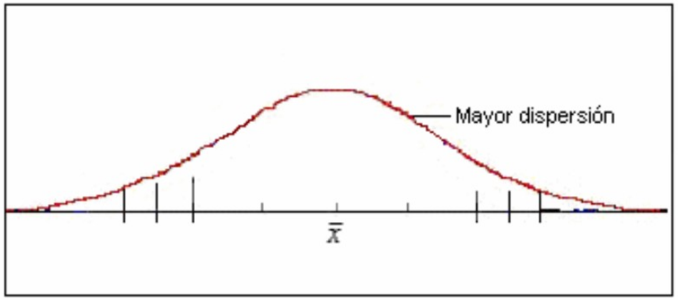
MEDIDAS DE DISPERSION

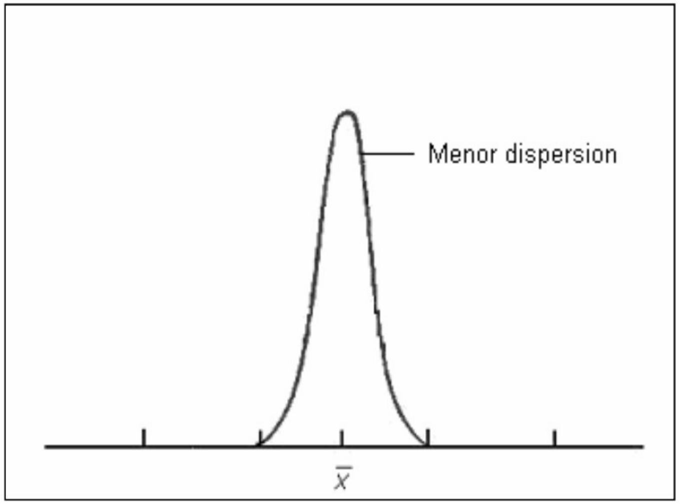
¿Es la media representativa?

Queremos cuantificar la separación de los valores de la distribución respecto a la media. Si todos los valores están cercanos al valor medio, la media es representativa.

MEDIDAS DE DISPERSIÓN

Miden qué tanto se dispersan las observaciones alrededor de su media.





MEDIDAS DE DISPERSIÓN

En algunos casos existen conjuntos de datos que tienen la misma media y la misma mediana, pero esto no refleja qué tan dispersos están los elementos de cada conjunto.

Ejemplo:

Conjunto 1. 80, 90, 100, 110, 120
Conjunto 2. 0, 50, 100, 150, 200

Conjunto 1

$$Media = \frac{80 + 90 + 100 + 110 + 120}{5} = 100$$

Conjunto 2

$$Media = \frac{0 + 50 + 100 + 150 + 200}{5} = 100$$

Observa que para ambos conjuntos la Mediana es igual a 100. También nota que los datos del conjunto 2 están más dispersos con respecto a su media que los datos del conjunto 1.

Medidas de dispersión

- Rango
- Varianza
- Desviación Estándar
- Coeficiente de Variación

El rango (R)

Llamado también recorrido, amplitud total o alcance.

a) Obtención: se obtiene de la influencia entre el dato mayor y el dato menor

Ejemplo:

Los siguientes datos representan el peso de 10 niños al nacer, (en Kg.). Calcule e interprete el rango.

2,860	3,150	3,450	2,950	3,780
4,170	3,920	3,280	4,050	3,120

Rango = (4,170 - 2,860)

Rango = 1,310 Kg.

b) Interpretación

La diferencia entre el bebe de mayor peso y el bebe menor peso es 1,310 Kg.

c) Cálculo a partir de datos agrupados, se utiliza la siguiente fórmula:

$$R = (L_s - L_i) + 1$$

donde: L_s : Limite superior de la última clase

L_i : Limite inferior de la primera clase

Ejemplo:

La distribución de frecuencias siguiente representa el tiempo que espera un paciente para ser atendido, en un consultorio externo. Calcule e interprete el rango

Tiempo (minutos)	Nº de Pacientes (por día)
12 - 16	4
17 - 21	8
22 - 26	15
27 - 31	23
32 - 36	10
Total	**60**

$$\text{Rango} = (36-12) + 1$$

$$R = 25 \text{ minutos}$$

Interpretación: la diferencia de tiempo entre el paciente que más espera y el que menos espera para ser atendido es 25 minutos.

f) Ventajas y desventajas del rango

Ventajas

- fácil de calcular
- fácil de entender e interpretar

Desventajas

- sólo considera los valores extremos
- no toma en cuenta ni el número de datos ni el valor de estos
- no es posible calcular en tablas con extremos abiertos.

La varianza

Es una medida de desviación promedio con respecto a la media aritmética

a) Cálculos a partir de datos no agrupados.

para una muestra

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

para un población

$$\sigma^2 = \frac{\sum_{i=1}^{N} \left(X_i - \mu \right)^2}{N}$$

Ejemplo:

La siguiente información se refiere al número de radiografías reprocesadas durante una semana. Calcule la varianza. 8, 10, 5, 12, 10, 15

Primero, elaboramos un cuadro de la forma siguiente:

x	$X_i - \bar{x}\ \chi$	$\left(X_i - \bar{x}\right)^2$
8	8 - 10 = 2	4
10	10 - 10 = 0	0
5	5 - 10 = 5	25
12	12 - 10 = 2	4
10	10 - 10 = 0	0
15	15 - 10 = 5	25
$\sum X = 60$	$\sum \left(X_i - \bar{x}\right) = 0$	$\sum \left(X_i - \bar{x}\right)^2 = 58$

$$\bar{x} = \frac{60}{6}$$

$$\bar{x} = 10$$

$$\sum \left(X_i - \bar{x} \right)^2 = 58$$

$$S^2 = \frac{\sum \left(X_i - \bar{x} \right)^2}{n-1}$$

$$S^2 = \frac{58}{6-1} = 11.6$$

La varianza mide la mayor o menor dispersión de los valores de la variable respecto a la media aritmética. Cuanto mayor sea la varianza mayor dispersión existirá y por tanto menor representatividad tendrá la media aritmética.

La varianza se expresa en las mismas unidades que la variable analizada, pero elevadas al cuadrado.

La varianza es positiva para un variable (Un constante tienen la varianza cero!)

Si sumamos a todos los valores de la variable una constante, la varianza no varía.

Si multiplicamos a todos los valores de la variable una constante, la varianza queda multiplicada por el cuadrado de la constante.

Ejercicio: Calcule la varianza de la siguiente distribución

12, 6, 7, 3, 15, 10, 18, 5.

Ejercicio

- La siguiente muestra representa las edades de 25 personan sometidas a un análisis de preferencias para un estudio de mercado.

25	19	21	35	44
20	27	32	38	33
18	30	19	29	33
26	24	28	39	31
31	18	17	30	27

Desviación Estandar

- La varianza transforma todas las distancias a valores positivos elevándolas al cuadrado, con el inconveniente de elevar consigo las unidades de los datos originales.

La desviación estándar

Llamada también desviación típica representa la variabilidad (o desviaciones) promedio de los datos con respecto a la media aritmética. Es la raíz cuadrada de la varianza, sea poblacional o muestral.

a) Cálculos a partir de datos no agrupados

para la muestra

$$S=\sqrt{S^{2}}=\sqrt{\frac{\sum_{i=1}^{n}\left(X_{i}-\bar{x}\right)^{2}}{n-1}}$$

para la población

$$\sigma=\sqrt{\sigma^{2}}=\sqrt{\frac{\sum_{i=1}^{N}\left(X_{i}-\mu\right)^{2}}{N}}$$

Ejemplo:

La siguiente información se refiere al número de radiografías reprocesadas durante una semana. Calcule la desviación estándar.

8, 10, 5, 12, 10, 15

Ya sabemos por el ejemplo anterior que $S^2 = 11,6$ Entonces

$$S = \sqrt{S^2}$$

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

$$S = 3,4 \text{ radiografias}$$

Ejercicio

Calcular la desviación estándar al siguiente conjunto de datos muestrales

220	215	218	210	210
219	208	207	213	225
213	204	225	211	221
218	200	205	220	215
217	209	207	211	218

- ¿Como podemos comparar la dispersión de dos variables distintas cuando la unidad de medida es diferente? (¿o cuando la media es diferente?).
- Necesitamos medidas adimensionales.

Coeficiente de Variación

El coeficiente de variación permite comparar la dispersión entre dos poblaciones distintas e incluso, comparar la variación producto de dos variables diferentes (que pueden provenir de una misma población).

Si comparamos la dispersión en varios conjuntos de observaciones tendrá menor dispersión aquella que tenga menor coeficiente de variación.

Compara la variabilidad entre dos o más grupos de datos en distintos sistemas de unidades de medida. Por ejemplo, talla, pesos en Kgmos, pesos en millones etc.

Determinar si la media es consistente con su desviación estándar.

$$C.V. = \frac{\sigma_x}{\bar{x}} \bullet 100\,\%$$

El coeficiente de variación

Muestra

$$CV = \frac{s}{\bar{x}} \times 100$$

Población

$$CV = \frac{\sigma}{\mu} \times 100$$

Ejemplo:

A continuación se presentan las tarifas (en unidades monetarias) de dos laboratorios de análisis clínicos. El laboratorio I tiene sus tarifas en soles y el laboratorio II en dólares ¿Cuál de ellos tiene un plan tarifario más homogéneo o estable?.

<u>Laboratorio I (soles)</u>	<u>Laboratorio II (dólares)</u>
40,70,60,48,52,65,58	70,35,150,140,82,110,140,120

Calculamos la media y desviación estándar por cada una de los laboratorios

Laboratorio I

$$\bar{x} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{393}{7} = 56.14$$

x	$X_i - \bar{x}\chi$	$\left(X_i - \bar{x}\right)^2$
40	-16.14	260.50
70	13.86	192.10
60	3.86	14.90
48	-8.14	66.26
52	-4.14	17.14
65	8.86	78.50
58	1.86	3.46
$\sum X = 393$	$\sum \left(X_i - \bar{x}\right) = 0$	$\sum \left(X_i - \bar{x}\right)^2 = 632,86$

Si $\sum\left(X_i-\bar{x}\right)^2=632.86$

$$S=\sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{x})^2}{n-1}}=\sqrt{\frac{632.86}{7-1}}=10.27$$

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

$$CV=\frac{10.27}{56.14}\times 100=18.29$$

Laboratorio II

$$\bar{x} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{847}{8} = 105.87$$

x	$X_i - \bar{x}\,\chi$	$\left(X_i - \bar{x}\right)^2$
70	-35.87	1286.6569
35	-70.87	5022.5569
150	44.13	1947.4569
140	34.13	1164.8569
82	-23.87	569.7769
110	4.13	17.0569
140	34.13	1164.8569
120	14.13	199.6569
$\sum X = 847$	$\sum \left(X_i - \bar{x}\right) = 0,04$	$\sum \left(X_i - \bar{x}\right)^2 = 11372,88$

$$\text{Si} \quad \sum_{i=1}^{n}\left(X_i - \bar{x}\right)^2 = 11372.88$$

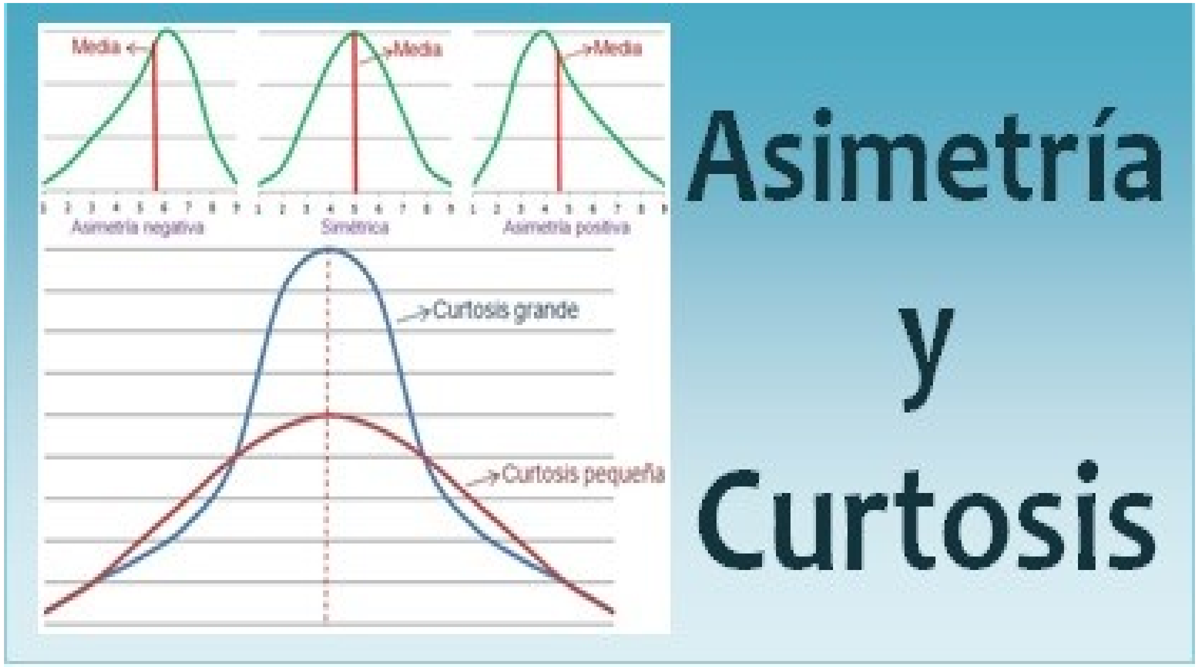
$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{x})^2}{n-1}} = \sqrt{\frac{11372,88}{8-1}} = 40.30$$

$$CV = \frac{S}{\bar{x}} \times 100 \qquad CV = \frac{40,30}{105,87} \times 100 = 30,06$$

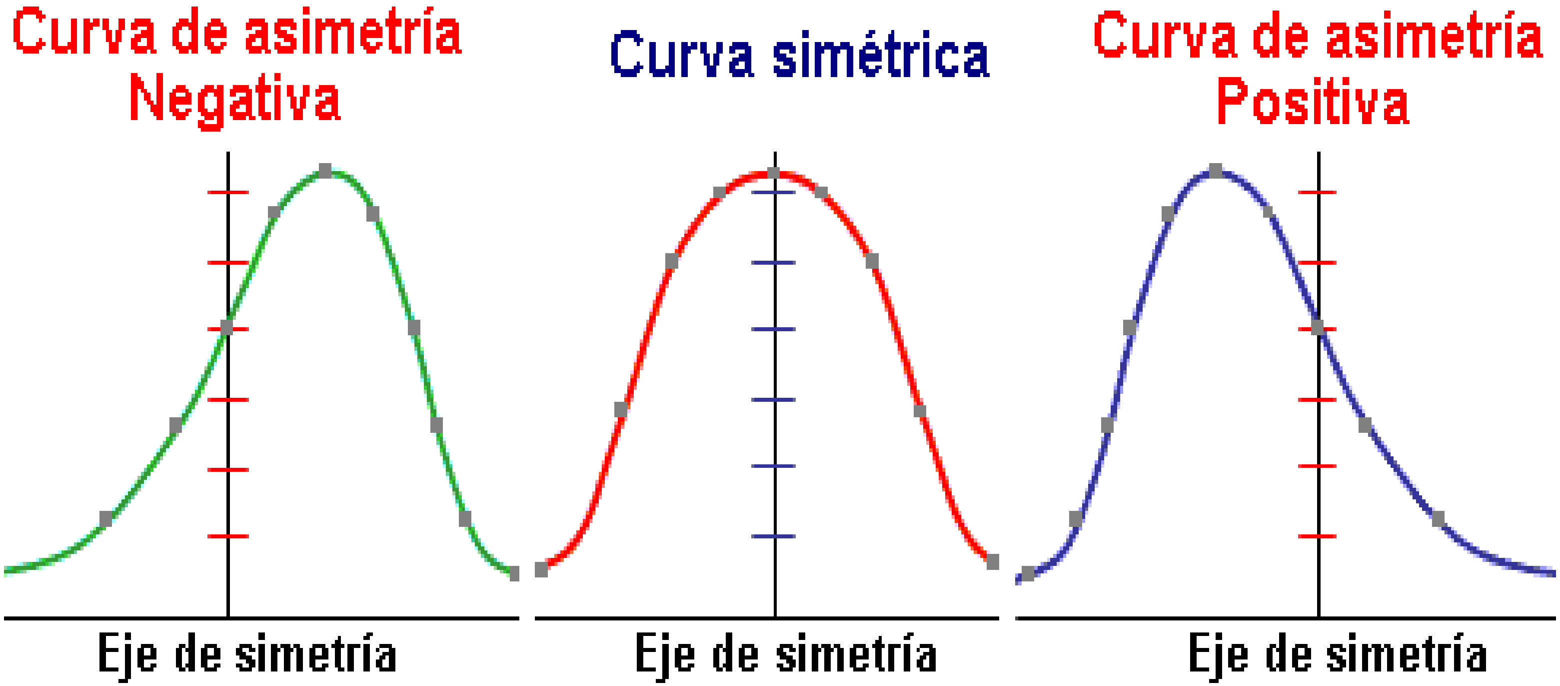
El Laboratorio II presenta una mayor variabilidad en el plan tarifario.

MEDIDAS DE FORMA

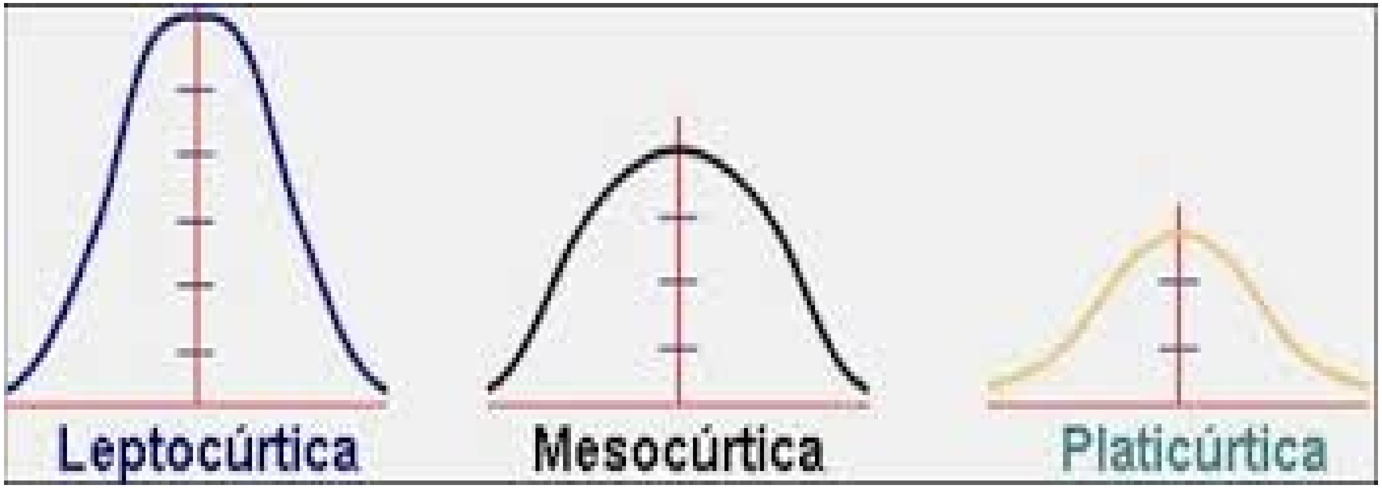
- ASIMETRIA
- CURTOSIS

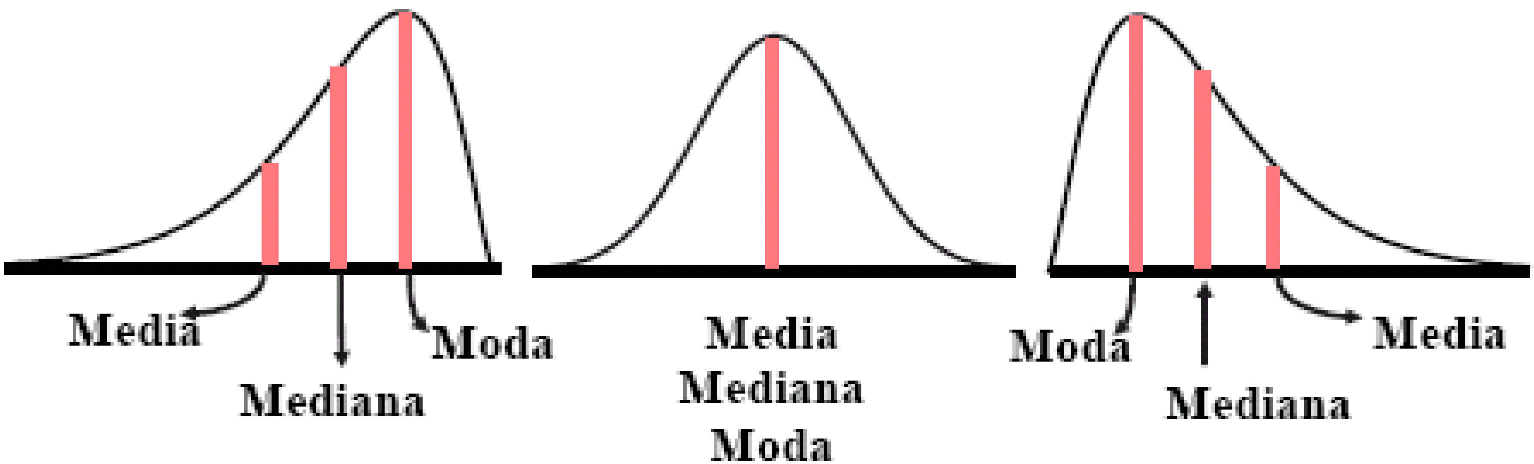


MEDIDAS DE FORMA



Curtosis





Asimétrica hacia la izquierda

Simétrica

Asimétrica hacia la derecha

Ejercicios

1. Considere la muestra 2, 4, 7, 8, 9. Encuentre:
   - Rango
   - Varianza
   - Desviación estándar
1. Dada la muestra 7, 6, 10, 7, 5, 9, 3, 7, 5, 13. Encuentre:
   - Varianza
   - Desviación estándar

Ejercicios.

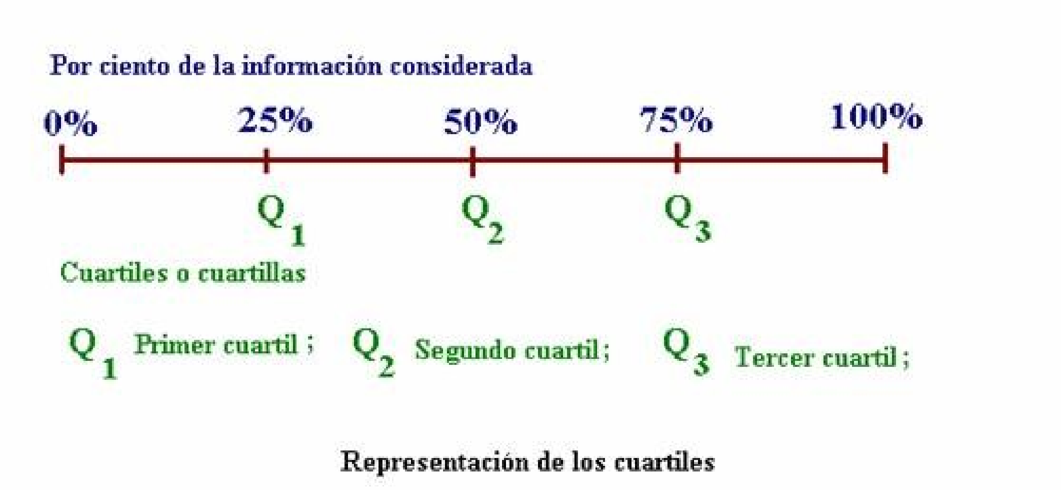
A los reclutas de una academia de policía se les solicitó presentar un examen que mide la capacidad que tienen para hacer ejercicio. Esta capacidad (medida en minutos) se obtuvo para cada uno de los 20 reclutas:

25	27	30	33	30	32	30	34	30	27
26	25	29	31	31	32	34	32	33	30

- Encuentre el rango
- Encuentre la varianza
- Encuentre la desviación estándar

Cuartiles

Los cuartiles son los **tres** valores de la variable que dividen a un conjunto de datos ordenados en **cuatro** partes iguales.



$$Q = (kn)/4 + 1/2$$

Número impar de datos

```
2, 5, 3, 6, 7, 4, 9
```

$$\begin{array}{ccccccc} 2, & 3, & 4, & 5, & 6, & 7, & 9 \\ & \downarrow & & \downarrow & & \downarrow & \\ & Q_1 & & Q_2 & & Q_3 & \end{array}$$

Número par de datos

```
2, 5, 3, 4, 6, 7, 1, 9
```

$$\begin{array}{ccccc} 1, & \underline{2,\ 3}, & \underline{4,\ 5}, & \underline{6,\ 7}, & 9 \\ & 2.5 & 4.5 & 6.5 & \\ & \downarrow & \downarrow & \downarrow & \\ & Q_1 & Q_2 & Q_3 & \end{array}$$

Los deciles son los nueve valores que dividen la serie de datos en diez partes iguales.

Los deciles dan los valores correspondientes al 10%, al 20%... y al 90% de los datos.

D5 coincide con la mediana.

$$D = (kn/10) + 1/2$$

El percentil es el valor de la variable por debajo del cual se encuentra un porcentaje dado de observaciones. Por ejemplo, el percentil 20º es el valor debajo del cual se encuentran el 20 por ciento de las observaciones.

$$P = (kn)/100 + 1/2$$

Ejercicio:

Dados los datos

16 11 14 11 8 8 5 11

Calcule:

- El primer cuartil
- El tercer decil
- Percentil 40