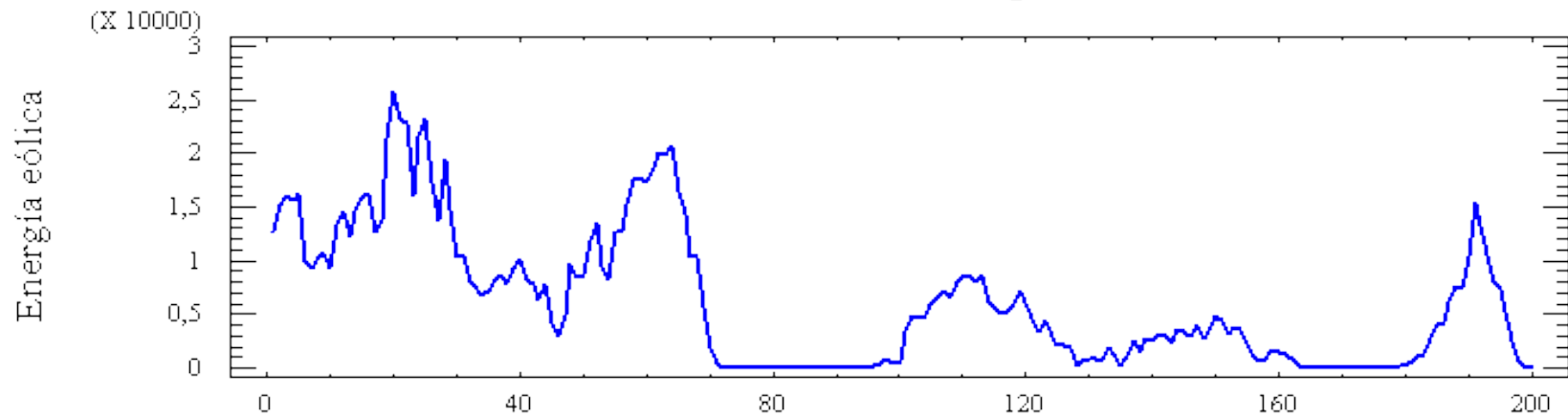
El eje X es el tiempo. Representa la evolución temporal de la variable



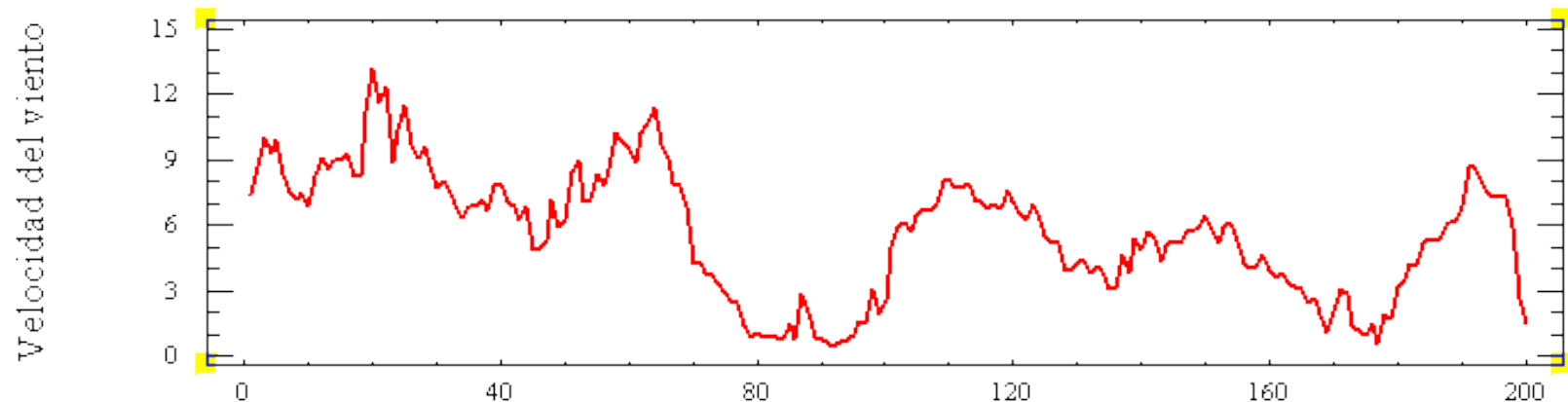
### 3.4 Series temporales

El eje X es el tiempo. Representa la evolución temporal de la variable

Time Series Plot for Energía eólica



Time Series Plot for Velocidad del viento



# Tema 1: Estadística descriptiva

1. Introducción. El propósito de la estadística
2. Descripción de datos mediante tablas
3. Descripción de datos mediante gráficos
4. Medidas características de una variable

# Medidas características de un conjunto de datos

Objetivo: Buscamos resumir las características más importantes de los datos en un conjunto reducido de números.

Cada característica  un número

## 4.1 Medidas de centralización

¿Cuál es el centro de los datos?

Hay muchas medidas alternativas.

Las más importantes son

- Media aritmética
- Mediana
- Moda

## 4.1 Medidas de centralización

- Media aritmética

Sea un conjunto de observaciones  $x_1, x_2, \dots, x_n$

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

Si hay  $J$  valores diferentes que se repiten:

$x_1$ , se repite  $n_1$  veces

$x_2$ , se repite  $n_2$  veces

...

$x_J$ , se repite  $n_J$  veces

$$\bar{x} = \sum_{j=1}^J x_j \text{ fr}(x_j)$$

Donde  $\text{fr}(x_j)$  es la frecuencia relativa del valor  $x_j$

## 4.1 Medidas de centralización

- Media aritmética

Ejemplo:  $x=\{1,2,3,3,5,5,5,6,6\}$

$$\bar{x} = \frac{1+2+3+3+5+5+5+6+6}{9} = 4$$

O bien:

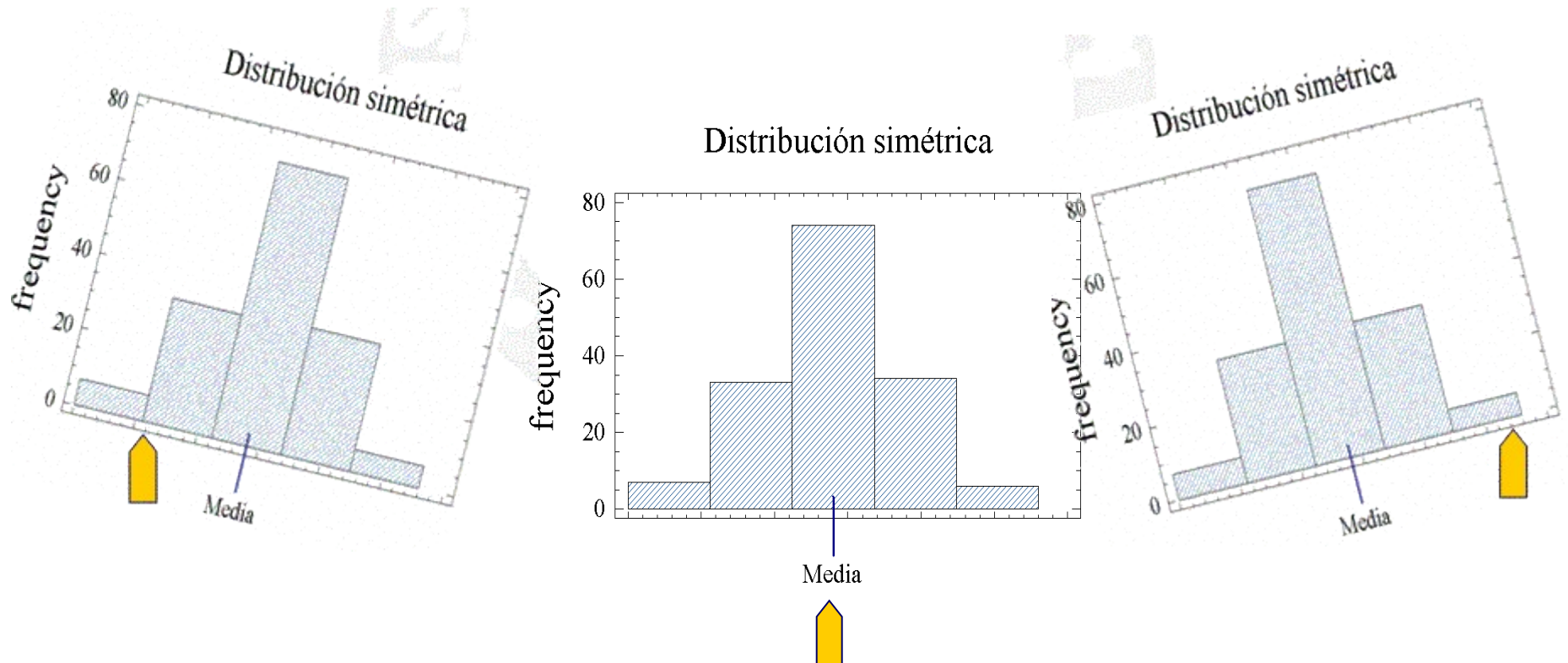
$$\bar{x} = \sum_{j=1}^J x_j \text{fr}(x_j) \quad \longrightarrow \quad \bar{x} = 1 \times \frac{1}{9} + 2 \times \frac{1}{9} + 3 \times \frac{2}{9} + 5 \times \frac{3}{9} + 6 \times \frac{1}{9} = 4$$

## 4.1 Medidas de centralización

- Media aritmética

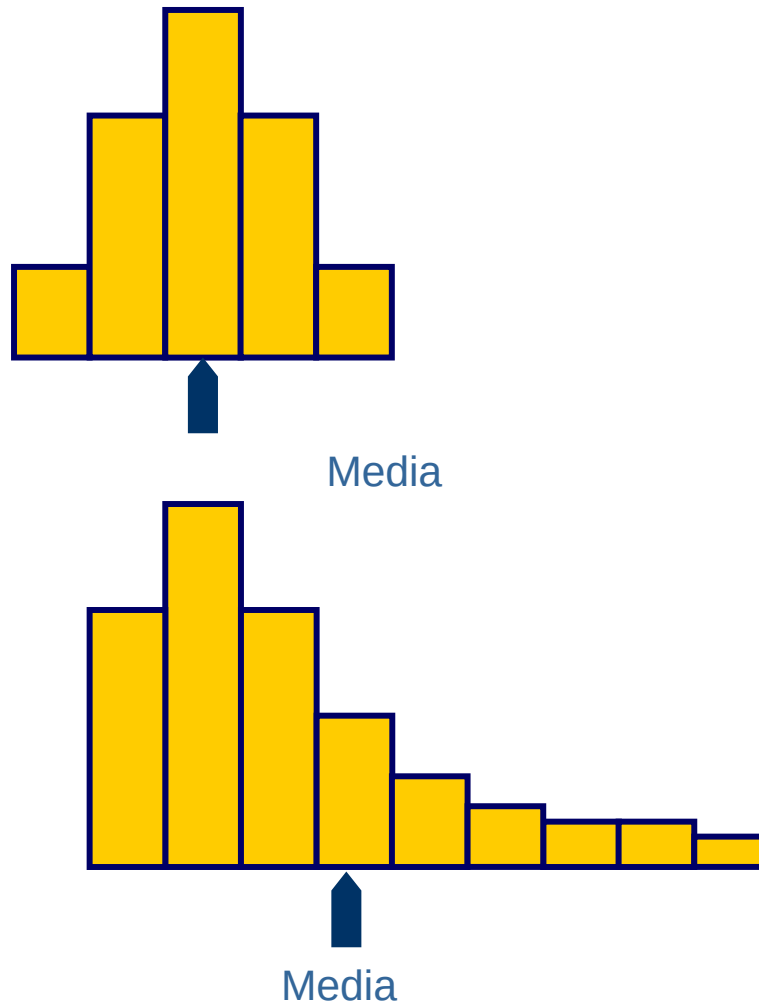
Puede interpretarse como el centro de gravedad de los datos.

Por ejemplo, en un histograma, es el punto de apoyo para que esté en equilibrio



- Media aritmética

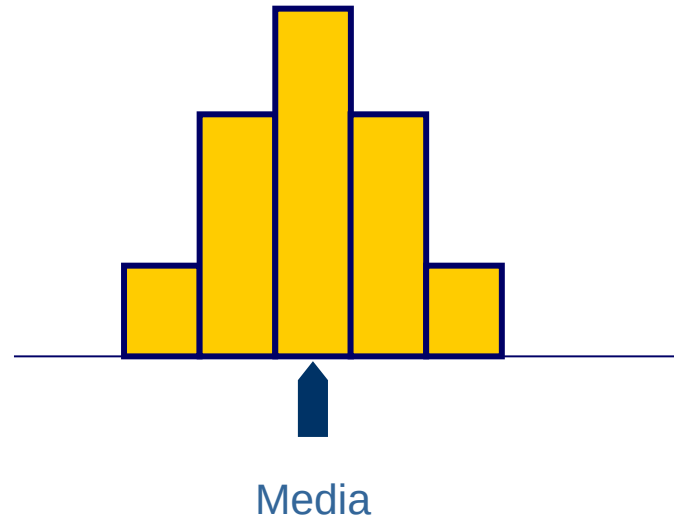
Cuanto más asimétrica sea más se desplaza la media hacia la cola





- Media aritmética

Es muy sensible a valores atípicos



Un solo dato puede desplazar la media hasta hacerla poco representativa del centro de los datos



- Mediana

Es el valor que deja a cada lado el 50% de los datos.

Es poco sensible a asimetrías

Es insensible a valores atípicos

1 2 5 8 11 13 24 28 31



Mediana=11

9 datos

Con un número impar de datos: el dato en posición central

1 2 3 5 8 11 13 24 28 31



Mediana=(8+11)/2=9,5

10 datos

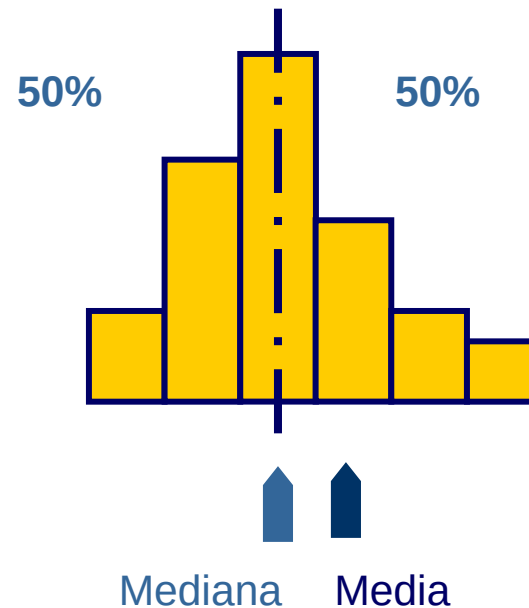
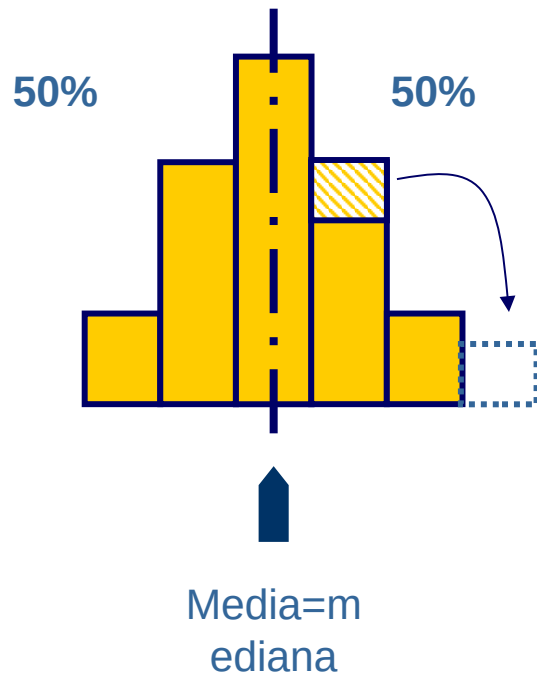
Con un número par de datos: la media de los dos centrales

- Mediana

Es el valor que deja a cada lado el 50% de los datos.

Es poco sensible a asimetrías

Es insensible a valores atípicos



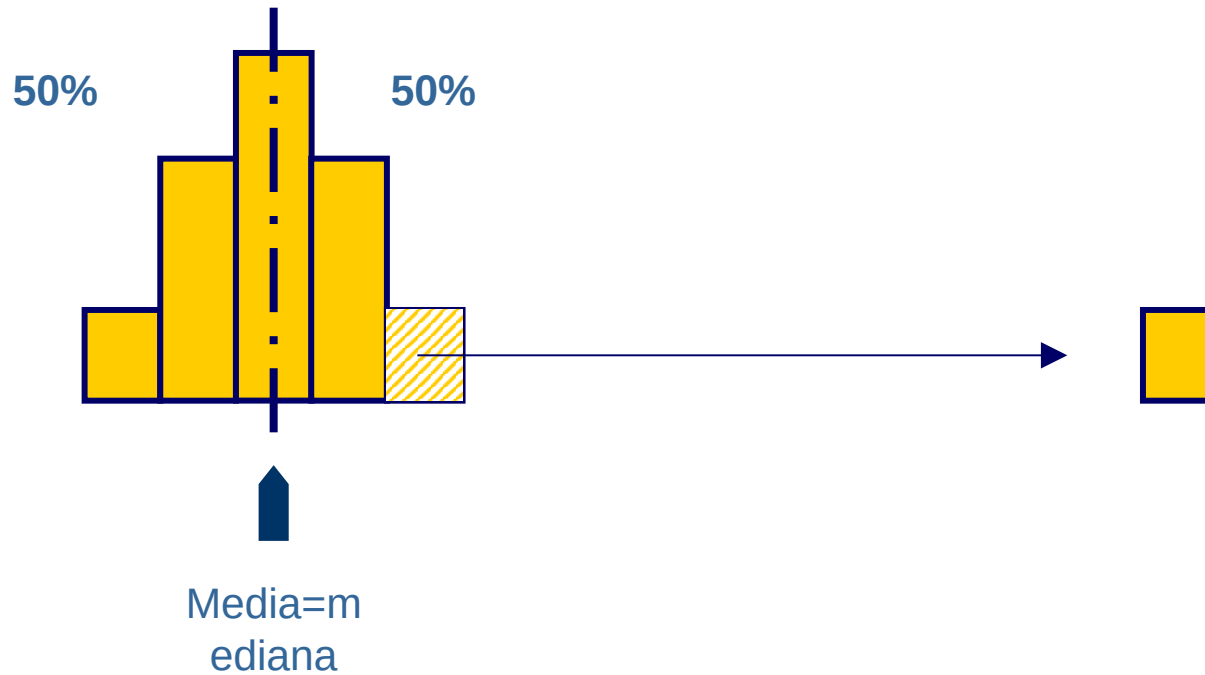
La mediana no varía, pero la media se desplaza

- Mediana

Es el valor que deja a cada lado el 50% de los datos.

Es poco sensible a asimetrías

Es insensible a valores atípicos

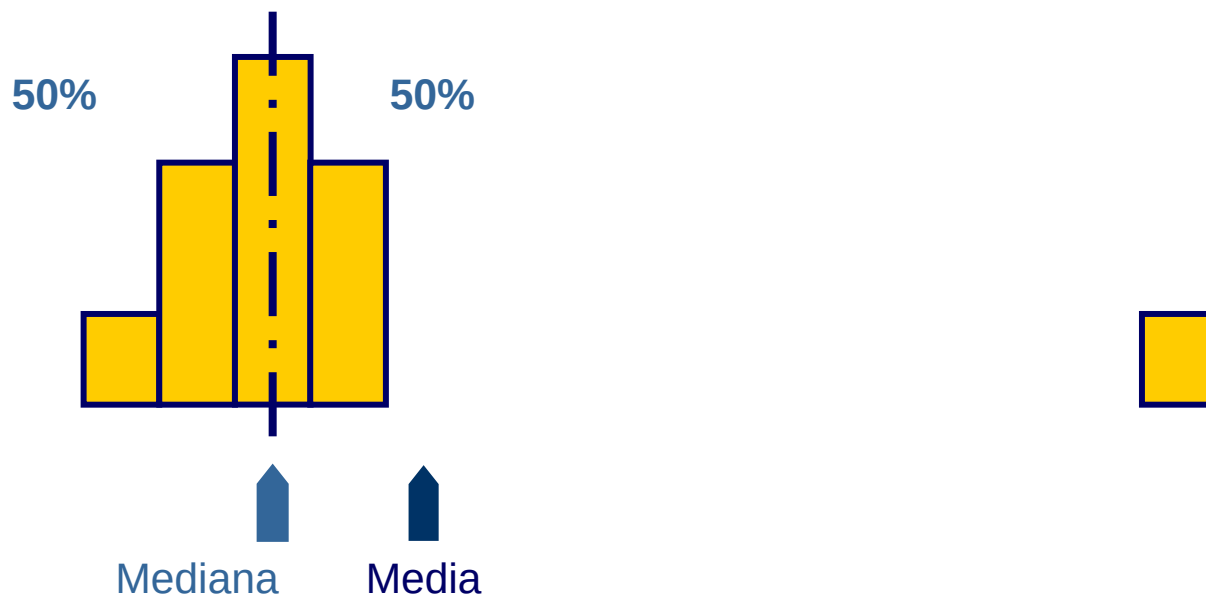


- Mediana

Es el valor que deja a cada lado el 50% de los datos.

Es poco sensible a asimetrías

Es insensible a valores atípicos



**Los valores atípicos no alteran la posición de la mediana**

**Con valores atípicos y asimetrías fuertes, la mediana es una medida de centralización más útil que la media**

- Moda

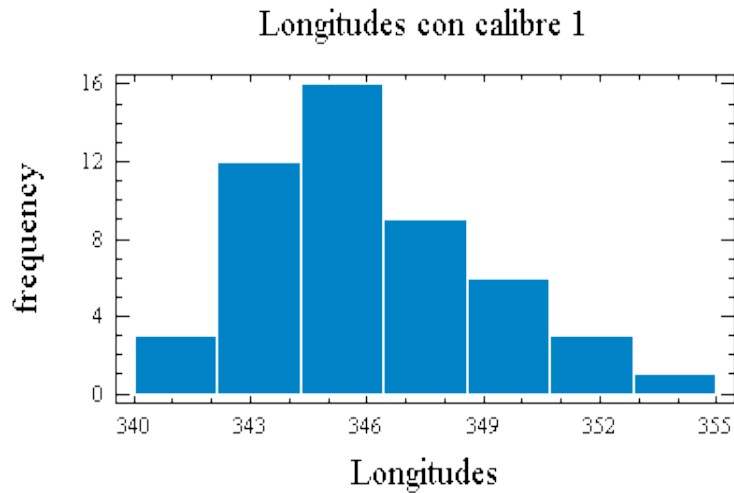
Es el valor más frecuente

1 2 2 2 2 5 5 5 8 8 11 13

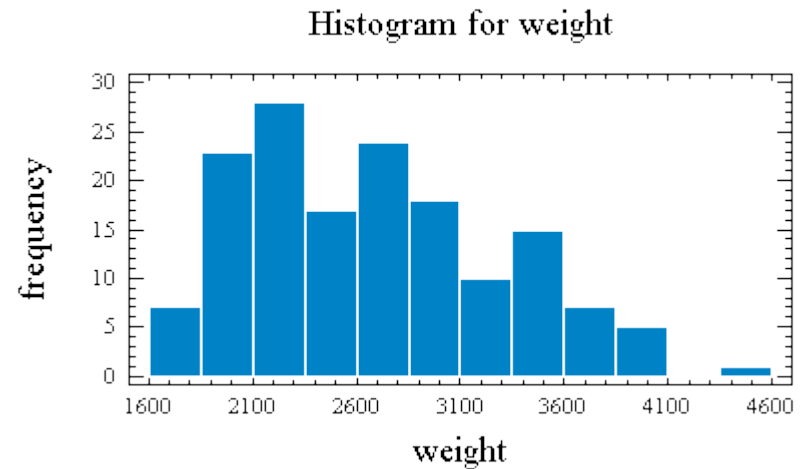
---

**Moda=2**

Con datos agrupados, es la clase más frecuente. Puede haber varias, sugiriendo la posible existencia de varios grupos



Distribución unimodal



Distribución trimodal

## 4.1 Medidas de centralización

media, mediana, moda

## 4.2 Medidas de dispersión

- Varianza (desviación típica)
- Cuasivarianza (cuasidesviación típica)
- Rango

### • Varianza

Promedio de desviaciones a la media, al cuadrado

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

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

Desviación típica

$$CV = \frac{s}{|\bar{x}|}$$

Coeficiente de variación

### • Cuasivarianza

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

$$\hat{s}_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

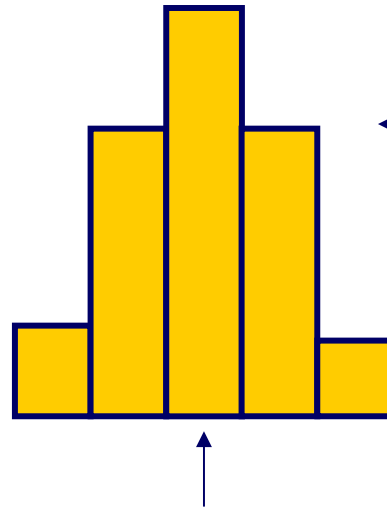
Se puede justificar teóricamente que miden mejor la dispersión de los datos.

Es la que usa R en la función var()

## • Varianza y Cuasivarianza

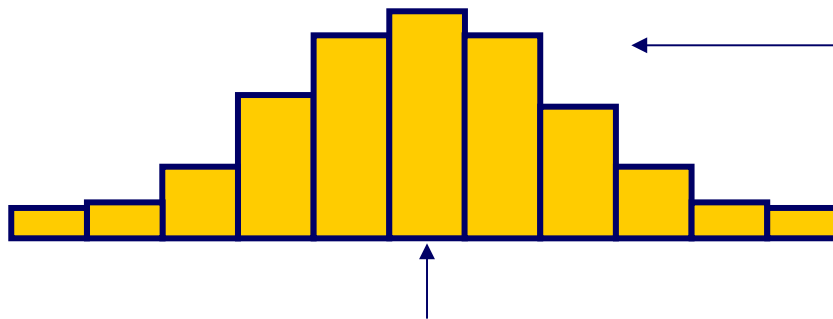
Promedio de desviaciones a la media, al cuadrado

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



media

Poca dispersión  
Baja varianza  
Baja desviación típica  
Bajo CV  
Baja cuasivarianza  
Baja cuasidesviación típica



media

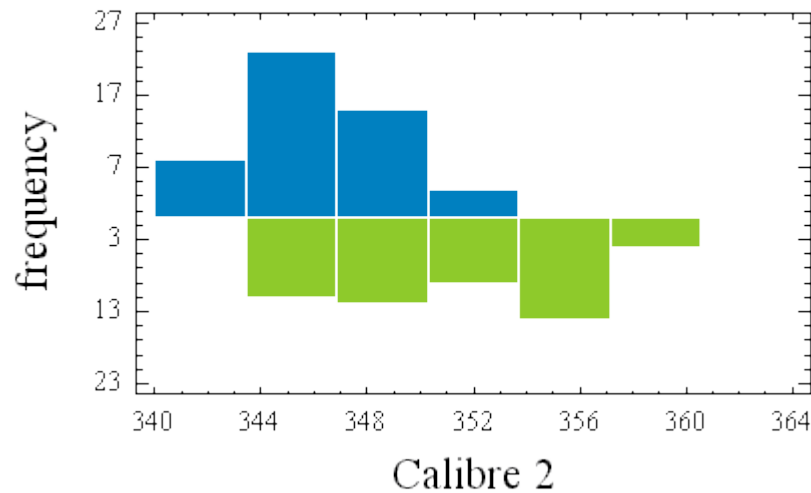
Mucha dispersión  
Altavarianza  
Alta desviación típica  
Alto CV  
Alta cuasivarianza  
Alta cuasidesviación típica



## Ejemplo:

Longitudes de 100 clavos del mismo tipo, medidos por dos personas, 50 clavos cada una, que usaron calibres diferentes.

Calibre 1



¿Qué calibre es mejor?

Varianza Calibre 1: **7.25 mm<sup>2</sup>**

Varianza Calibre 2: **21.47 mm<sup>2</sup>**

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

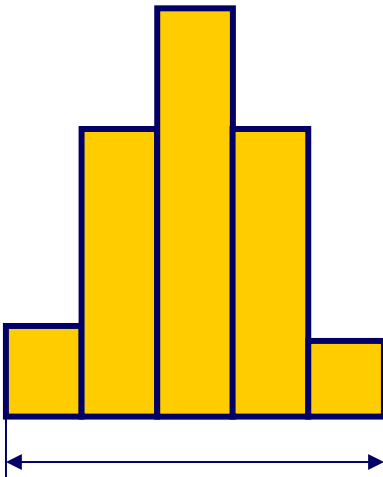
- **Rango**

Valor máximo menos valor mínimo

X: 1 2 5 8 11 13 24 28 31

Rango:  $31 - 1 = 30$

**A mayor rango mayor dispersión**



## • Percentiles (Medidas de posición)

Los percentiles son los 99 valores de los datos correspondientes al 1%, al 2%, ... y al 99% de los datos.

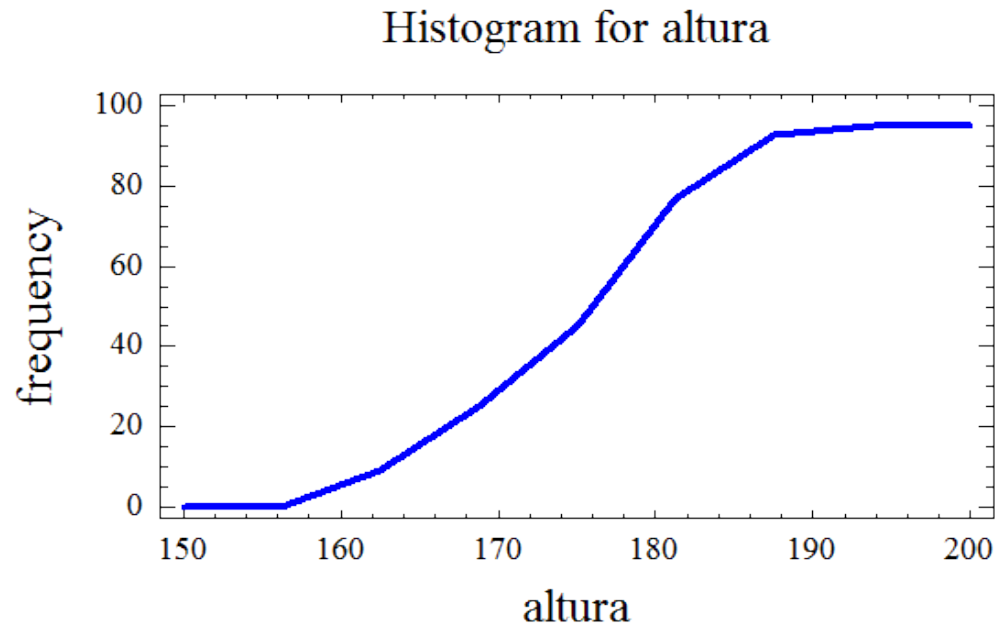
Dado un conjunto de datos, el valor **m** es el percentil **p**, si el **px100%** de los datos son menores a **m**

### Ejemplo:

Si un estudiante que mide 1,69 m tiene una altura que es el percentil 30 de su clase, quiere decir que el 30% de sus compañeros mide menos de 1,69 m

### Ejemplo:

La siguiente figura es el polígono de frecuencias acumulado de las estaturas del fichero AlumnosIndustriales.xls. A la vista de esta figura, determina los percentiles 20 y 80.



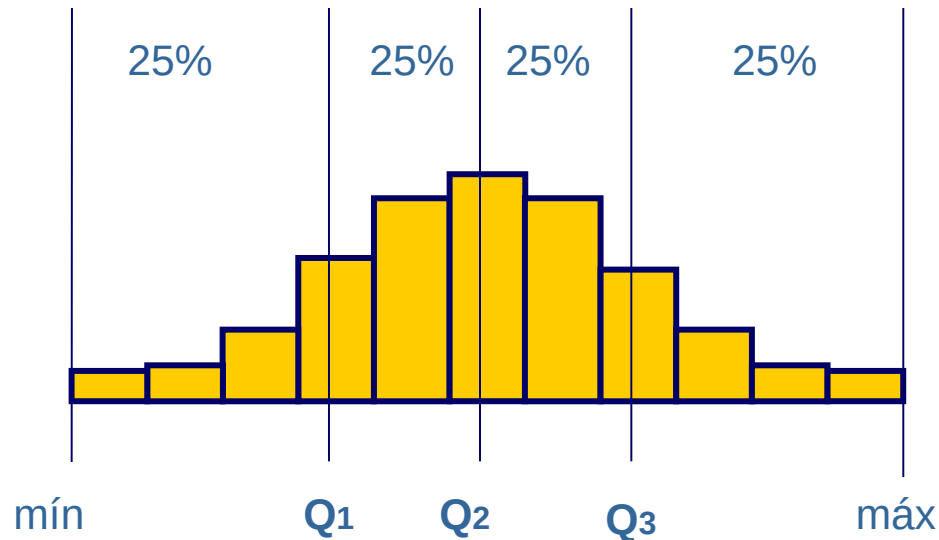
- Cuartiles  $Q_1$ ,  $Q_2$ ,  $Q_3$

Son los valores que dividen la muestra en 4 grupos, cada uno con el 25% de los datos (aproximadamente)

Entre el mínimo y $Q_1$	→	25% de los datos	}	50%
Entre $Q_1$ y $Q_2$	→	25% de los datos		
Entre $Q_2$ y $Q_3$	→	25% de los datos	}	50%
Entre $Q_3$ y el máximo	→	25% de los datos		

$Q_1$ =percentil 25;  $Q_2$  = mediana;  $Q_3$ =percentil 75

$Q_3 - Q_1$ =Rango Intercuartílico (R.I.)



Hay varios métodos para calcular  $Q_1$  y  $Q_3$ . Con pocos datos pueden dar valores diferentes

- Cuartiles  $Q_1$ ,  $Q_2$ ,  $Q_3$

$x:\{1,1,3,3,5,9,11,14,15\}$

Método sencillo para calcular cuartiles

1º: Obtenemos la mediana  $Q_2$



5

2º: Excluimos ese valor y nos quedamos con dos grupos de datos, uno a cada lado de la mediana



izda.:  $\{1,1,3,3\}$

dcha.:  $\{9,11,14,15\}$

3º:  $Q_1$  es la mediana del grupo que queda a la izquierda



$Q_1=(1+3)/2=2$

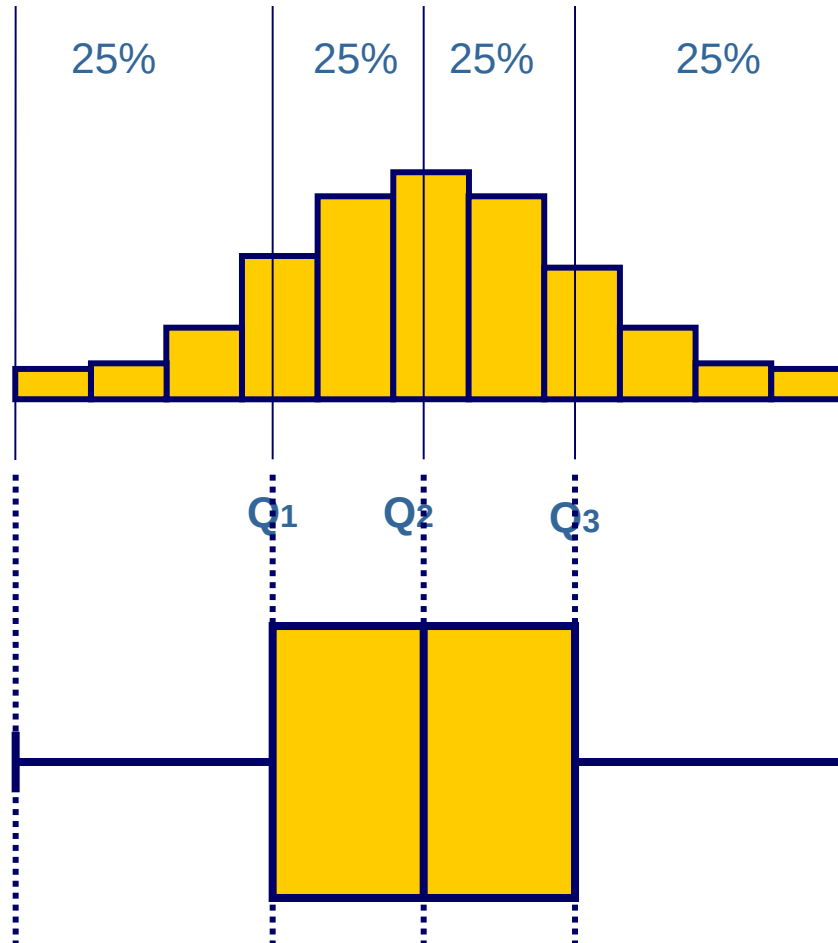
4º:  $Q_3$  es la mediana del grupo que queda a la derecha



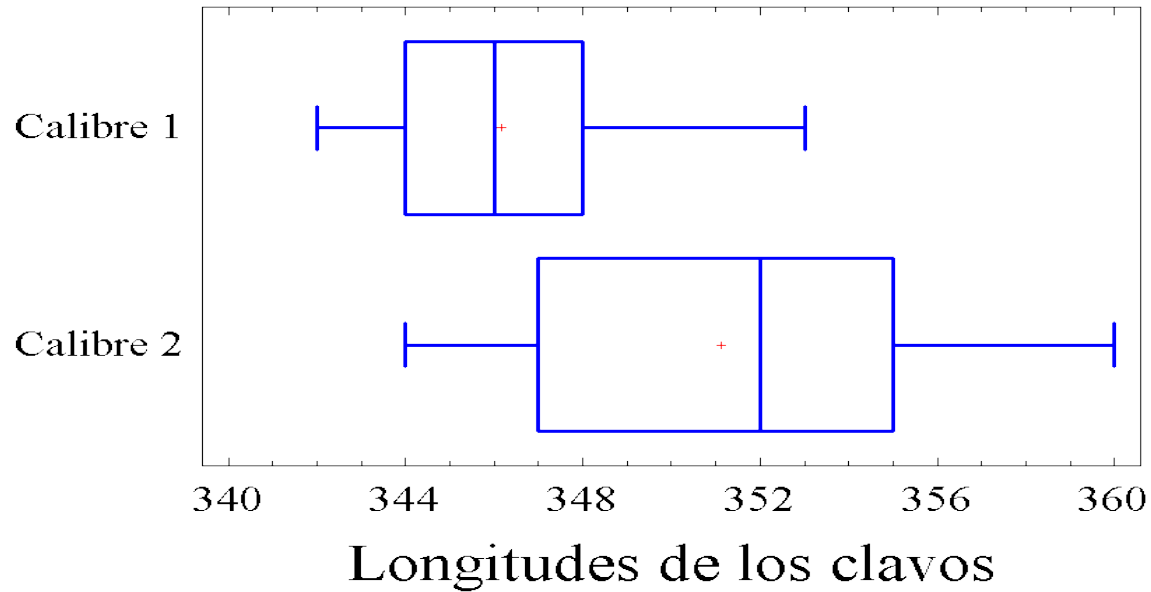
$Q_3=(11+14)/2=12.5$

- **Diagrama de caja (Box-Plot)**

Es la representación gráfica de los cuartiles



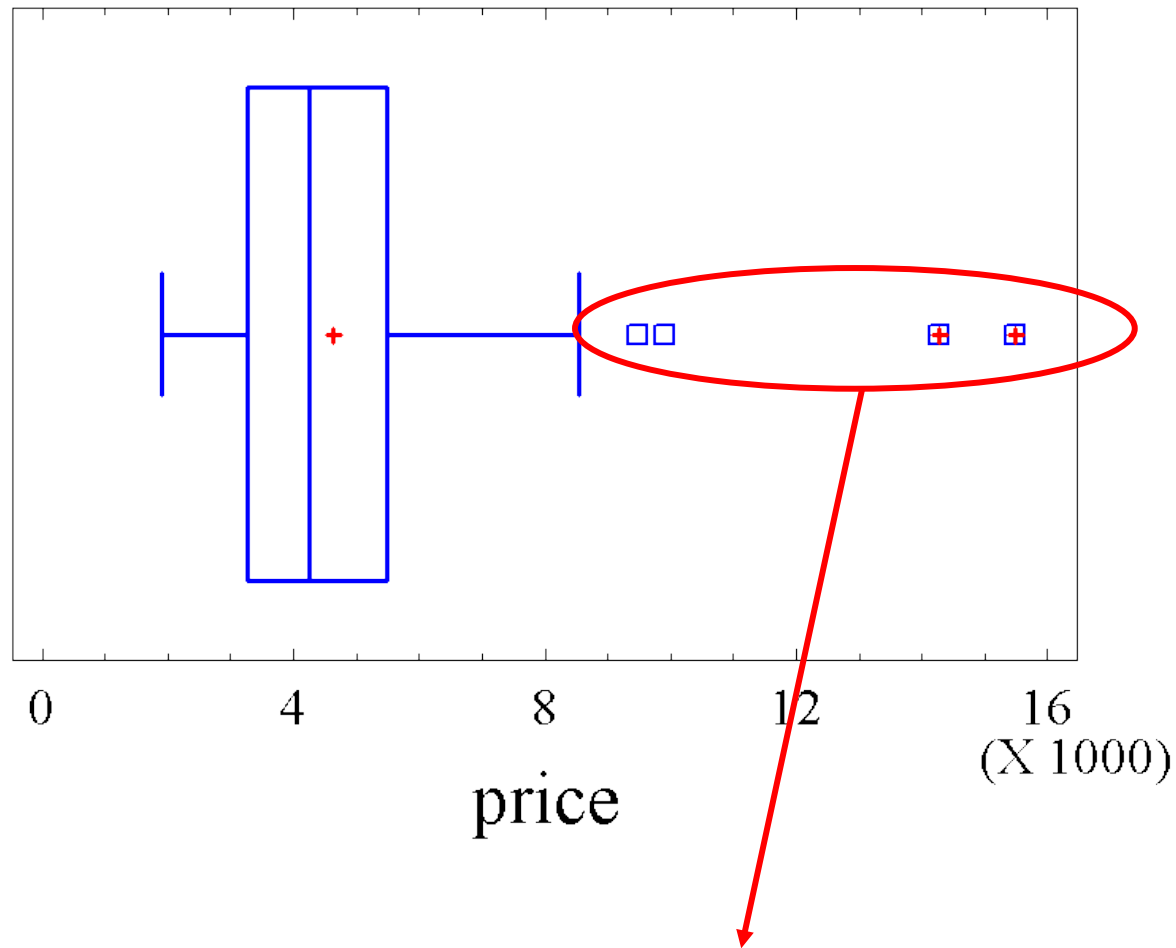
## Box-and-Whisker Plot



Los gráficos Box-plot son muy útiles para:

- **Comparar grupos**
- Ver asimetrías
- Detectar atípicos \*\*

# Box-and-Whisker Plot

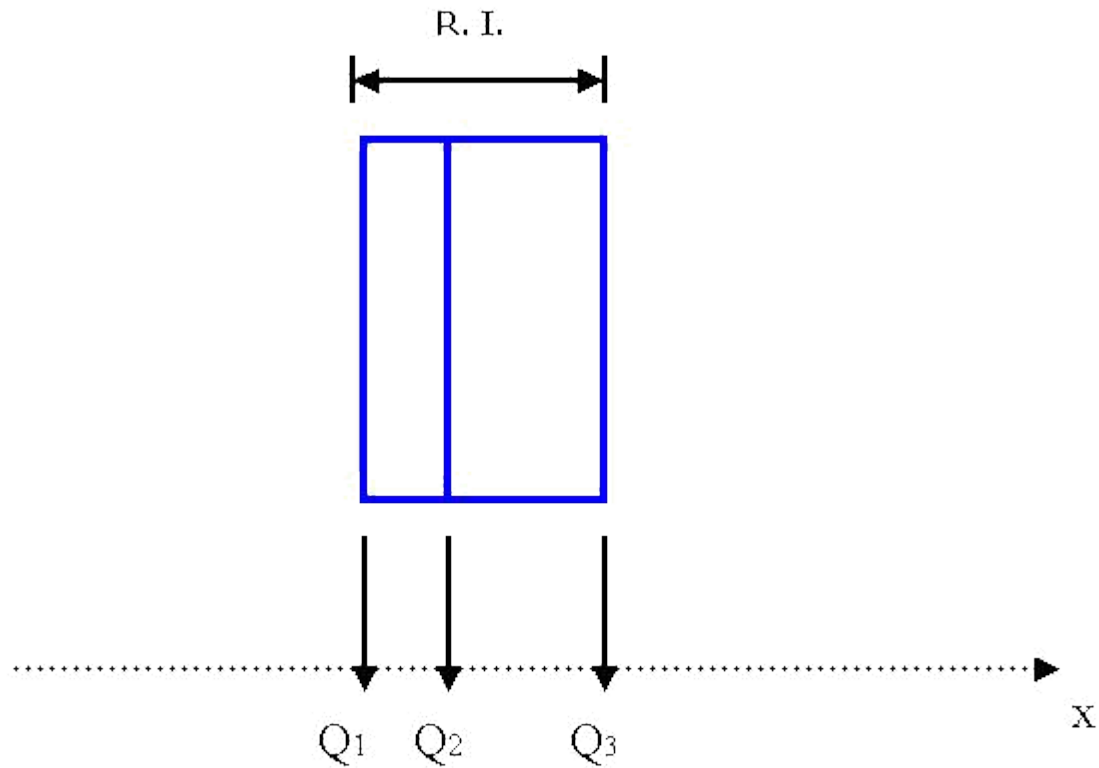


Datos extremos (o 'atípicos')

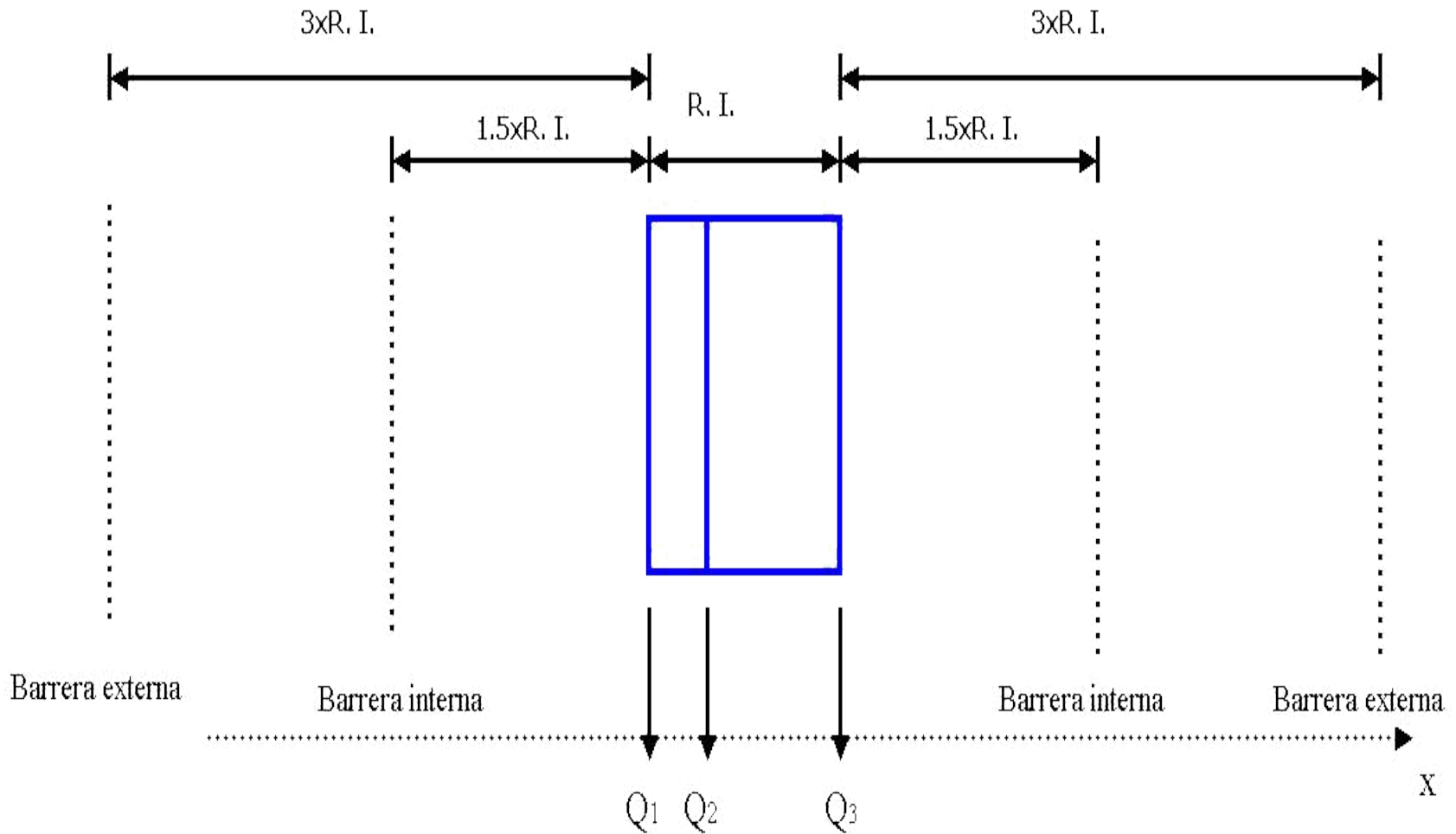


## Para construir un Box-plot con marcas de atípicos

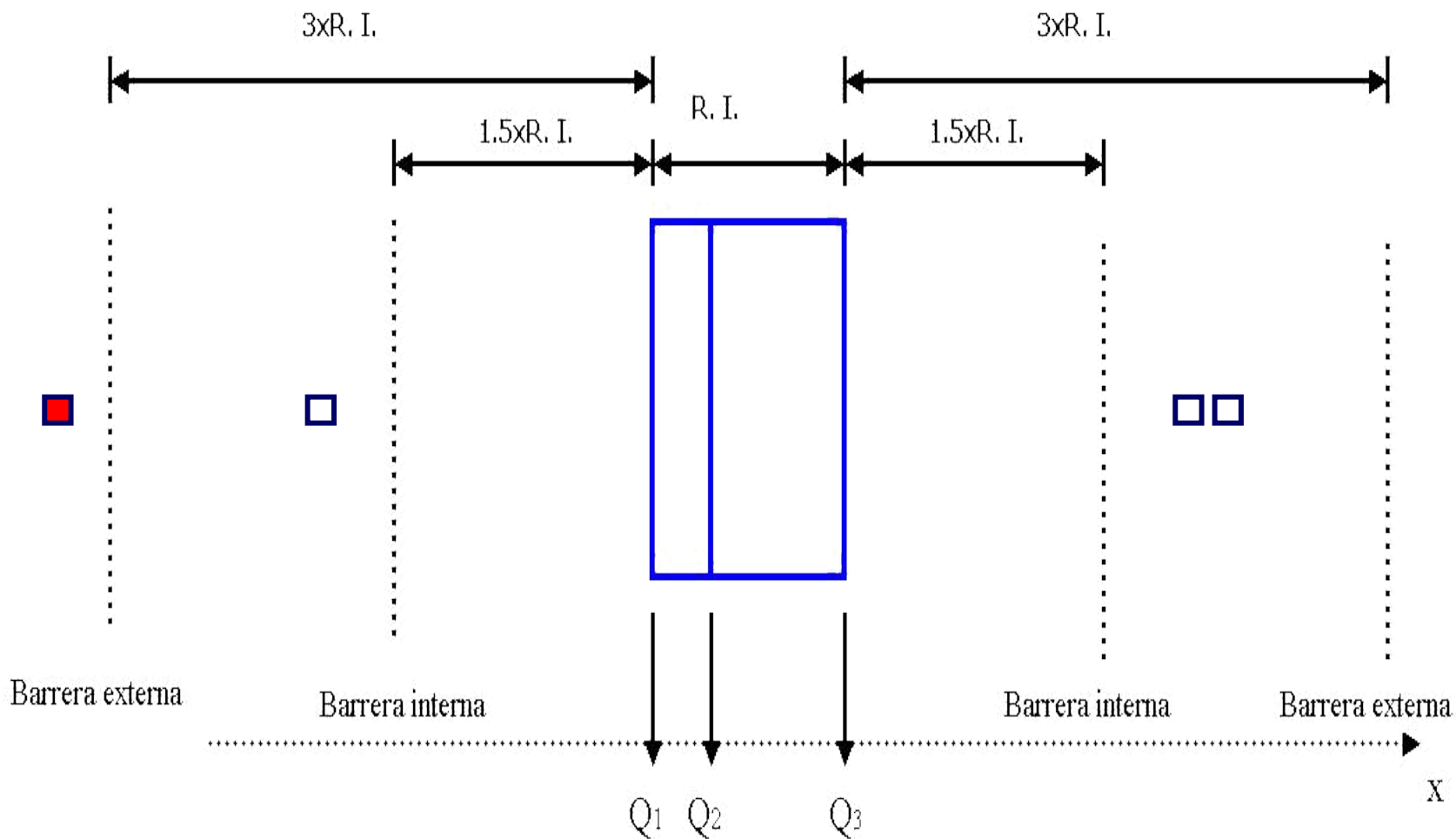
### Primer paso



## Segundo paso

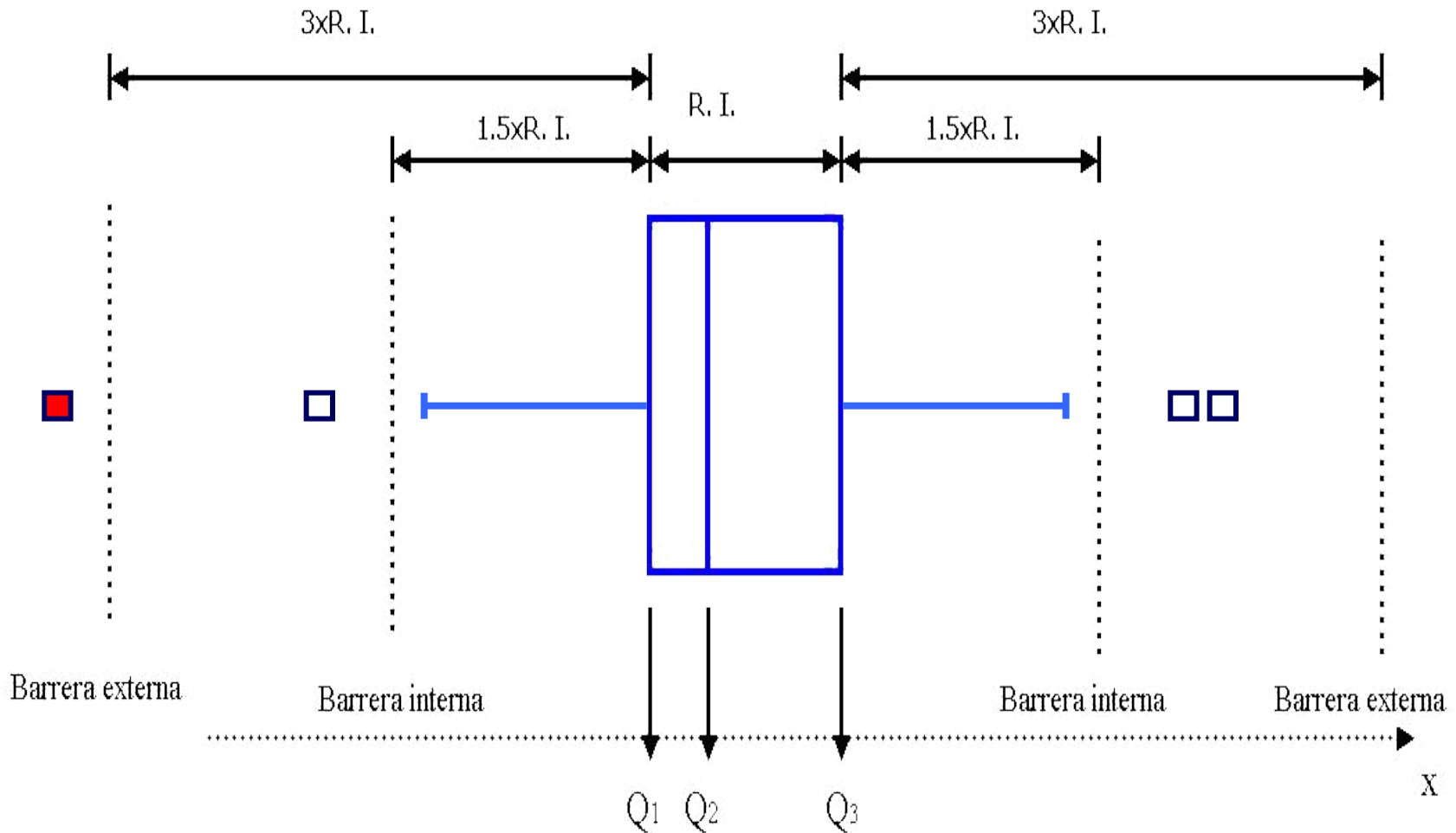


### Tercer paso



**Los puntos que caigan en estas zonas se marcan**

### Tercer paso

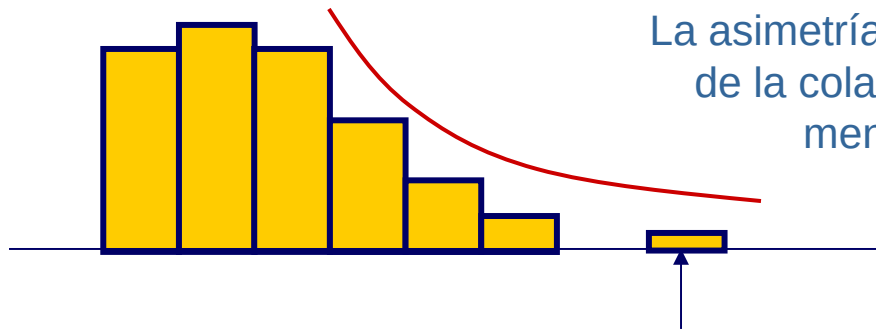


Las líneas laterales se extienden sólo hasta el último punto dentro de la barrera interna

Ojo!! Cuando hay asimetrías, un dato extremo no debe necesariamente catalogarse como atípico

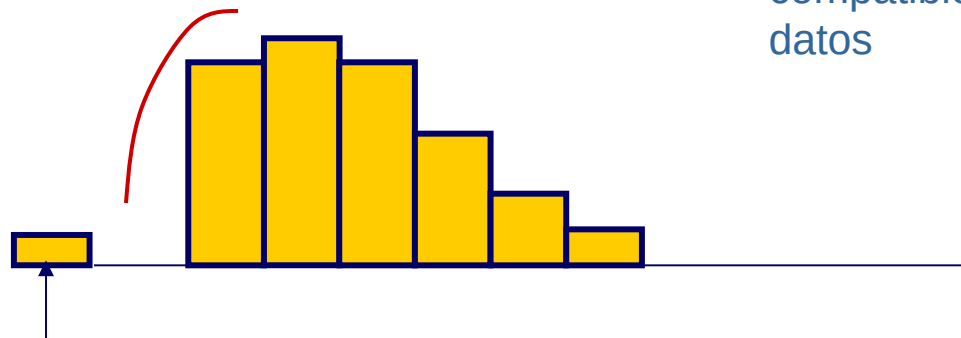


Dato atípico: dato que se sale del patrón general de los datos



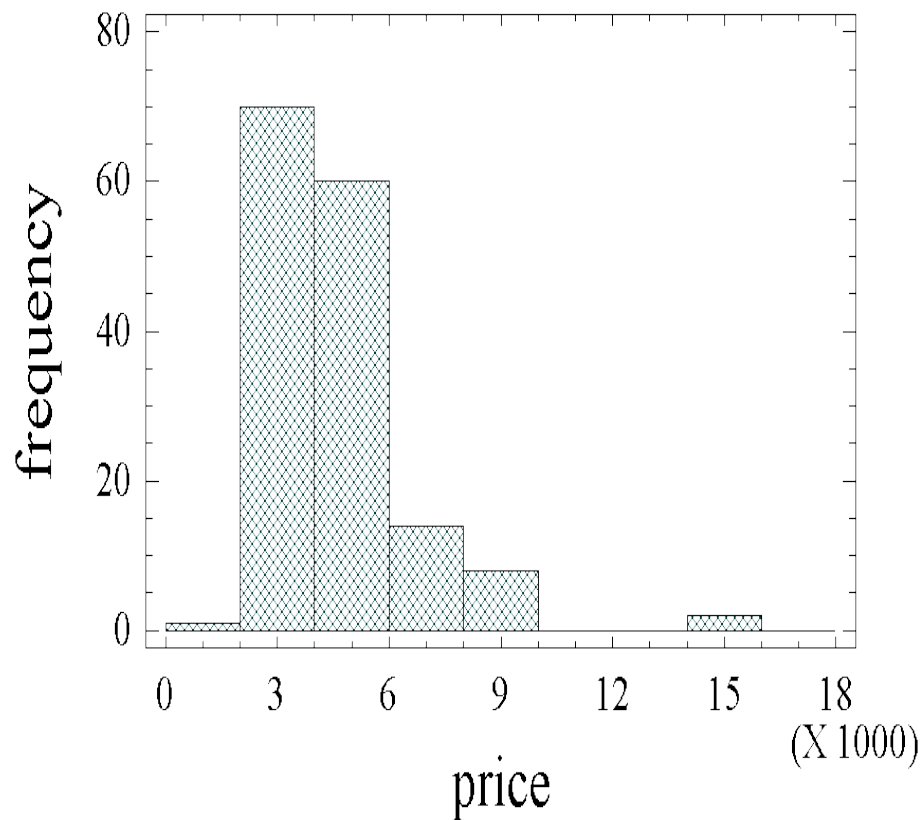
La asimetría indica que al final de la cola habrá cada vez menos datos.

No es necesariamente un atípico. Es compatible con el patrón de los datos

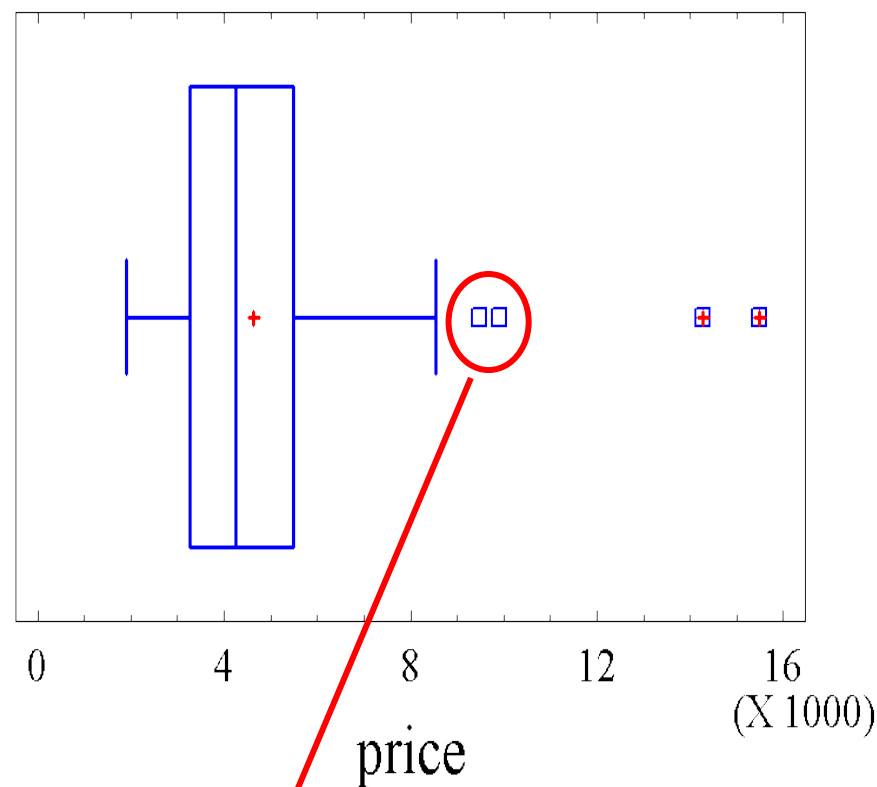


Sí es atípico

# Histogram for price



# Box-and-Whisker Plot



Son compatibles con la asimetría positiva

#### 4.1 Medidas de centralización

media, mediana, moda

#### 4.2 Medidas de dispersión

varianza, desv. típica, coeficiente de  
variación, meda, rango, rango  
intercuartílico

#### 4.3 Otras medidas de forma

- Medidas de asimetría
- Medidas de apuntamiento

#### • Medidas de asimetría

Coeficiente de  
asimetría

$$CA = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n s_x^3}$$

- $CA = 0$ ; si la distribución es perfectamente simétrica
- $CA > 0$ ; si hay asimetría positiva
- $CA < 0$ ; si hay asimetría negativa

- **Medidas de asimetría**

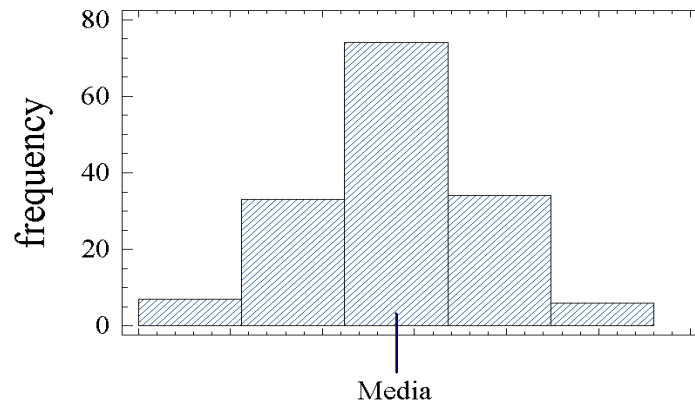
Coeficiente de  
asimetría



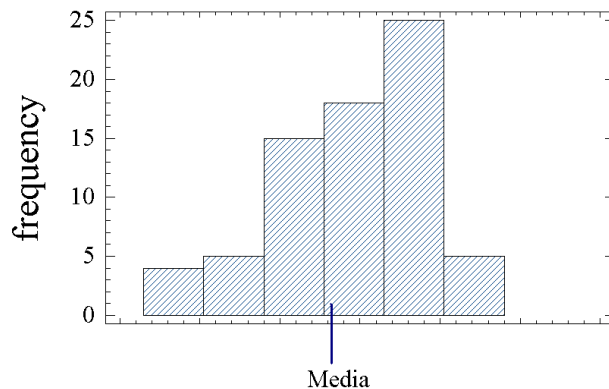
$$CA = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n s_x^3}$$

- $CA = 0$ ; si la distribución es perfectamente simétrica
- $CA > 0$ ; si hay asimetría positiva
- $CA < 0$ ; si hay asimetría negativa

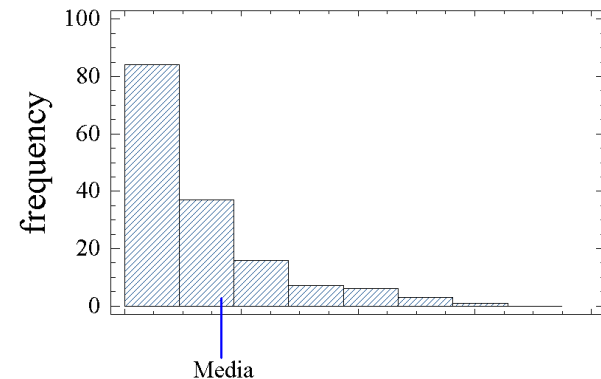
Distribución simétrica



Distribución asimétrica negativa

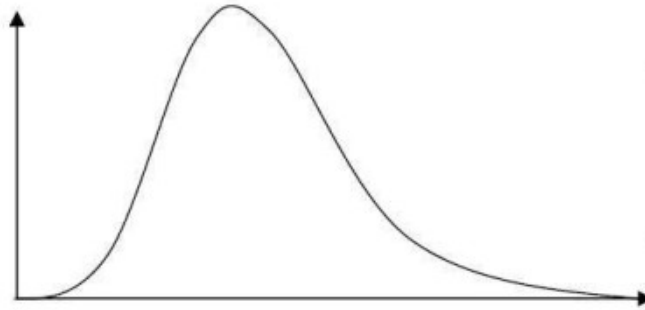


Distribución asimétrica positiva

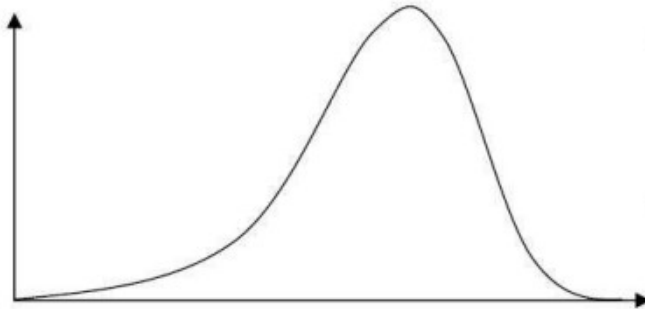




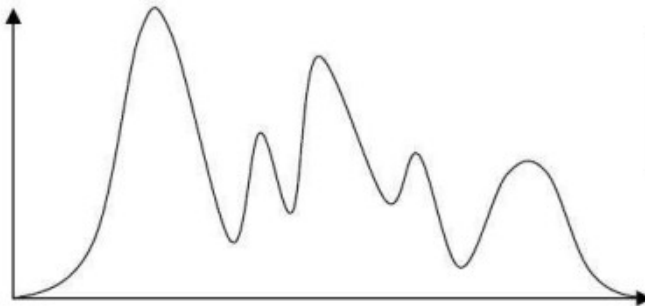
## DISTRIBUCIÓN ASIMÉTRICA



**Distribución Sesgada hacia la Izquierda:** Los datos se concentran hacia la izquierda de la distribución.



**Distribución Sesgada hacia la Derecha:** Los datos se concentran hacia la derecha de la distribución.

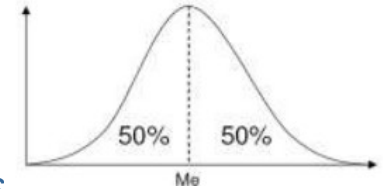


**Distribución asimétrica:** No presenta uniformidad en la distribución de los datos.

## RELACIÓN ENTRE LA MEDIA, MEDIANA Y MODA

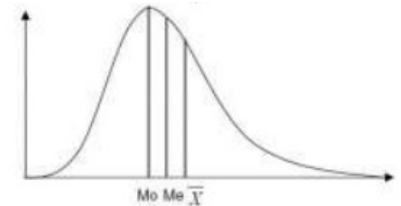
$$X = Me = Mo$$

Cuando una distribución de frecuencia es simétrica, la media, mediana y moda coinciden en su valor ( $X = Me = Mo$ ). En el caso de una distribución binomial simétrica, es necesario calcular el promedio de las modas.



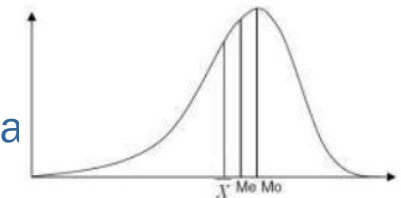
$$Mo < Me < X$$

En una distribución sesgada a la izquierda, la moda es menor a la mediana, y esta a su vez menor que la media.



$$Mo > Me > X$$

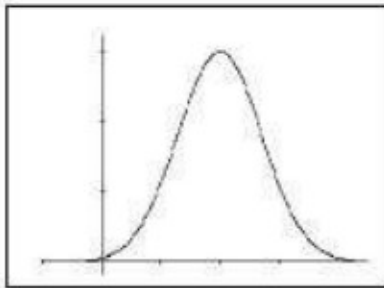
En una distribución sesgada a la derecha la relación se invierte, la moda es mayor a la mediana, y esta a su vez mayor que la media.



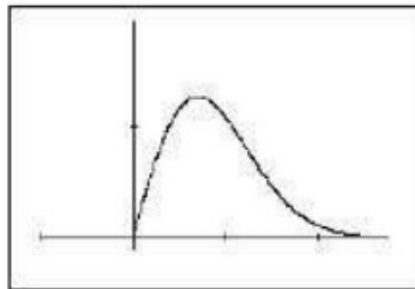
## COEFICIENTE DE ASIMETRÍA

Mide el grado de asimetría de la distribución con respecto a la media. Un valor positivo de este indicador significa que la distribución se encuentra sesgada hacia la izquierda (orientación positiva). Un resultado negativo significa que la distribución se sesga a la derecha.

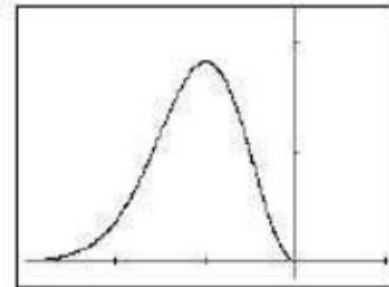
Sesgo = 0



Sesgo > 0



Sesgo < 0



- Medidas de apuntamiento

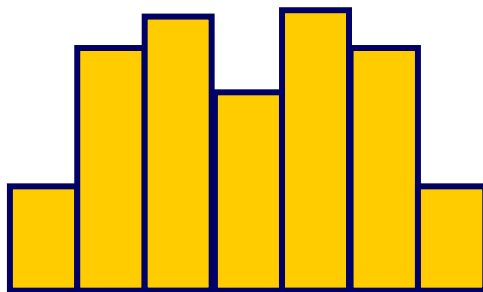
Coeficiente de  
apuntamiento o  
Curtosis



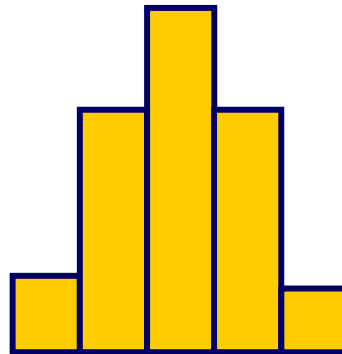
$$CAp = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n s_x^4}$$

- $CAp = 3$ ; distribución con forma de campana
- $CAp > 3$ ; distribución más apuntada que una campana
- $CAp < 3$ ; distribución menos apuntada que una campana

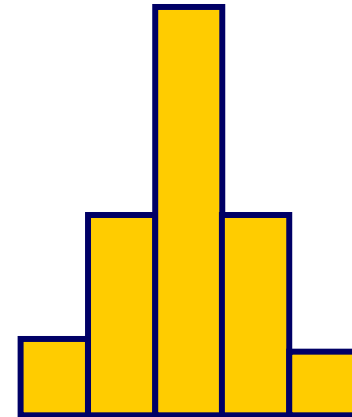
$CAp < 3$



$CAp = 3$



$CAp > 3$



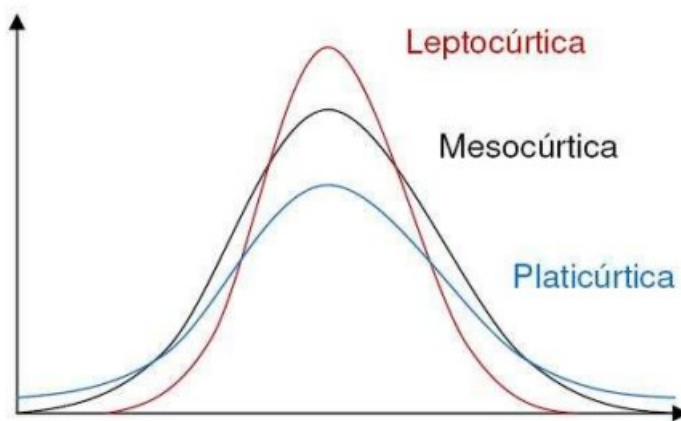
Muchos softwares estadísticos definen curtosis como  **$CAp-3$** .

Esta medida recibe también el nombre de 'Exceso de curtosis'.

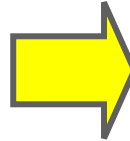
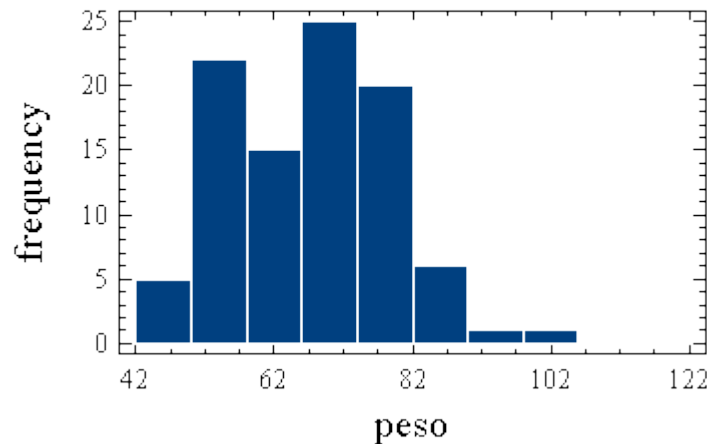
## COEFICIENTE DE APUNTAMIENTO O CURTOSIS

Indica que tan apuntada o achatada se encuentra una distribución respecto a un comportamiento normal (distribución normal).

- ✓ LEPTOCÚRTICA: los datos están muy concentrados hacia la media  $CAp > 3$
- ✓ PLATICÚRTICA: Si los datos están muy dispersos,  $CAp < 3$
- ✓ MESOCÚRTICA: El comportamiento normal exige que la curtosis sea  $CAp = 3$



Histogram for peso



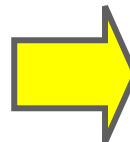
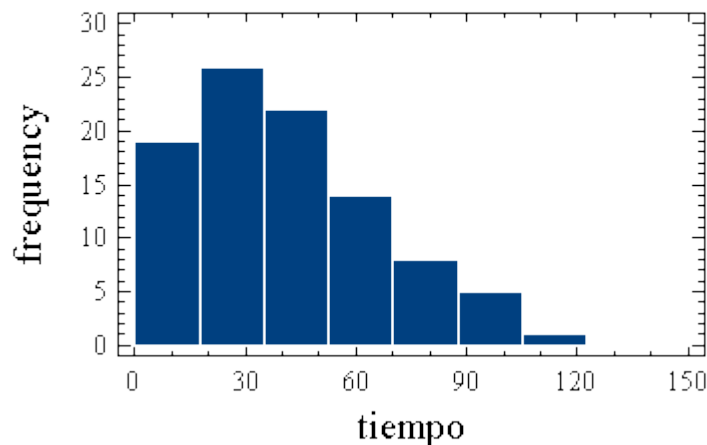
Summary Statistics for peso

```
Count = 95
Average = 67,7684
Median = 69,0
Skewness = 0,261155
Kurtosis = -0,502931
```

(Curtosis-3: 'exceso de curtosis')

Valores de curtosis muy bajos son indicio de 'multimodalidad'

Histogram for 'Tiempo en llegar a la Universidad'



Summary Statistics for tiempo

```
Count = 95
Average = 41,4211
Median = 40,0
Skewness = 0,651076
Kurtosis = 0,0915265
```