

## Apuntes de estadística.

### 3. ESTADÍSTICA.

#### 3.1. Conceptos básicos.

La Estadística es la parte de las Matemáticas que estudia una serie de datos para compararlos y sacar conclusiones.

**Población:** Es el conjunto total de individuos sobre los que se quiere estudiar unos datos determinados. [Por ej., si queremos estudiar la estatura media de los españoles, la población la constituirán todos los españoles. O si queremos estudiar la nota media de Matemáticas en 1º de la ESO en el Colegio Internacional SEK Alborán, la población la constituyen todos los alumnos de 1º de ESO del colegio].

**Muestra:** Cuando la población es muy grande (por ejemplo, todos los españoles) o difícil de estudiar (por ej., la calidad de las bombillas o la deformación de los coches en un choque) se elige una muestra, que es una parte de la población representativa de la misma, es decir, con unas características similares. Ha de elegirse al azar. [Por ejemplo, para estudiar la estatura de los españoles en la muestra deberíamos incluir a individuos de diferentes edades, de diferente estrato social y económico, de distinto lugar geográfico, de ambos sexos].

**Variable estadística:** Es el dato o característica que se quiere estudiar. Por ejemplo: la estatura, la calidad de las bombillas, la nota de Matemáticas, etc.

**Variable cuantitativa:** Es aquella que estudia algo que se expresa mediante números. Por ej., la estatura, la nota de Matemáticas, etc. Puede ser:

- **discreta**, si toma valores aislados (números enteros), como, por ejemplo, el número de hermanos de los alumnos de 1º de ESO. [Los valores de la variable son números enteros; un alumno no puede tener 2,3 hermanos, por ejemplo].

- **continua**, si toma todos los valores dentro de un intervalo, como por ejemplo estatura de los alumnos de 1º ESO. [Los valores de la variable pueden ser, por ejemplo, desde 1,50 metros hasta 2 metros de estatura, pudiendo tomar cualquier valor intermedio].

**Variable cualitativa:** Es aquella que estudia algo que no puede expresarse por números. Por ej., qué programa de TV se ve más, qué libro de lectura es el preferido por los alumnos de 1º de ESO, etc.

**Encuesta:** Procedimiento que nos permite obtener los datos para hacer un estudio de ellos (puede ser oral o escrita).

### Tablas de frecuencias:

En un estudio estadístico, una vez obtenidos los datos hay que recontarlos, ordenarlos y tabularlos, esto es, colocarlos en **tablas** en las que se aprecie información sobre las frecuencias de cada valor o cada cualidad de la variable. Por ejemplo, si estamos estudiando las notas de Matemáticas de 30 alumnos (**N**) de 1º de la ESO habríamos de proceder de una forma parecida a la siguiente:

<b>Variable</b> (Notas)	<b>Frecuencia</b> (Alumnos)	<b>Frecuencia</b> <b>relativa</b>	<b>Frecuencia</b> <b>acumulada</b>	<b><math>x_i \cdot n_i</math></b>
<b><math>x_i</math></b>	<b><math>n_i</math></b>	<b><math>f_i</math></b>	<b><math>N_i</math></b>	
<b>0</b>	2	0,07	2	<b>0</b>
<b>1</b>	2	0,07	4	2
<b>2</b>	1	0,03	5	2
<b>3</b>	4	0,13	9	12
<b>4</b>	2	0,07	11	8
<b>5</b>	3	0,1	14	15
<b>6</b>	3	0,1	17	18
<b>7</b>	4	0,13	21	28
<b>8</b>	3	0,1	24	24
<b>9</b>	4	0,13	28	36
<b>10</b>	2	0,07	30	20
	<b>30</b>	<b>1</b>		<b>165</b>

**Frecuencia absoluta ( $n_i$ ):** número de veces que aparece cada valor ( $x_i$ ) de la variable (p. ej., la frecuencia de la nota 7 en este ejemplo es 4). La suma de todas las frecuencias absolutas es igual al número total de datos (**N**).

**Frecuencia relativa ( $f_i$ ):** es el resultado de dividir la frecuencia absoluta entre el número total de datos (**N**):

$$f_i = \frac{n_i}{N}$$

(Así, p. ej., la nota 8 tiene una frecuencia relativa:  $f_i = \frac{3}{0,1} = 0,1$ ).

La suma de todas las frecuencias relativas es igual a la unidad.

La frecuencia relativa se puede convertir en **porcentaje** multiplicándola por 100. Así, p. ej., el porcentaje de alumnos que han obtenido un 7 en Matemáticas es del 13 % ( $0,13 \cdot 100$ )

### 3.2. Gráficos estadísticos

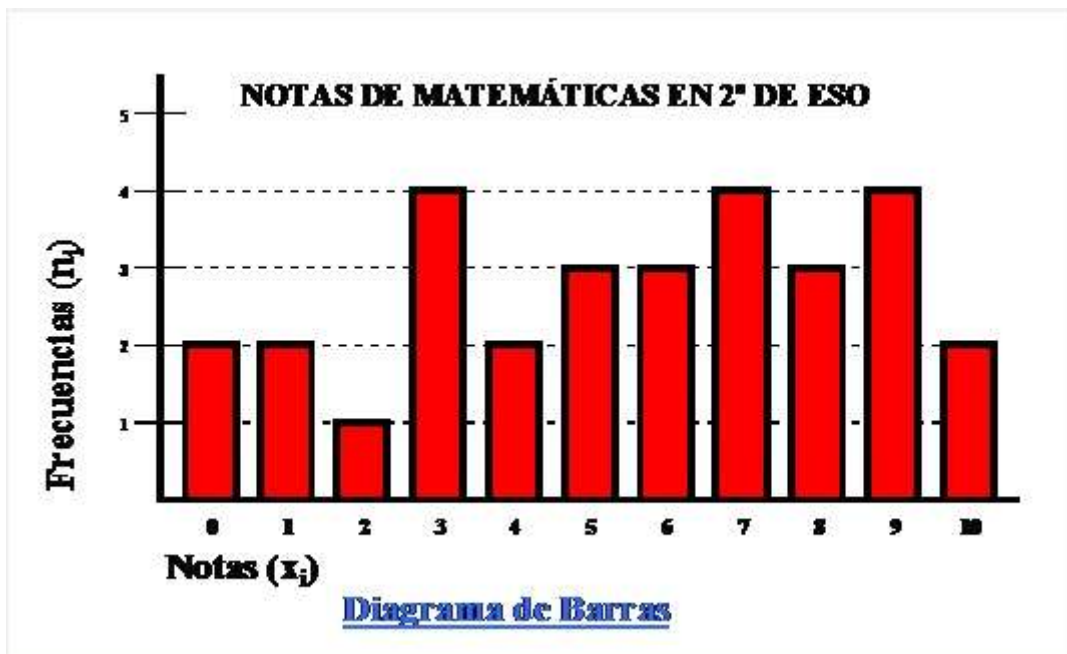
Para dar a conocer los datos de un estudio estadístico se confeccionan gráficas estadísticas.

#### Diagrama de barras:

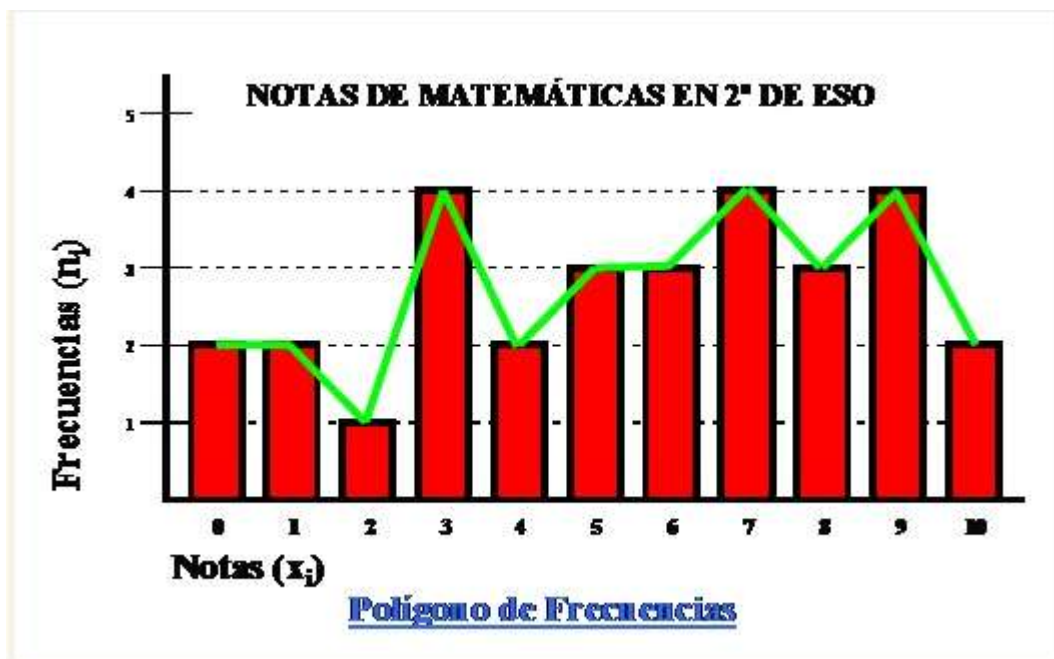
Un diagrama de barras consiste en la representación mediante barras de los valores de la variable, con una altura de la barra proporcional a su frecuencia absoluta.

Las barras se colocan en unos ejes de coordenadas: en el eje de las abscisas se ponen los valores de la variable y en el eje de ordenadas su frecuencia.

Se pueden representar variables cualitativas o cuantitativas.



Si unimos los puntos medios de los extremos de las barras por una línea obtenemos un polígono de frecuencias, que se utiliza con variables cuantitativas.



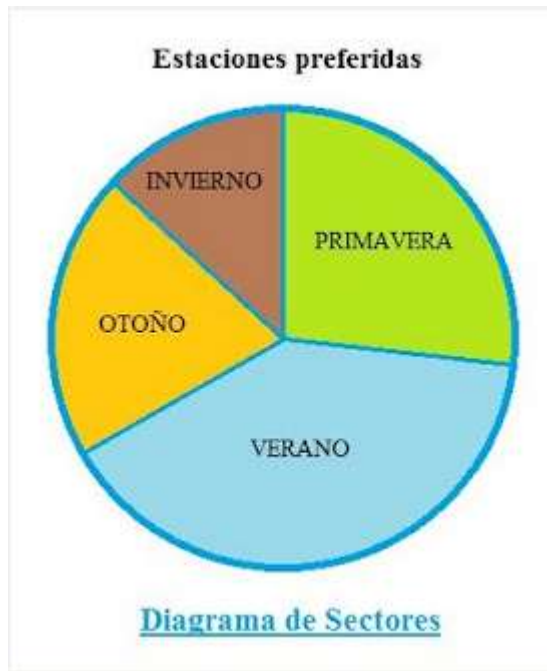
### Diagrama de Sectores:

Un **diagrama de sectores** consiste en representar los valores o cualidades de la variable en sectores circulares.

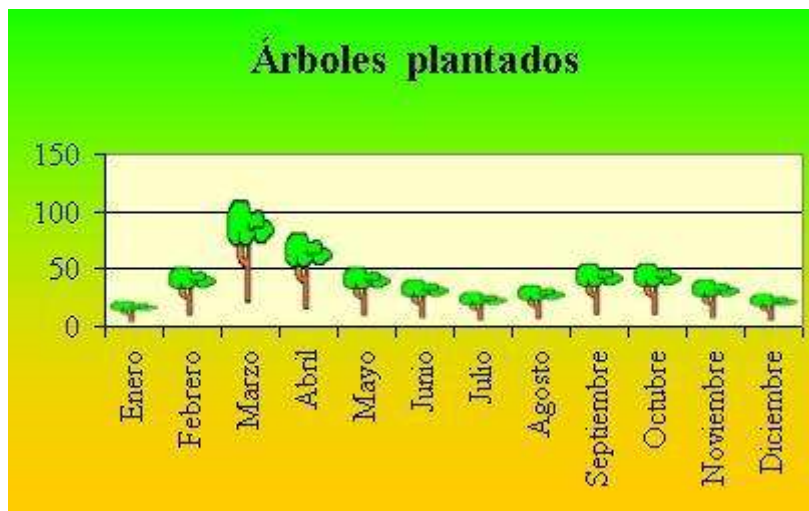
La amplitud o área de cada sector ha de ser proporcional a la frecuencia de cada valor (para ello se dividen los 360° de la circunferencia entre el número total de datos, N, para saber cuántos grados corresponden a cada dato, y el resultado se va multiplicando por cada frecuencia absoluta de los respectivos valores de la variable).

**Ejemplo:** Se ha preguntado a 30 alumnos de una clase de 1º de ESO qué estación del año preferían: Primavera (P), Verano (V), Otoño (O) o Invierno (I). Los resultados han sido éstos: P, V, V, P, O, O, V, V, I, P, I, P, I, O, V, V, V, I, O, P, P, V, V, O, O, P, V, V, V, P. Tabula los datos en una tabla de frecuencias y representa los resultados en un diagrama de sectores.

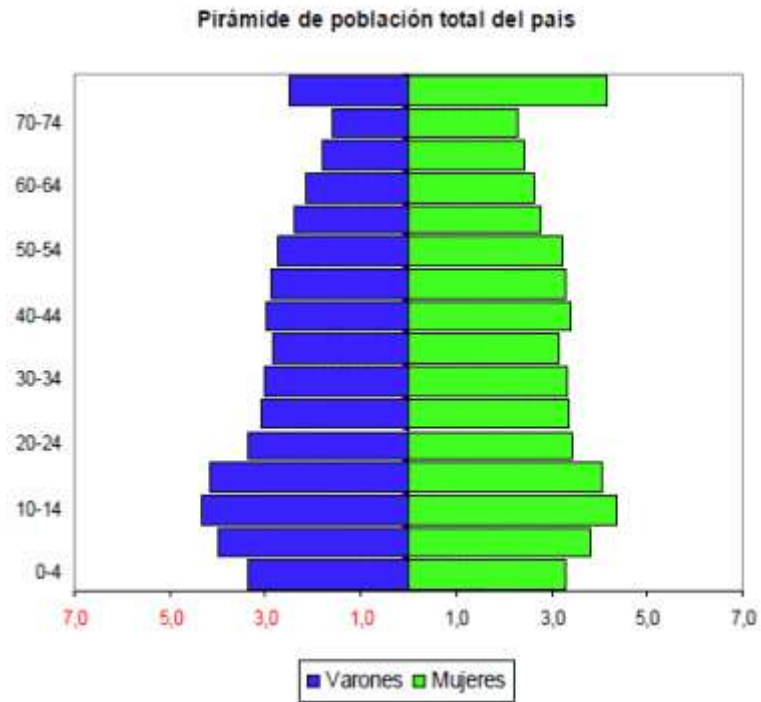
Variable (Estación del año) xi	Frecuencia ni	Amplitud de sectores (Para cada dato: $360^\circ : 30 = 12^\circ$ )
P	8	$8 \cdot 12^\circ = 96^\circ$
V	12	$12 \cdot 12 = 144^\circ$
O	6	$6 \cdot 12^\circ = 72^\circ$
I	4	$4 \cdot 12^\circ = 48^\circ$
	<b>30</b>	$30 \cdot 12^\circ = 360^\circ$



También se utilizan mucho en los medios de comunicación los **pictogramas**, que consisten en representar los datos con dibujos referidos a la variable que se estudia y de tamaño proporcional a la frecuencia de los valores de la variable.

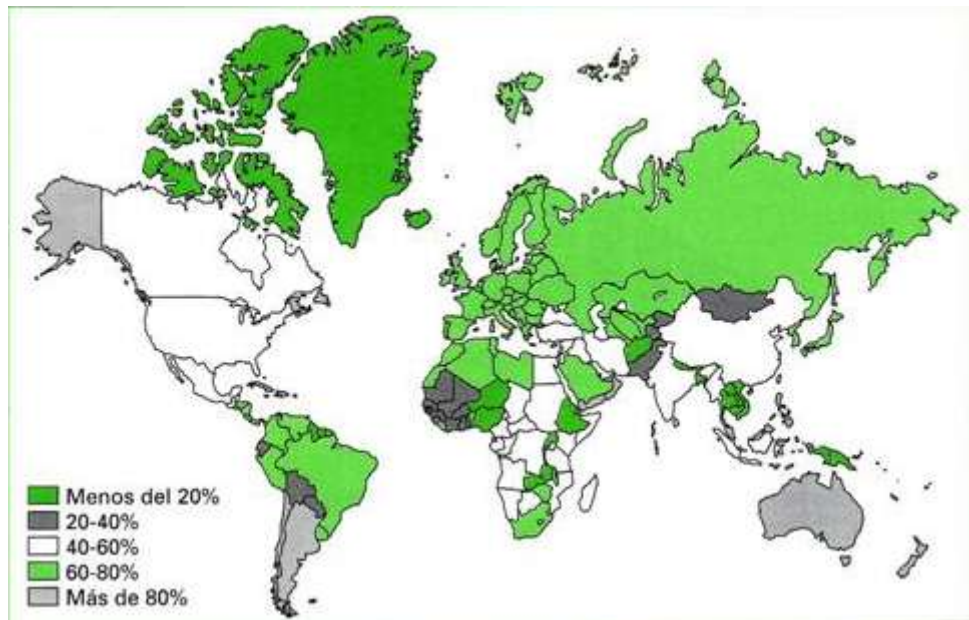


Dependiendo de la información que estemos estudiando, se pueden utilizar otros tipos de gráficos. Uno de ellos es por ejemplo, la **pirámide de población**. Sirve para analizar cómo va evolucionando (con respecto a su edad) una población determinada. Consiste en dos diagramas de barras, uno de ellos para representar los datos de los hombres y el otro para los de las mujeres, pero dispuestos de forma horizontal y por edades.



Los **cartogramas y mapas de coropletas** son gráficos realizados sobre mapas, en los que aparecen indicados sobre las distintas zonas cantidades o colores de acuerdo con el carácter que representan.

En el siguiente cartograma observamos la urbanización en el mundo atendiendo a la industrialización:



### 3.3. Parámetros estadísticos

#### 3.3.1. De centralización

En muchas ocasiones es poco práctico ofrecer todos los datos obtenidos y lo que se hace es facilitar los que resumen las características que estamos estudiando. Para ello se suelen utilizar los llamados **valores centrales**, que son: Media aritmética, Mediana y Moda.

#### **Media aritmética:**

La **media aritmética** se calcula sumando todos los valores obtenidos de la variable estudiada y dividiéndolos por el número de datos que haya. La media se representa con la letra  $x$  y una rayita encima. Lógicamente, la media sólo se puede calcular con datos cuantitativos.

**Ejemplo:** Las notas de Matemáticas de un alumno de 1º de ESO en la 3ª evaluación han sido éstas: 7, 6, 5, 8. ¿Cuál es su nota media?

$$\begin{aligned}\bar{x} &= \frac{7+6+5+8}{4} \\ x &= \frac{26}{4} \\ x &= 6,5\end{aligned}$$

Si los datos los hemos ordenado **en una tabla de frecuencias**, la media se calcula multiplicando cada valor de la variable por su frecuencia absoluta, sumando los productos obtenidos y dividiendo el resultado entre el número de datos que haya.

**Ejemplo:** Calcular la media aritmética de las notas de Matemáticas de 30 alumnos (**N**) de 1º de la ESO que aparecen en esta tabla de frecuencias:

Variable (Notas)	Frecuencia (Alumnos)	$x_i \cdot n_i$
$x_i$	$n_i$	
0	2	0
1	2	2
2	1	2
3	4	12
4	2	8
5	3	15
6	3	18
7	4	28
8	3	24
9	4	36
10	2	20
	<b>30</b>	<b>165</b>

$$\bar{x} = \frac{165}{30}$$

$$\bar{x} = 5,5$$

### Mediana:

La **mediana** es el valor que ocupa el lugar del medio, el central, después de haber ordenado todos los datos de menor a mayor.

Si el número de datos es impar, sólo habrá uno en el medio. Si es par, habrá dos que ocupen el lugar central y habremos de hallar la media aritmética de ambos.

**Ejemplo:** ¿Cuál es la mediana de las siguientes notas: 8, 7, 9, 4, 6, 8, 5, 7, 6?

Notas ordenadas: 4, 5, 6, 6, **7**, 7, 8, 8, 9.

$$M_e = 7$$

**Otro ejemplo:** ¿Cuál es la mediana de las siguientes notas: 6, 5, 8, 3, 9, 6, 5, 7, 7, 8?

Notas ordenadas: 3, 5, 5, 6, **6, 7**, 7, 8, 8, 9.

$$M_e = \frac{6+7}{2}$$

$$M_e = 6,5$$



Si los datos están **en una tabla de frecuencias**, la mediana es el primer valor cuya frecuencia acumulada sea mayor que  $N/2$ , **si N es impar**.

**Ejemplo:** Al tirar un dado 15 veces se han obtenido los valores (puntuaciones) de la siguiente tabla. Calcula la Mediana.

<b>Variable (Puntos) <math>x_i</math></b>	<b>Frecuencia <math>n_i</math></b>	<b>Frecuencia acumulada <math>N_i</math></b>
1	2	2
2	3	5
3	1	6
<b>4</b>	3	<b>9</b>
5	4	13
6	2	15
	<b>15</b>	

Como el número de datos es impar,  $N = 15$ ,  $\rightarrow \frac{N}{2} = \frac{15}{2} = 7,5$

El primer valor de la variable cuya frecuencia acumulada es mayor que 7,5 es  $x_i = 4$  ( $N_i = 9 > 7,5$ ). Por lo tanto:

$$M_e = 4$$

**Si N es par**, la mediana será la media aritmética de los primeros valores cuya frecuencia acumulada contenga a

$$\frac{N}{2} \text{ y } \frac{N}{2} + 1$$

**Ejemplo:** Al tirar un dado 16 veces se han obtenido los valores (puntuaciones) de la siguiente tabla. Calcula la Mediana.

<u>Variable</u> (Puntos) $x_i$	<u>Frecuencia</u> $n_i$	<u>Frecuencia</u> <u>acumulada</u> $N_i$
1	2	2
2	3	5
3	1	6
<b>4</b>	2	<b>8</b>
<b>5</b>	4	<b>12</b>
6	4	16
	<b>16</b>	

Como el número de datos es par,  $N = 16$ ,

$$\frac{N}{2} = 8$$

El primer valor de la variable cuya frecuencia acumulada contiene a 8 es  $x_i = 4$  ( $N_i = 8$ ) y el primer valor de la variable cuya frecuencia acumulada contiene a 9 ( $8 + 1$ ) es  $x_i = 5$  ( $N_i = 12$ , que contiene a 9). Por lo tanto:

$$Me = \frac{4+5}{2}$$

$$Me = 4,5$$

### **Moda:**

La **moda** es el valor que más se repite, que está “de moda”. Por lo tanto, es el valor de la variable que tiene mayor frecuencia absoluta. Hay moda tanto en las variables cuantitativas como en las cualitativas.

Puede haber más de una moda. Si hubiera dos modas, la distribución sería bimodal; si hubiera tres, sería trimodal; y si hubiera más de tres, sería multimodal.

**Ejemplo:** Se ha preguntado a 30 alumnos de una clase de 1º de ESO qué estación del año preferían: Primavera (P), Verano (V), Otoño (O) o Invierno (I). Los resultados han sido los recogidos en esta tabla:

Variable (Estación del año) $x_i$	Frecuencia $n_i$
P	8
V	12
O	6
I	4
	30

¿Cuál es la moda?

Respuesta: La moda es Verano (la estación preferida por los alumnos, la que mayor frecuencia absoluta tiene).

### 3.3.2. De dispersión

Estudia la distribución de los valores de la serie, analizando si estos se encuentran más o menos concentrados, o más o menos dispersos.

Existen diversas **medidas de dispersión**, entre las más utilizadas podemos destacar las siguientes:

**1.- Rango:** mide la amplitud de los valores de la muestra y se calcula por diferencia entre el valor más elevado y el valor más bajo.

**2.- Varianza:** Mide la distancia existente entre los valores de la serie y la media. Se calcula como sumatorio de las diferencias al cuadrado entre cada valor y la media, multiplicadas por el número de veces que se ha repetido cada valor. El sumatorio obtenido se divide por el tamaño de la muestra.

La varianza siempre será mayor que cero. Mientras más se aproxima a cero, más concentrados están los valores de la serie alrededor de la media. Por el contrario, mientras mayor sea la varianza, más dispersos están.

**3.- Desviación típica:** Se calcula como raíz cuadrada de la varianza.

**4.- Coeficiente de varización de Pearson:** se calcula como cociente entre la desviación típica y la media.

**Ejemplo:** vamos a utilizar la serie de datos de la estatura de los alumnos de una clase (lección 2ª) y vamos a calcular sus medidas de dispersión.

Variable (Valor)	Frecuencias absolutas		Frecuencias relativas	
	Simple	Acumulada	Simple	Acumulada
1,20	1	1	3,3%	3,3%
1,21	4	5	13,3%	16,6%
1,22	4	9	13,3%	30,0%
1,23	2	11	6,6%	36,6%
1,24	1	12	3,3%	40,0%
1,25	2	14	6,6%	46,6%
1,26	3	17	10,0%	56,6%
1,27	3	20	10,0%	66,6%
1,28	4	24	13,3%	80,0%
1,29	3	27	10,0%	90,0%
1,30	3	30	10,0%	100,0%

1.- **Rango:** Diferencia entre el mayor valor de la muestra (1,30) y el menor valor (1,20). Luego el rango de esta muestra es 10 cm.

2.- **Varianza:** recordemos que la media de esta muestra es 1,253. Luego, aplicamos la fórmula:

$$S_x^2 = \frac{((1,20-1,253)^2 * 1) + ((1,21-1,253)^2 * 4) + ((1,22-1,253)^2 * 4) + \dots + ((1,30-1,253)^2 * 3)}{30}$$

Por lo tanto, la varianza es 0,0010

3.- **Desviación típica:** es la raíz cuadrada de la varianza.

$$\sigma = (S_x^2)^{(1/2)}$$

Luego:

$$\sigma = (0,010)^{(1/2)} = 0,0320$$

4.- **Coefficiente de variación de Pearson:** se calcula como cociente entre la desviación típica y la media de la muestra.

$$Cv = 0,0320 / 1,253$$

Luego,

$$Cv = 0,0255$$

El interés del coeficiente de variación es que al ser un porcentaje permite comparar el nivel de dispersión de dos muestras. Esto no ocurre con la desviación típica, ya que viene expresada en las mismas unidades que los datos de la serie.

Por ejemplo, para comparar el nivel de dispersión de una serie de datos de la altura de los alumnos de una clase y otra serie con el peso de dichos alumnos, no se puede utilizar las desviaciones típicas (una viene vienes expresada en cm y la otra en kg). En cambio, sus coeficientes de variación son ambos porcentajes, por lo que sí se pueden comparar.