

Estadística General

Tema 2: Estadística descriptiva

Estadísticos Centrales



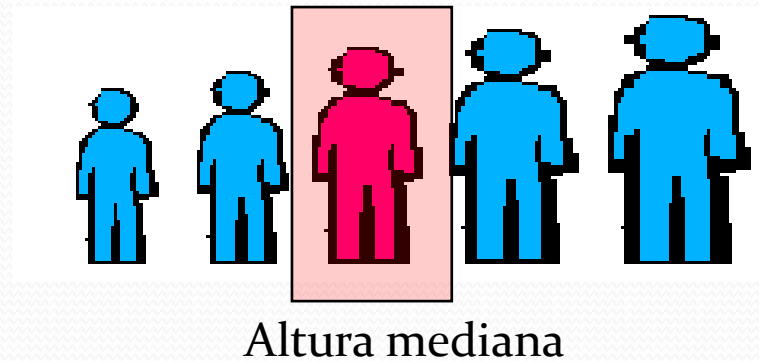
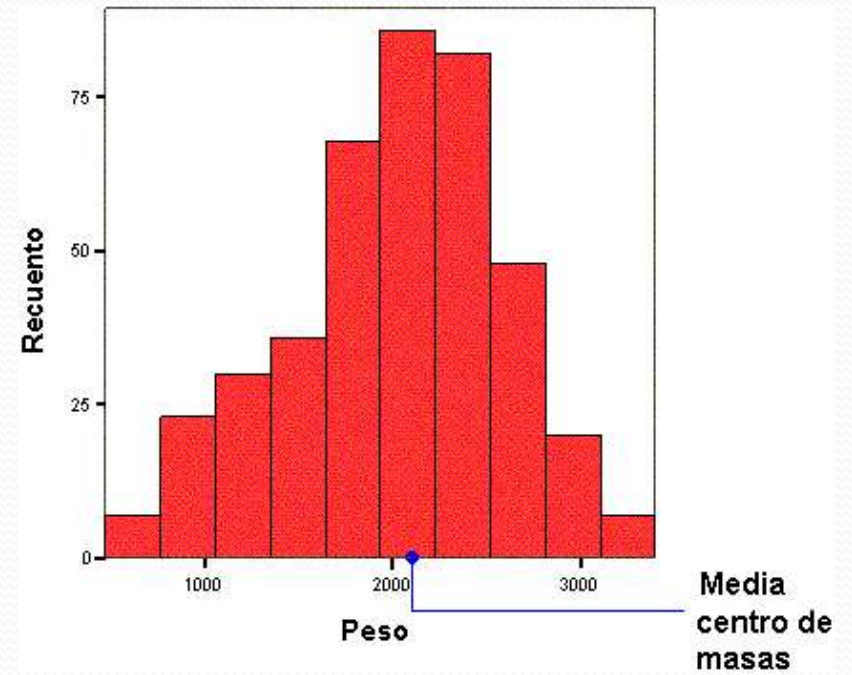
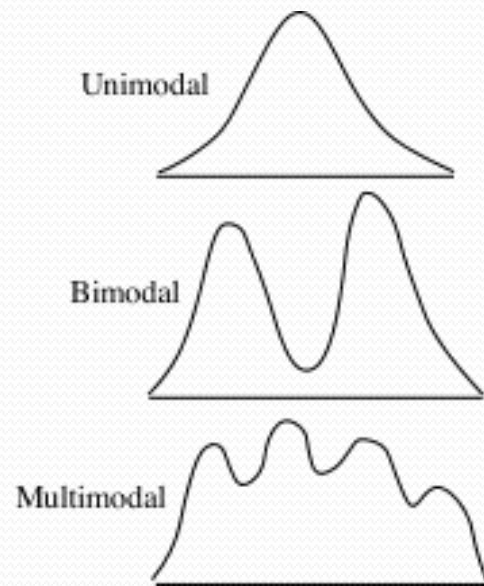
Prof. José G. Páez

Estadísticos de centralización

Añaden unos cuantos casos particulares a las medidas de posición. En este caso son medidas que buscan posiciones (valores) con respecto a los cuales los datos muestran tendencia a agruparse.

- **Media** ('mean') Es la media aritmética (promedio) de los valores de una variable. Suma de los valores dividido por el tamaño muestral.
 - Media de 2,2,3,7 es $(2+2+3+7)/4=3,5$
 - Conveniente cuando los datos se concentran simétricamente con respecto a ese valor. Muy sensible a valores extremos.
 - Centro de gravedad de los datos
- **Mediana** ('median') Es un valor que divide a las observaciones en dos grupos con el mismo número de individuos (percentil 50). Si el número de datos es par, se elige la media de los dos datos centrales.
 - Mediana de 1,2,4,5,6,6,8 es 5
 - Mediana de 1,2,4,5,6,6,8,9 es $(5+6)/2=5,5$
 - Es conveniente cuando los datos son asimétricos. No es sensible a valores extremos.
 - Mediana de 1,2,4,5,6,6,800 es 5. ¡La media es 117,7!
- **Moda** ('mode') Es el/los valor/es donde la distribución de frecuencia alcanza un máximo.

Estadísticos de centralización



Algunas fórmulas

- **Datos sin agrupar:** x_1, x_2, \dots, x_n

- Media

$$\bar{x} = \frac{\sum_i x_i}{n}$$

- **Datos organizados en tabla**

- si está en intervalos usar como x_i las marcas de clase.
Si no ignorar la columna de intervalos.

- Media $\bar{x} = \frac{\sum_i x_i n_i}{n}$

- Cuantil de orden α

- i es el menor intervalo que tiene frecuencia acumulada superior a $\alpha \cdot n$
- $\alpha=0,5$ es mediana

$$C_\alpha = L_{i-1} + \frac{\alpha \cdot n - N_{i-1}}{n_i} (L_i - L_{i-1})$$

Variable	fr.	fr. ac.	
$L_0 - L_1$	x_1	n_1	N_1
$L_1 - L_2$	x_2	n_2	N_2
...			
$L_{k-1} - L_k$	x_k	n_k	N_k
n			

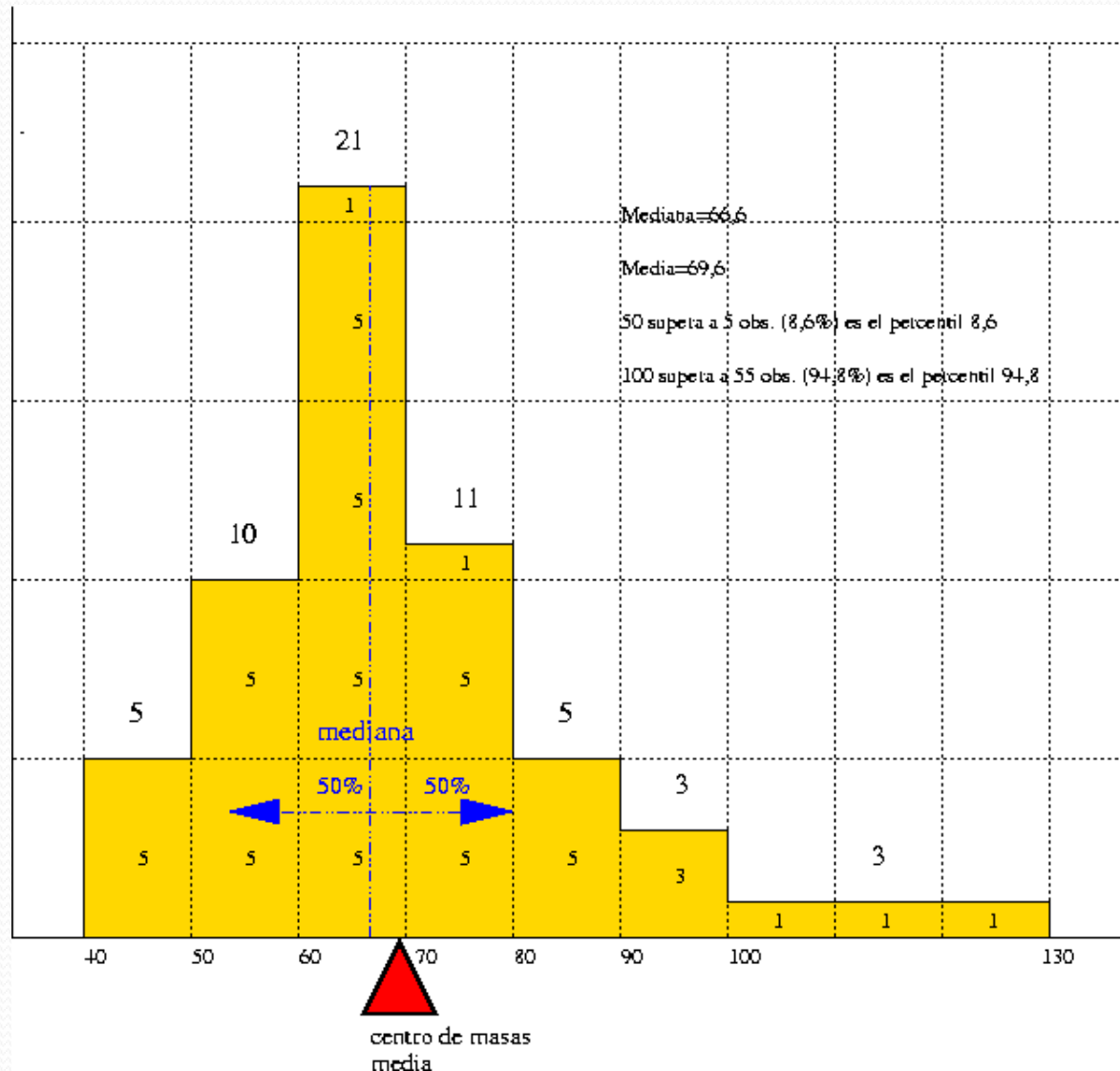
Ejemplo con variable en intervalos

Peso	M. Clase	frec	Fr. acum.
40 – 50	45	5	5
50 – 60	55	10	15
60 – 70	65	21	36
70 - 80	75	11	47
80 - 90	85	5	52
90 - 100	95	3	55
100 – 130	115	3	58

En el histograma se identifica “unidad de área” con “individuo”.

Para calcular la media es necesario elegir un punto representante del intervalo: La marca de clase.

La media se desplaza hacia los valores extremos. No coincide con la mediana. Es un punto donde el histograma “estaría en equilibrio” si tuviese masa.



Ejemplo (continuación)

Peso	M. Clase	Fr.	Fr. ac.
40 - 50	45	5	5
50 - 60	55	10	15
60 - 70	65	21	36
70 - 80	75	11	47
80 - 90	85	5	52
90 - 100	95	3	55
100 - 130	115	3	58
			58

$$\bar{x} = \frac{\sum_i x_i n_i}{n} = \frac{45 \cdot 5 + 55 \cdot 10 + \dots + 115 \cdot 3}{58} = 69,3$$

$$\begin{aligned} \text{Mediana} &= C_{0,5} = L_{i-1} + \frac{0,5 \cdot 58 - N_{i-1}}{n_i} (L_i - L_{i-1}) \\ &= 60 + \frac{0,5 \cdot 58 - 15}{21} (70 - 60) = 66,6 \end{aligned}$$

$$P_{75} = C_{0,75} = L_{i-1} + \frac{0,75 \cdot 58 - N_{i-1}}{n_i} (L_i - L_{i-1}) = 70 + \frac{43,5 - 36}{11} (80 - 70) = 76,8$$

- Moda = marca de clase de (60,70] = 65
 - Cada libro ofrece una fórmula diferente para la moda (difícil estar al día.)