ESTADÍSTICA 4º ESO

A) INICIACIÓN A LA ESTADÍSTICA

1.- ¿QUÉ ES LA ESTADÍSTICA?

La **Estadística** es la rama de las Matemáticas que se ocupa del estudio de una determinada característica en una población, recogiendo los datos, organizándolos en tablas, representándolos gráficamente y analizándolos para sacar conclusiones de dicha población.

Hoy día la Estadística ha adquirido gran importancia como herramienta para el desarrollo de multitud de disciplinas científicas; además el alumnado que siga distintos estudios universitarios, en un futuro próximo, va a encontrarse con asignaturas relacionadas con la Estadística.

Por otra parte, su utilización en la vida cotidiana se ha popularizado tanto que constituye un vehículo de comunicación usual y casi imprescindible. Por ello, se quiere presentar al alumnado la Estadística como un elemento auxiliar básico para la investigación experimental de cara a una posible especialización universitaria (Económicas, Biología, Sociología, Ingenierías, Medicina, ...) o profesional y a la vez aportar las claves necesarias para comprender los elementos esenciales de una investigación estadística, prevenir ante posibles abusos de la estadística (presentes a veces en los medios de comunicación) y comprender mejor la naturaleza y el significado de los diferentes indicadores sociales que ayuden a formar una visión fundamentada de la panorámica social en un determinado momento.

Con esta materia se aborda el estudio de la Estadística como saber estratégico, como herramienta procedimental para la investigación científica y tecnológica, y como campo de conocimiento imprescindible para la descripción de fenómenos sociales y culturales. Puede tomar distintos aspectos según el itinerario de las Modalidades de Bachillerato a las que se oferta. Por ejemplo en el itinerario de Ciencias puede ayudar en el perfeccionamiento de métodos de investigación experimental; en Tecnología industrial, a resolver problemas de control de calidad; en Ciencias de la Salud, contribuir al conocimiento de investigaciones médicas y farmacológicas; en Ciencias Sociales, profundizar en el estudio sobre la población e indicadores sociales.

Las técnicas estadísticas se aplican de manera amplia en mercadotecnia, contabilidad, control de calidad y en otras actividades; estudios de consumidores; análisis de resultados en deportes; administradores de instituciones; en la educación; organismos políticos; médicos; y por otras personas que intervienen en la toma de decisiones.

2.- MÉTODO ESTADÍSTICO

El conjunto de los métodos que se utilizan para medir las características de la información, para resumir los valores individuales, y para analizar los datos a fin de extraerles el máximo de información, es lo que se llama **métodos estadísticos**. Los métodos de análisis para la información cuantitativa se pueden dividir en los siguientes seis pasos:

1. Definición del problema.
2. Recopilación de la información existente.
3. Obtención de información original.
4. Clasificación.
5. Presentación.
6. Análisis.

Según se haga el estudio sobre todos los elementos de la población o sobre un grupo de ella, vamos a diferenciar dos tipos de Estadística:

Estadística descriptiva. Realiza el estudio sobre la población completa, observando una característica de la misma y calculando unos parámetros que den información global de toda la población.

Estadística inferencial. Realiza el estudio descriptivo sobre un subconjunto de la población llamado muestra y, posteriormente, extiende los resultados obtenidos a toda la población.

3.- LENGUAJE ESTADÍSTICO. Conceptos básicos.

Es obvio que todo estudio estadístico ha de estar referido a un conjunto o colección de personas o cosas. Este conjunto de personas o cosas es lo que denominaremos **población o universo**.

Las personas o cosas que forman parte de la población se denominan **elementos o individuos**. En sentido estadístico un elemento puede ser algo con existencia real, como un automóvil o una casa, o algo más abstracto como la temperatura, un voto, o un intervalo de tiempo.

Carácter: Es el aspecto, fenómeno, rasgo o cualidad que se va estudiar en cada uno de los individuos de la población. Así por ejemplo si consideramos como elemento a una persona, podemos distinguir en ella los siguientes caracteres: sexo, edad, nivel de estudios, profesión, peso, altura, color de pelo, etc.

Por tanto de cada elemento de la población podremos estudiar uno o más aspectos cualidades o caracteres.

La población puede ser según su **tamaño** de dos tipos:

Población finita: cuando el número de elementos que la forman es finito, por ejemplo el número de alumnos de un centro de enseñanza, o grupo clase.

Población infinita: cuando el número de elementos que la forman es infinito, o tan grande que pudiesen considerarse infinitos. Por ejemplo, si se realizase un estudio sobre los productos que hay en el mercado, hay tantos y de tantas calidades que esta población podría considerarse infinita.

Ahora bien, normalmente en un estudio estadístico, no se puede trabajar con todos los elementos de la población sino que se realiza sobre un subconjunto de la misma. Este subconjunto es la **muestra**.

Los caracteres de un elemento pueden ser de muy diversos tipos, por lo que los podemos clasificar en dos grandes clases:

Variables Cualitativas (o atributos): si las distintas modalidades de los individuos no son medibles numéricamente, como la profesión de una persona.

Variables cuantitativas: son las que se describen por medio de números, como por ejemplo el peso, altura, edad, número de suspensos…

A su vez este tipo de variables se puede dividir en dos subclases:

Cuantitativas discretas. Aquellas a las que se les puede asociar un número entero, es decir, aquellas que por su naturaleza no admiten un fraccionamiento de la unidad, por ejemplo número de hermanos, páginas de un libro, etc.

Cuantitativas continuas: Aquellas que no se pueden expresar mediante un número entero, es decir, aquellas que por su naturaleza admiten que entre dos valores cualesquiera la variable pueda tomar cualquier valor intermedio, por ejemplo peso, tiempo. etc. No obstante en muchos casos el tratamiento estadístico hace que a variables discretas las trabajemos como si fuesen continuas y viceversa.

B) TABLAS Y GRÁFICOS ESTADÍSTICOS

1.- RECUENTO Y AGRUPAMIENTO DE DATOS

Una vez que hemos observado y recogidos los datos, bien a través de encuestas, bien con bases de datos ya almacenados, debemos resumir la información de forma adecuada y útil para su posterior estudio.

Atendiendo al problema que estemos estudiando, realizamos el agrupamiento de datos de una forma u otra:

- Si la variable es **cualitativa**, observamos y contamos el número de individuos de la población que presentan cada una de las distintas modalidades del carácter.

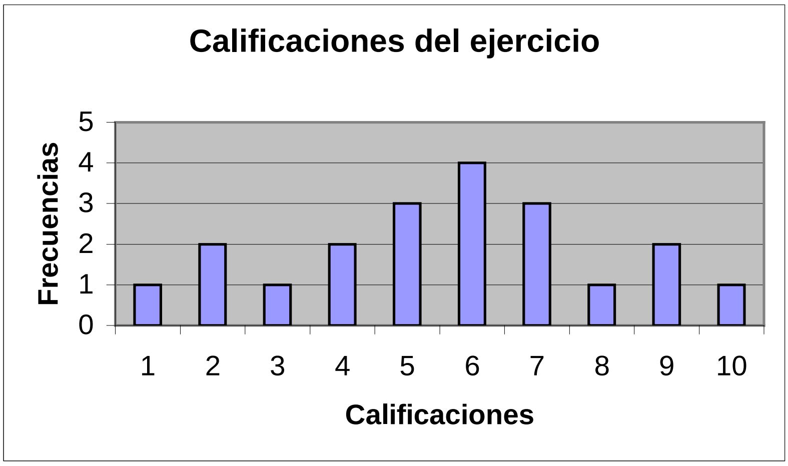
Ejemplo 1: *Un estudio hecho en un conjunto de 24 personas con objeto de determinar su grupo sanguíneo ha conducido a los siguientes resultados:*

A, B, A, A, A, AB, O, A, A, A, O, B, O, A, B, O, B, O, A, B, B, A, A, O.

Modalidad	**Recuento**	**Frecuencia (nº de veces que aparece)**
A	||||| ||||| |	11
B	||||| |	6
O	||||| |	6
AB	|	1

- Si la variable es **cuantitativa discreta,** observamos y contamos el número de individuos de la población que presentan cada uno de los distintos valores del carácter o variable, si son pocos valores.

<u>Ejemplo 2:</u> *Las calificaciones obtenidas en un ejercicio se recogen en el siguiente gráfico:*



Calificaciones	**Recuento**	**Frecuencia**
1	|	1
2	||	2

3	|	1
4	||	2
5	|||	3
6	||||	4
7	|||	3
8	|	1
9	||	2
10	|	1

- Si la variable es **cuantitativa continua** sus valores se distribuyen en intervalos de clase. Al agrupar los valores de una variable en intervalos de clase, se simplifica su descripción, pero se pierde precisión: cuánto menor sea el número de intervalos que se consideren, menor será la precisión, y viceversa.

Ejemplo 3: S*i a un grupo de 30 personas le preguntamos el dinero que en ese momento llevan encima (en €), nos encontramos con los siguientes datos:*

45	115	25	30	17	8	2	268	60	78	159	230	500	120	10
5	18	20	67	50	37	150	20	98	18	12	31	42	56	110

Evidentemente, la variable estadística tiene un recorrido (diferencia entre el mayor y menor valor) muy grande, 498 euros, por lo que si queremos hacer una tabla con estos datos tendremos que tomar intervalos. Para decidir la amplitud de los intervalos, necesitaremos decidir cuántos intervalos queremos. Normalmente se suele trabajar con no más de 10 o 12 intervalos.

Amplitud =498/10 = 49,8, por lo que tomaremos intervalos de amplitud 50

Procuraremos que en la decisión de intervalos los valores observados no coincidan con los valores de los extremos del intervalo y si esto ocurre que no sea en más de un 5% del total de observaciones.

[L_{i-1} , L_i)	**Recuento**	**Frecuencia**
[0,50)	||||| ||||| |||||		16
[50, 100)	|||||		6
[100,150)	|||	3
[150, 200)	||	2
[200, 250)			1
[250, 300)			1
[300, 350)		0
[350, 400)		0
[400, 450)		0
[450, 500)		0
[500,550)			1

2.- TABLAS ESTADÍSTICAS

Una vez realizado el recuento de datos interesa recoger la información de la muestra resumida en una tabla en la que a cada valor de la variable se le asocian determinados números que representan el número de veces que ha aparecido, su proporción con respecto a otros valores de la variable, etc. Estos números se denominan **frecuencias:** Así tenemos los siguientes tipos de frecuencia:

Frecuencia absoluta: La frecuencia absoluta de una variable estadística es el número de veces que aparece en la muestra dicho valor de la variable, la representaremos por $\mathbf{n_i}$

Frecuencia relativa: La frecuencia absoluta, es una medida que está influida por el tamaño de la muestra, al aumentar el tamaño de la muestra aumentará también el tamaño de la frecuencia absoluta. Esto hace que no sea una medida útil para poder comparar. Para esto es necesario introducir el concepto de *frecuencia relativa*, que es el cociente entre la frecuencia absoluta y el tamaño de la muestra. La denotaremos por $\mathbf{f_i}$

$$f_i = \frac{n_i}{N}$$ N es el tamaño de la muestra.

Porcentaje: La frecuencia relativa es un *tanto por uno*, sin embargo, hoy día es bastante frecuente hablar siempre en términos de *tantos por ciento o porcentajes*, por lo que esta medida resulta de multiplicar la frecuencia relativa por 100. La denotaremos por $\mathbf{p_i}$. $p_i = f_i \cdot 100$

Frecuencia Absoluta Acumulada: Para poder calcular este tipo de frecuencias hay que tener en cuenta que la variable estadística ha de ser cuantitativa. Es el número de veces que ha aparecido en la muestra un valor menor o igual que el de la variable y lo representaremos por $\mathbf{N_i}$.

$$N_i = \sum_{k=1}^{i} n_k$$

Frecuencia Relativa Acumulada: Al igual que en el caso anterior la frecuencia relativa acumulada es la frecuencia absoluta acumulada dividido por el tamaño de la muestra, y la denotaremos por $\mathbf{F_i}$

$$F_i = \frac{N_i}{N}$$

Ejemplo 4: Se le ha preguntado a 50 familias el número de miembros activos (personas que trabajan); una vez realizado el recuento se ha confeccionado la tabla de frecuencias así:

Personas Activas	**Número Familias**				
X_i	n_i	f_i	p_i	N_i	F_i
1	16	16/50	32%	16	16/50
2	20	20/50	40%	36	36/50
3	9	9/50	18%	45	45/50

4	5	5/50	10%	50	50/50
Total	50	1	100%		

Ejercicio 1.- Realiza las tablas de frecuencias de los ejemplos 1, 2 y 3.

Ejercicio 2.- Las notas obtenidas por un curso en un control de Matemáticas han sido: 2, 3, 4, 3, 5 5, 6, 5, 4, 3 2, 6, 7, 7, 5 8, 8, 9, 3, 4 4, 5, 6, 5, 4

Distribúyelas en una tabla de frecuencias.

Ejercicio 3.- Estudiando el número de hijos de 30 familias elegidas al azar en una ciudad se han obtenido los siguientes datos:

1, 2, 3, 5, 6, 0, 7, 8, 4, 1, 3, 4, 5, 2, 6, 5, 2, 3, 4, 6, 2, 3, 4, 6, 4, 3, 6, 6, 3, 3

Distribuye los datos en una tabla de frecuencias.

Ejercicio 4.- Las puntuaciones obtenidas por 30 personas de un instituto en un test de inteligencia han sido:

100, 102, 98, 95, 92, 105, 121, 110, 84, 87, 94, 99, 98, 112, 123,

145, 116, 93, 89, 86, 97, 114, 127, 103, 104, 135, 128, 109, 110, 85

Agrupa los datos en intervalos de clase y construye una tabla completa de frecuencias (con sus respectivas marcas de clase).

3.- TABLAS ESTADÍSTICAS CON LA HOJA DECÁLCULO (Variables discretas) (Mediante GeoGebra)

Hemos anotado el número de hermanos y hermanas que tiene el alumnado de dos clases de un Colegio. Queremos construir la tabla de frecuencias.

1 3 1 4 2 1 2 1 3 2

2 1 3 1 0 2 3 2 1 1

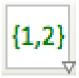
3 2 0 4 2 1 0 1 2 3

1 1 4 2 1 3 1 2 3 2

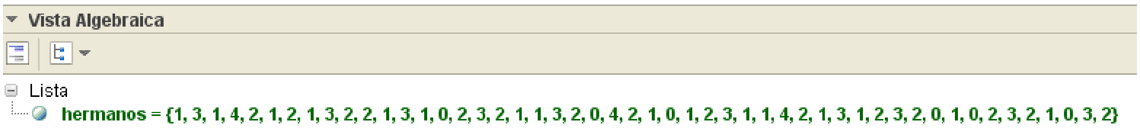
0 1 0 2 3 2 1 0 3 2

En GeoGebra accedemos a **Hoja de cálculo** a través del menú disponible en **Vista**. Introducimos los 50 datos anteriores.

	A	B	C	D	E	F	G	H	I	J
1	1	3	1	4	2	1	2	1	3	2
2	2	1	3	1	0	2	3	2	1	1
3	3	2	0	4	2	1	0	1	2	3
4	1	1	4	2	1	3	1	2	3	2
5	0	1	0	2	3	2	1	0	3	2

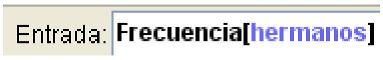
Seleccionamos ahora todas las celdas de la hoja y usamos la herramienta 

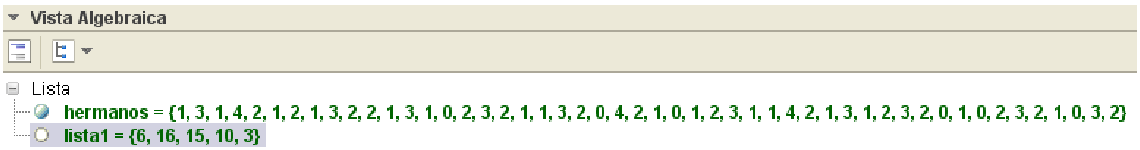
para crear una lista de datos a la que llamaremos **hermanos**. Debemos tener activa la **Vista Algebraica** para que aparezca dicha lista en ella.



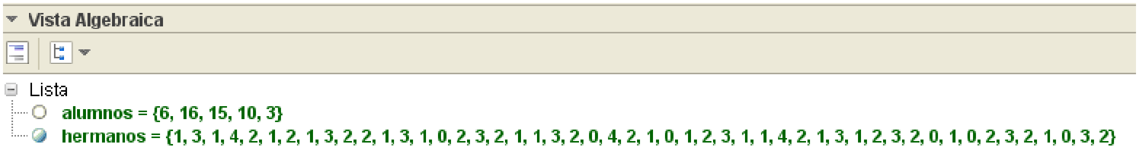
Para obtener los valores de las frecuencias absolutas usamos la barra de entrada en la forma

Escribimos el nombre de nuestra lista

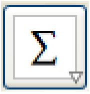




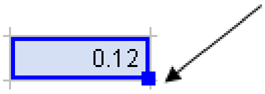
Cambiamos el nombre de la **lista1** con el botón derecho y usando **Renombra.**



GeoGebra presenta las frecuencias con datos ordenados de la lista *hermanos* de menor a mayor, o sea, hay 6 alumnos con 0 hermanos, 16 con uno, etc.

Ahora vamos a construir una tabla de frecuencias al "estilo clásico" pero ayudándonos con GeoGebra. Para ello podemos abrir una **Nueva Ventana** para tener una nueva **Hoja de cálculo** o seguir trabajando en la ventana actual para tener los datos a la vista. Haremos esto último. A partir de la celda A7 ponemos los datos (número de hermanos) y en la columna de la derecha, a partir de B7, las frecuencias absolutas (número alumnos con ese número de hermanos). En B12 escribimos **="N = " Suma[B7:B11]** para presentar la suma. (Atención a las comillas y a los espacios. Todo lo que va entre comillas es texto, lo demás son fórmulas). También podríamos haber resaltado todos los datos de la columna y usar la herramienta de la Hoja de cálculo  .

A partir de la columna C7 ponemos las frecuencias relativas escribiendo **=B7 / 50** y arrastrando el botón de control de la celda para rellenar el resto de celdas.



Por último para tener una columna de porcentajes escribimos en la celda D7 la siguiente expresión **= C7 *100 " %"** y volvemos a arrastrar su botón de control.

Nos debe quedar algo como ésto:

6	Nº hermanos	F. absolutas	F. relativas	Porcentajes
7	0	6	0.12	12 %
8	1	16	0.32	32 %
9	2	15	0.3	30 %
10	3	10	0.2	20 %
11	4	3	0.06	6 %
12		N = 50	Total = 1	

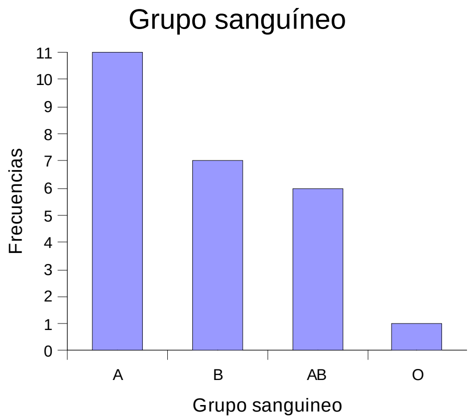
4.- GRÁFICOS ESTADÍSTICOS

Una vez recogida la información en tablas, puede ser útil resumir dicha información a través de gráficos; debemos conseguir que con un simple vistazo se nos presente la mayor información posible.

Según el carácter que estemos estudiando (cualitativo, cuantitativo discreto o cuantitativo continuo) utilizaremos distintos tipos de gráficos.

A) Diagrama de barras o rectángulos. Consiste en dos ejes perpendiculares y una barra o rectángulo para cada valor de la variable. Normalmente, se suele colocar en el eje horizontal los valores de la variable (aunque también se puede hacer en el vertical). El otro eje se gradúa según los valores de las frecuencias. La representación gráfica consiste en dibujar una barra o un rectángulo para cada uno de los valores de la variable de altura igual a su frecuencia.

El diagrama de barras asociado al ejemplo 1 sería de este tipo:

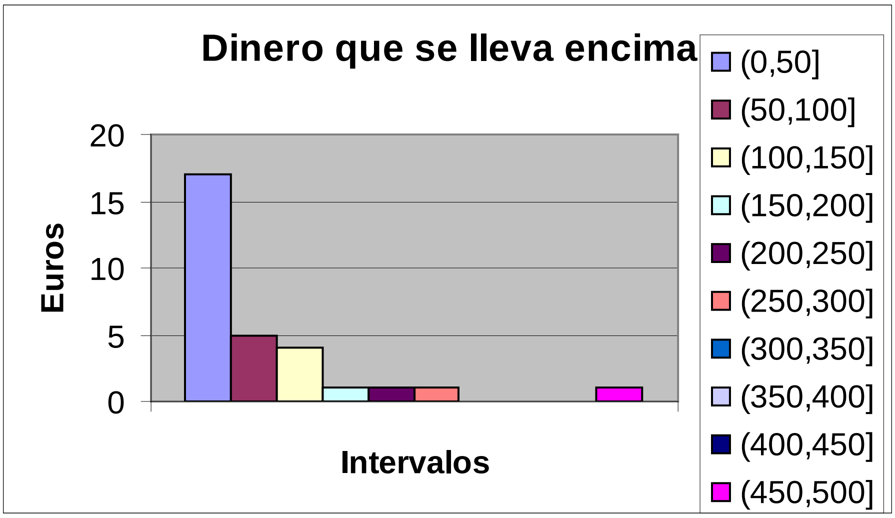


Ejercicio 5.- Construye un diagrama de barras para los datos del *ejercicio 2.*

Ejercicio 6.- Recupera la tabla del *ejercicio 4* e inserta un diagrama de barras.

B) Histograma de frecuencias. Es un caso particular del diagrama anterior en el caso de variables continuas. Si los intervalos son correlativos, los rectángulos aparecen pegados en la representación gráfica. En caso de que la amplitud de los intervalos no se igual para todos, hay que hacer coincidir el área del rectángulo con la frecuencia del intervalo.

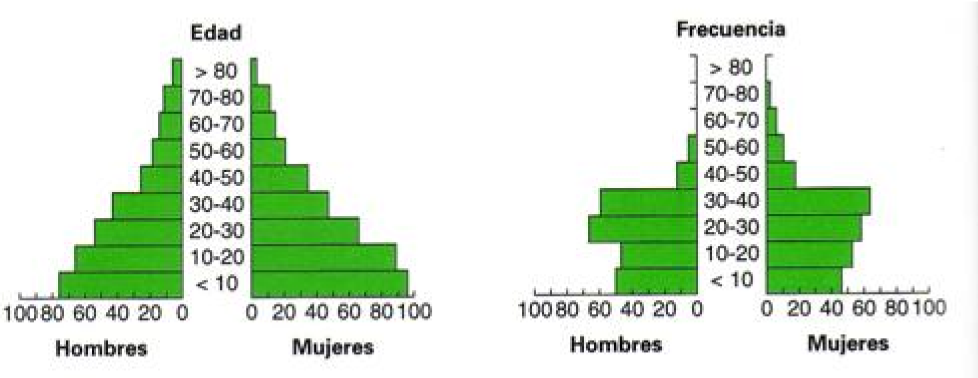
El histograma asociado al ejemplo 3 es:



Este gráfico se ha realizado con la hoja de cálculo Excel de Microsoft Office.

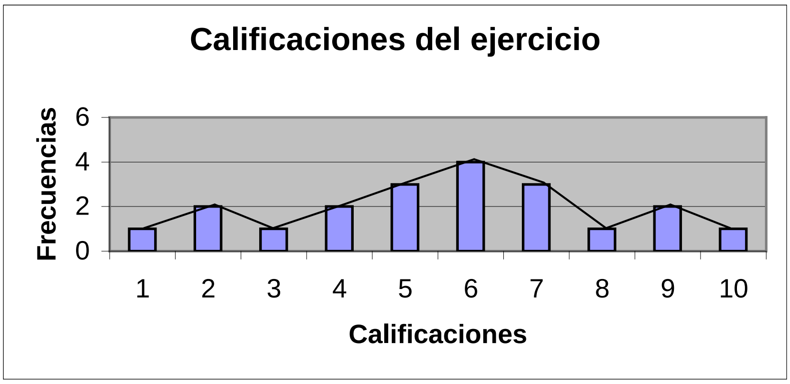
Ejercicio 7.- Recupera la tabla del *ejercicio 5* e inserta un histograma.

C) **Pirámides de población.** Cuando se realizan representaciones correspondientes a edades de población, cambiamos el eje Y por el eje X para obtener las llamadas pirámides de población, que no son más que 2 histogramas a izquierda y derecha, para hombres y mujeres. Veamos un ejemplo:

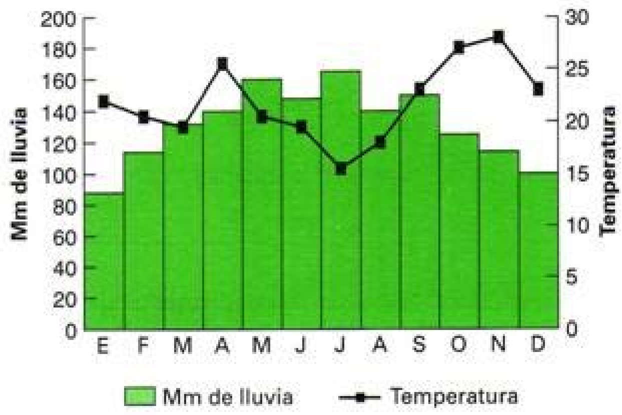


D) **Polígono de frecuencias**. Los polígonos de frecuencias son líneas poligonales que unen los vértices superiores de las barras de un diagrama de barras o los puntos medios de las bases superiores de los rectángulos de un histograma, según sea la variable agrupada o no agrupada.

El polígono de frecuencia asociado al ejemplo 2 es:



Un caso particular de aplicación de los histogramas y los polígonos de frecuencias es el **climograma**, que representa la marcha anual de las temperaturas y de las lluvias medias, sobre un mismo sistema de coordenadas. Veamos un ejemplo:



E) Diagrama de sectores. Es un gráfico estadístico formado por un círculo dividido en sectores circulares cuyas amplitudes son proporcionales a las frecuencias de los datos representados.

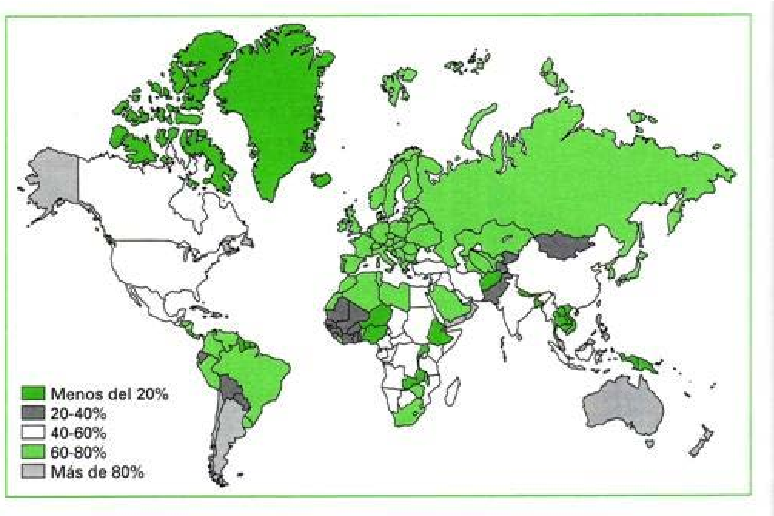
El diagrama de sectores asociado al ejemplo 1 es:

F) Pictogramas. Son gráficos con dibujos alusivos al carácter que se está estudiando y cuyo tamaño es proporcional a la frecuencia que representan; dicha frecuencia se suele representar.

En el siguiente ejemplo hemos representado el número de partidos ganados, perdidos o empatados de un equipo:

G) Cartogramas. 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



Ejercicio 8.- Los resultados finales de una evaluación de matemáticas han sido:

S, S, S. B, I   S, I, B, N, N   S, S, I, I, I   S, S, S, Sb, N   N, N, S, I, S   S, B, B.

a) ¿De qué tipo es esta variable?
b) Construye una tabla con las frecuencias absolutas, relativas y porcentuales.
c) Realiza un diagrama de barras que describa la distribución anterior.
d) Elabora un diagrama de sectores.
e) Elabora un diagrama de sectores que muestre los alumnos que promocionan y los que no.

Ejercicio 9.- Los siguientes datos corresponden a la superficie, en metros cuadrados, de 26 viviendas elegidas al azar en una localidad:

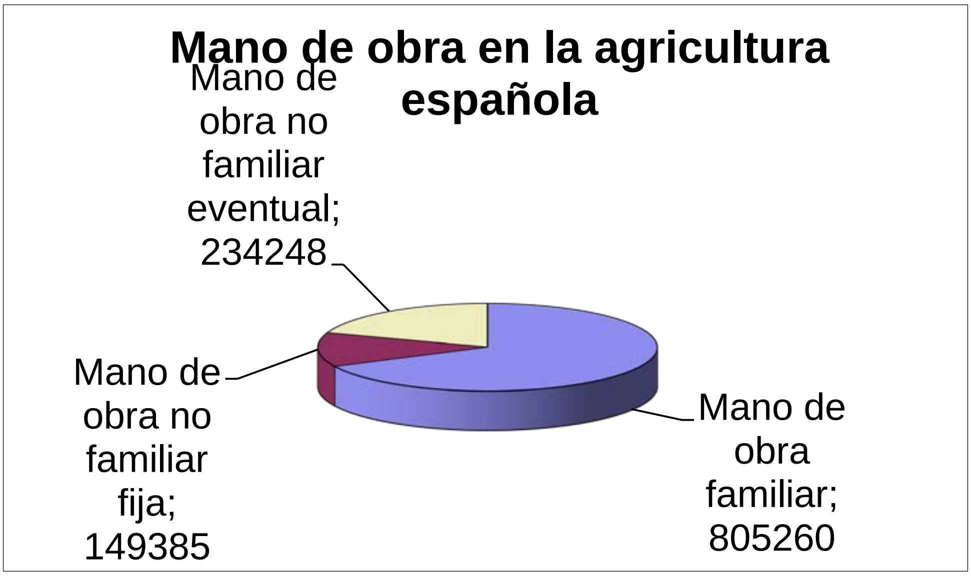
124, 100’25, 67, 78’56, 83   154’76, 59’89, 105’4, 124, 100’5   95, 98’75, 89’35, 87

135, 126’78, 105, 197, 156’8   134’25, 67’45, 74’72, 58’9, 50’57, 73, 74’57

a) Distribuye los datos en tres intervalos que clasifiquen las viviendas según su superficie entre 50 m^2 y 90 m^2, entre 90 m^2 y 120 m^2, y entre 120 m^2 y 200 m^2.

b) Elabora una tabla de frecuencias y represéntala mediante un histograma y un diagrama de sectores.

Ejercicio 10.- El siguiente diagrama de sectores muestra la procedencia de la mano de obra en la agricultura española durante el año 1999. Con los datos del diagrama construye una tabla de frecuencias absolutas, relativas y porcentuales y halla el valor de los ángulos de cada sector circular.



Ejercicio 11. Lanza un dado 40 veces y anota los resultados. Después haz un recuento y organiza los datos en una hoja de cálculo y construye una tabla de frecuencias en la que aparezcan los valores de la variable, la frecuencia absoluta, la frecuencia relativa y los porcentajes correspondientes.

Ejercicio 12.-

a) Localiza en la página del Instituto de Estadística y Cartografía de Andalucía http://www.juntadeandalucia.es/institutodeestadisticaycartografia el apartado **Padrón Municipal de Habitantes. Cifras Oficiales de Población Municipal** y en él consulta Último año disponible 2015

b) Abre el apartado 3.1. Población por municipio de residencia y sexo. Guarda la tabla (expórtala a Excel o Calc) de 3.1.9. Andalucía; ábrela con la hoja de cálculo y distribuye todos los municipios en intervalos: (0,2000), (2000,5000), (5000,10000), (10000 ,20000), (20000, 50000), (50000, 100000) y 100000 en adelante. c) Confecciona un gráfico adecuado.

C) MEDIDAS ESTADÍSTICAS

Ante la gran cantidad de datos que se utilizan al realizar un estudio estadístico se hace necesario resumir toda la información en unos pocos datos. En esta unidad estudiaremos las **medidas estadísticas** que nos ayudan a emitir conclusiones sobre las poblaciones y a hacer comparaciones entre ellas.

1.- MEDIDAS DE CENTRALIZACIÓN

Son medidas estadísticas que buscan características del centro de la distribución: **media, moda y mediana.**

a) **Media**: Se define la media aritmética de los valores x_i de una distribución con frecuencias absolutas n_i como:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i \cdot n_i}{n}$$

siendo n la suma de las frecuencias absolutas.

Propiedades:

- Es el centro de gravedad de la distribución, siendo única en cada distribución.
- Si aparecen valores extremos y poco significativos, la media puede no ser representativa.
- No tiene sentido en el caso de un carácter cualitativo ni en el caso de datos agrupados con algún intervalo no acotado.
- Si se suma una constante a todos los valores de una variable, su media aumenta en dicha constante.
- Si se multiplican todos los valores de la variable por una constante, la media queda multiplicada por dicha constante.

Ejemplo 1: La siguiente tabla muestra las notas en un examen de 25 personas:

Nota	**n_i**
3	3
4	4
5	6
6	6
7	4
8	2
Suma	**25**

Su nota media es: $\bar{x} = \dfrac{3\cdot 3 + 4\cdot 4 + 5\cdot 6 + 6\cdot 6 + 7\cdot 4 + 8\cdot 2}{25} = \dfrac{135}{25} = 5'4$

Se podría confeccionar la tabla así:

Nota	**n_i**	**$x_i \, . \, n_i$**
3	3	9
4	4	16
5	6	30
6	6	36
7	4	28
8	2	16
Suma	**25**	**135**

Si los datos de la variable se agrupan en intervalos, la media aritmética es la de las marcas de clase.

Ejemplo 2: La siguiente tabla muestra la talla en centímetros de 36 personas:

Tallas	**x_i**	**ni**	**$x_i \, . \, n_i$**
[150,170)	160	9	1440
[170,180)	175	18	3150
[180,200)	190	9	1710
Suma		36	6300

$\bar{x} = \dfrac{6300}{36} = 175$ cm.

Ejercicio 1.- Se ha preguntado a un grupo de 70 alumnos de un IES sobre el número de zapato que calzan, obteniendo los resultados de la siguiente tabla:

Nº de calzado	**Nº de alumnos**
35	4
36	15
37	17
38	20
40	10
42	4

¿Cuál es el número medio de calzado?

Ejercicio 2.- El consumo de carburantes de una flota de camiones a lo largo de un día está en la siguiente tabla de frecuencias. ¿Cuál es el consumo medio de carburante?

Consumo	**Camiones**
(0,10]	8
(10,20]	12
(20,30]	10
(30,40]	14
(40,50]	21
(50,60]	16
(60,70]	9

b) **Moda**: En el caso de variables no agrupadas, la moda es el valor de la variable que más se repite, esto es, el de mayor frecuencia absoluta. En el caso de variables agrupadas en intervalos, si son de igual amplitud buscamos el intervalo de mayor frecuencia (clase modal)

La moda depende de la frecuencia y no de que la variable sea cualitativa o cuantitativa. Así, la moda de la variable cualitativa “grupo sanguíneo” del ejemplo 1 del tema 2 es el grupo A.

Una variable puede tener más de una moda o ninguna. En el **ejemplo 1** de la distribución de las notas de 25 personas, las modas son 5 y 6; es bimodal.

Ejercicio 3.- Se ha preguntado a un grupo de 70 alumnos de un IES sobre el número de zapato que calzan, obteniendo los resultados de la siguiente tabla:

Nº de calzado	**Nº de alumnos**
35	4
36	15
37	17
38	20
40	10
42	4

¿Cuál es la moda?

Ejercicio 4.- El consumo de carburantes de una flota de camiones a lo largo de un día está en la siguiente tabla de frecuencias.

Consumo	**Camiones**
(0,10]	8
(10,20]	12
(20,30]	10
(30,40]	14
(40,50]	21
(50,60]	16
(60,70]	9

¿Cuál es el intervalo modal? Aproxima la moda utilizando la fórmula.

c) **Mediana:** En el caso de una variable no agrupada, una vez ordenados los datos, la mediana es el valor central si el nº de observaciones es impar y la media de los valores centrales si es par.

En el **ejemplo 1** de la distribución de notas de notas, como tenemos 25 elementos, al dividirlos en dos partes iguales nos queda como el valor central la posición 13; por tanto la mediana es $M_e=5$, que es la nota que corresponde a ese lugar.

Ejercicio 5.- Se ha preguntado a un grupo de 70 alumnos de un IES sobre el número de zapato que calzan, obteniendo los resultados de la siguiente tabla. ¿Cuál es la mediana?

Nº de calzado	**Nº de alumnos**
35	4
36	15
37	17
38	20
40	10
42	4

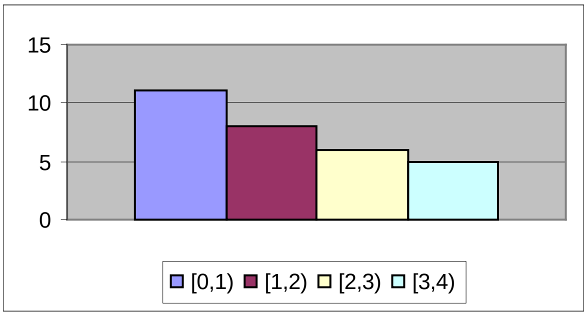
Ejercicio 6.- El consumo de carburantes de una flota de camiones a lo largo de un día está en la siguiente tabla de frecuencias.

Consumo	**Camiones**
(0,10]	8
(10,20]	12
(20,30]	10
(30,40]	14
(40,50]	21
(50,60]	16
(60,70]	9

¿Cuál es la clase mediana? Aproxima la mediana utilizando la fórmula.

Interpretación gráfica de los parámetros de centralización.

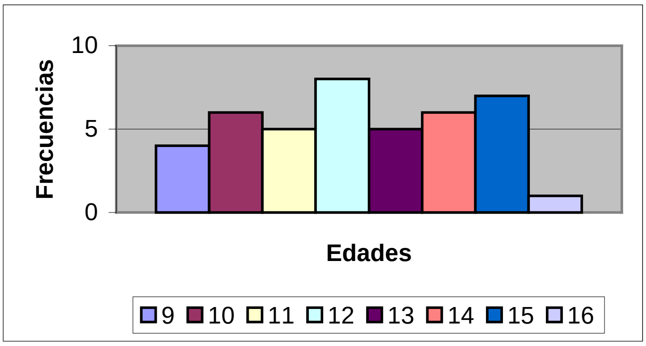
A partir de los datos representados en los gráficos se pueden calcular algunos parámetros de centralización. Por ejemplo, el siguiente histograma muestra las horas de estudio de 30 alumnos.



La clase modal es [0,1), puesto que la mayor parte del alumnado estudia menos de 1 hora; la clase mediana corresponde a los alumnos que estudian de 1 h a 2 h diarias, y la media aritmética es el promedio de las marcas de clase:

$$\bar{x} = \frac{0'5 \cdot 11 + 1'5 \cdot 8 + 2'5 \cdot 6 + 3'5 \cdot 5}{30} = 1'67 \text{ horas}$$

Ejercicio 7.- Halla la media, la moda y la mediana de las edades representadas en el diagrama de barras siguiente.



2.- MEDIDAS DE POSICIÓN

Son medidas estadísticas que indican, una vez ordenados, cuántos elementos quedan a la izquierda o a la derecha de uno dado: centiles o percentiles, cuartiles y deciles**.**

a) **Centiles o percentiles.** Una vez ordenados los datos, los centiles o percentiles son los valores que dejan a su izquierda un porcentaje determinado de la población. Se representan por C_h o P_h, donde h indica el porcentaje. Por ejemplo, el C_{32} deja a su izquierda al 32% de la población.

En el **ejemplo 1** sobre las notas de 25 personas, si calculamos el C_{45}, tendremos que obtener el 45 % de 25 y saber así cuántos elementos hemos de dejar a la izquierda del que buscamos: 45 % de 25 es 11’25; hemos de dejar a la izquierda 11’25 personas.

Notas	ni	Ni
3	3	3
4	4	7
5	6	13
6	6	19
7	4	23
8	2	25

Observando las frecuencias absolutas acumuladas C_{45}=5 puesto que las personas que ocupan el lugar 11 y 12 han obtenido un 5.

b) **Cuartiles.** Una vez ordenados los datos, son los valores de la variable que dividen a los datos en cuatro grupos iguales. En cada uno de ellos hay un 25 % de individuos de la población o muestra. Se representan por Q_1, Q_2 y Q_3. Evidentemente se tiene:

$$Q_1 = C_{25} \qquad Q_2 = C_{50} = M_e \qquad Q_3 = C_{75}$$

Vamos a obtener el cuartil inferior del ejemplo1 y el cuartil superior del ejemplo 2.

En el **ejemplo 1**, como hay 25 elementos, obtenemos que Q_1 corresponde al que ocupa el lugar 6’25 (25 % de 25). Según la columna de las frecuencias absolutas acumuladas, Q_1=4.

c) **Deciles.** Una vez ordenados los datos, son los valores de la variable que dividen a los datos en diez grupos iguales, de modo que entre dos deciles hay un 10% de los individuos de la población o muestra. Se representan por D_1, D_2, D_3, …, D_{10}

$D_1=C_{10}$, $D_2=C_{20}$, …, $D_9=C_{90}$

Vamos a obtener el decil primero del ejemplo1 y el decil tercero que coincide con el percentil 30 del ejemplo 2.

En el **ejemplo 1,** como hay 25 elementos, el decil primero será la media aritmética entre los elementos que ocupan los lugares 2 y 3 (10 % de 25 es 2’5). Ambos son 3, por tanto $D_1=3$.

Ejercicio 8.- Se ha preguntado a un grupo de 70 alumnos de un IES sobre el número de zapato que calzan, obteniendo los resultados de la siguiente tabla:

Nº de calzado	**Nº de alumnos**
35	4
36	15
37	17
38	20
40	10
42	4

a) Añade la columna de frecuencias absolutas acumuladas y calcula sobre ella el centil 14, el cuartil primero y el decil cuarto.
b) Calcula el centil 14 y el cuartil primero utilizando la hoja de cálculo y las funciones apropiadas.

Ejercicio 9.- El consumo de carburantes de una flota de camiones a lo largo de un día está en la siguiente tabla de frecuencias.

Consumo	**Camiones**
(0,10]	8
(10,20]	12
(20,30]	10
(30,40]	14
(40,50]	21
(50,60]	16
(60,70]	9

Añade la columna de frecuencias absolutas acumuladas y calcula sobre ella el centil 45, el cuartil superior y el decil tercero.

3.- MEDIDAS DE DISPERSIÓN

Proporcionan una idea sobre la separación de los datos: rango o recorrido, desviación media, varianza, desviación típica y coeficiente de variación.

a) **Rango o recorrido.** Si la variable es no agrupada, el rango es la diferencia entre los valores mayor y menor de la variable. Si es agrupada, el rango es la diferencia entre el extremo superior del último intervalo y el extremo inferior del primer intervalo.

El rango sólo tiene en cuenta los valores extremos, por lo que no influyen en él los demás elementos de la distribución.

En el **ejemplo 1**, el rango o recorrido es: 8 – 3 = 5

En el **ejemplo 2**, como es agrupada será: 200 – 150 = 50

b) **Rango intercuartílico.** El rango intercuartílico es la diferencia entre el tercer y el primer cuartil: $Q_3 - Q_1$. Nos da una franja en la que se encuentra el 50% de la población.

En el **ejemplo 1**, sabemos que Q_1 =4. Si calculamos Q_3, como 75 % de 25 es 18’75, según la columna de frecuencias absolutas acumuladas tendremos que Q_3 = 6.

Por tanto, el rango intercuartílico es $Q_3 - Q_1 =$ 6-4 =2.

En el **ejemplo 2,** conocemos Q_3 =180; vamos a calcular Q_1, que estará en el intervalo [150, 170) puesto que 25 % de 36 es 9, y según la tabla el valor que ocupa el valor 9 se encuentra en dicho intervalo y aplicando la fórmula nos queda:

$$Q_1 = 150 + \frac{25 \cdot \frac{36}{100} - 0}{9} \cdot 20 = 170$$

Luego, $Q_3 - Q_1$ = 180 – 170=10

c) **Desviación media.** Es la media aritmética de las desviaciones de los valores de la variable respecto de la media de la distribución.

Se llama *desviación* respecto de la media al valor absoluto de la diferencia de los valores de la variable y la media. Desviaciones respecto de la media: $| x_i - \overline{x} |$

Por tanto, la **desviación media** *DM* es la media aritmética de los valores absolutos de las diferencias entre los distintos datos y la media aritmética. $DM = \frac{\sum_{i=1}^{k} | x_i - \overline{x} | \cdot n_i}{n}$

El cálculo de la desviación media se suele hacer mediante tablas, con columnas útiles para dicho cálculo. Vamos a calcular la del **ejemplo 1**:

Notas	ni	$\lvert x_i - \overline{x} \rvert \cdot n_i$
3	3	7,2
4	4	5,6
5	6	2,4
6	6	3,6
7	4	6,4
8	2	5,2
SUMA	**25**	**30,4**

Por tanto: $DM = \dfrac{30'4}{25} = 1'216$

Para calcular la desviación media en variables agrupadas tomamos la marca de clase y realizamos los cálculos de forma análoga.

d) **Varianza.** Es la media de los cuadrados de las desviaciones respecto de la media. Se representa por S^2 o σ^2.

$$S^2 = \frac{\sum_{i=1}^{k} (x_i - \overline{x})^2 \cdot n_i}{n}$$

Propiedades:

- Tiene la desventaja de que las desviaciones grandes afectan más al resultado.
- Las unidades de la varianza no coinciden con los valores de la muestra, ya que estamos elevando las desviaciones al cuadrado.
- Siempre es positiva, siendo nula cuanto todos los valores coinciden con la media.

El cálculo de la varianza se suele hacer mediante tablas, con columnas útiles para dicho cálculo. Vamos a calcular la del **ejemplo 1**:

Notas	ni	$(x_i - \overline{x})^2 \cdot n_i$
3	3	17,28
4	4	7,84
5	6	0,96
6	6	2,16
7	4	10,24
8	2	13,52
SUMA	**25**	**52**

Luego: $S^2 = \dfrac{52}{25} = 2'08$

Para calcular la varianza en variables agrupadas tomamos la marca de clase y realizamos los cálculos de forma análoga.

e) **Desviación típica.** Se define la desviación típica *S* como la raíz cuadrada positiva de la varianza. Es la unidad de dispersión más utilizada, siendo su unidad la misma que la de los valores de la muestra

En el **ejemplo 1:** $S = \sqrt{2'08} = 1'44$

.

Relación entre la media aritmética y la desviación típica.

Las tallas x, e y, de dos grupos de alumnos A y B, vienen dadas por las siguientes tablas:

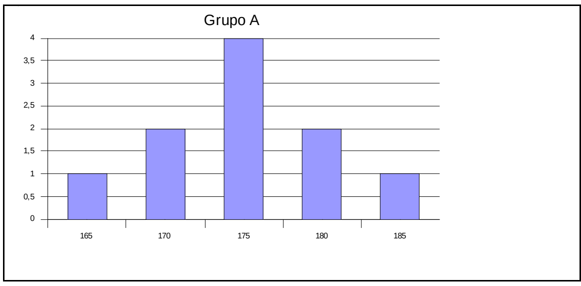
GRUPO A		
165	1	165
170	2	340
175	4	700
180	2	360
185	1	185

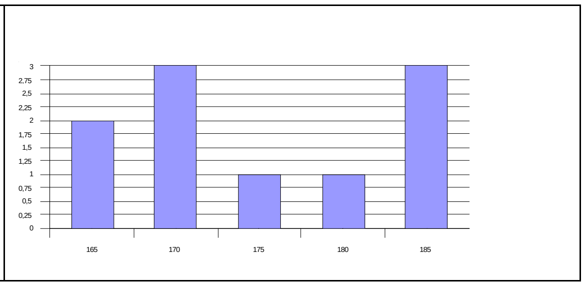
GRUPO B		
165	2	330
170	3	510
175	1	175
180	1	180
185	3	555

La talla media es la misma en ambos $\bar{x} = \bar{y} = 1750 \div 10 = 175$ cm

El grado de dispersión de las tallas de cada grupo respecto de la talla media viene dado por sus desviaciones típicas, que son: $S_x = 5'48$, $S_y = 7'75$

Como la media es la misma, y dado que $S_x < S_y$ la dispersión de las tallas de los alumnos del grupo B es mayor que la de los del grupo A, tal como se aprecia en los correspondientes gráficos.





f) **Coeficiente de variación de Pearson.** Las dispersiones de aquellas distribuciones que tienen medias aritméticas diferentes o cuyos datos vienen dados en unidades diferentes se pueden comparar mediante el coeficiente de variación, que se define como el cociente entre la desviación típica y la media. $CV = \frac{S}{\bar{x}}$.

Ejemplo 3: Imagina, por ejemplo, que los alumnos de un curso tienen una talla media $\bar{x} = 175$ cm., con una desviación típica $S_x = 5$ cm., y que la nota media obtenida en una asignatura por estos alumnos es $\bar{y} = 5$, con una desviación típica $S_y = 1'25$. Los coeficientes de variación son:

$$CV(x) = \frac{S_x}{\bar{x}} = 5 \div 175 = 0'03 \qquad CV(y) = \frac{Sy}{\bar{y}} = 1'25 \div 5 = 0'25$$

Es decir, el coeficiente de variación de las tallas es del 3 % y el de las notas, del 25 %.

Las tallas están, pues, menos dispersas que las notas.

Ejercicio 10.- Se ha preguntado a un grupo de 70 alumnos de un IES sobre el número de zapato que calzan, obteniendo los resultados de la siguiente tabla:

Nº de calzado	**Nº de alumnos**
35	4
36	15
37	17
38	20
40	10
42	4

a) Halla el rango o recorrido, el rango intercuartílico, la desviación media, la varianza, la desviación típica y el coeficiente de variación.
b) Comprueba los resultados utilizando las funciones apropiadas de la hoja de cálculo.

Ejercicio 11- El consumo de carburantes de una flota de camiones a lo largo de un día está en la siguiente tabla de frecuencias.

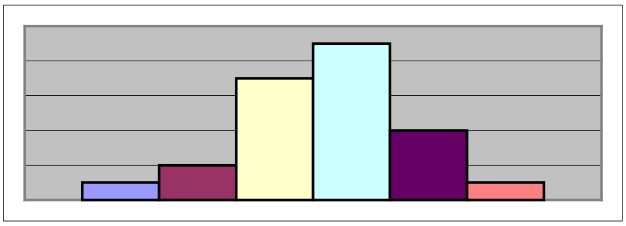
Consumo	**Camiones**
(0,10]	8
(10,20]	12
(20,30]	10
(30,40]	14
(40,50]	21
(50,60]	16
(60,70]	9

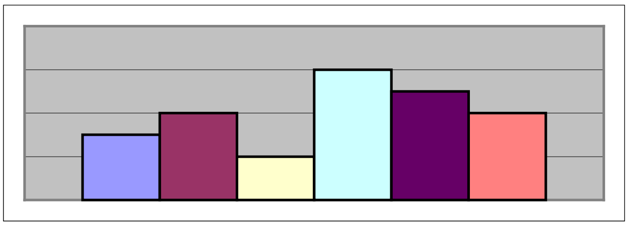
a) Halla el rango o recorrido, el rango intercuartílico, la desviación media, la varianza, la desviación típica y el coeficiente de variación.

b) Comprueba los resultados utilizando las funciones apropiadas de la hoja de cálculo.

$$\bar{x}_1 = 5'4\,;\ S_1 = 3'3\,;\ \bar{x}_2 = 5'4\,;\ S_2 = 2'5\,.$$

Averigua el gráfico correspondiente a cada par $(\bar{x}, S)$, explicando el razonamiento seguido.





Ejercicio 12.- Los jóvenes, a los 17 años, tienen un peso medio de 60’8 Kg. y una desviación de 6’69 Kg. Los niños a los 10 años tienen un peso medio de 30’5 Kg. y una desviación de 5’37 Kg. ¿Se puede afirmar que el peso es más variable a los 10 años que a los 17? ¿Por qué?

Ejercicio 13.- La siguiente tabla muestra las calificaciones obtenidas por Paco y Eva en diez controles de matemáticas:

Paco	4	5	5	4	6	7	8	9	3	9
Eva	5	6	6	5	6	7	7	6	5	7

Halla sus medias y desviaciones típicas. ¿Quién es más regular?