

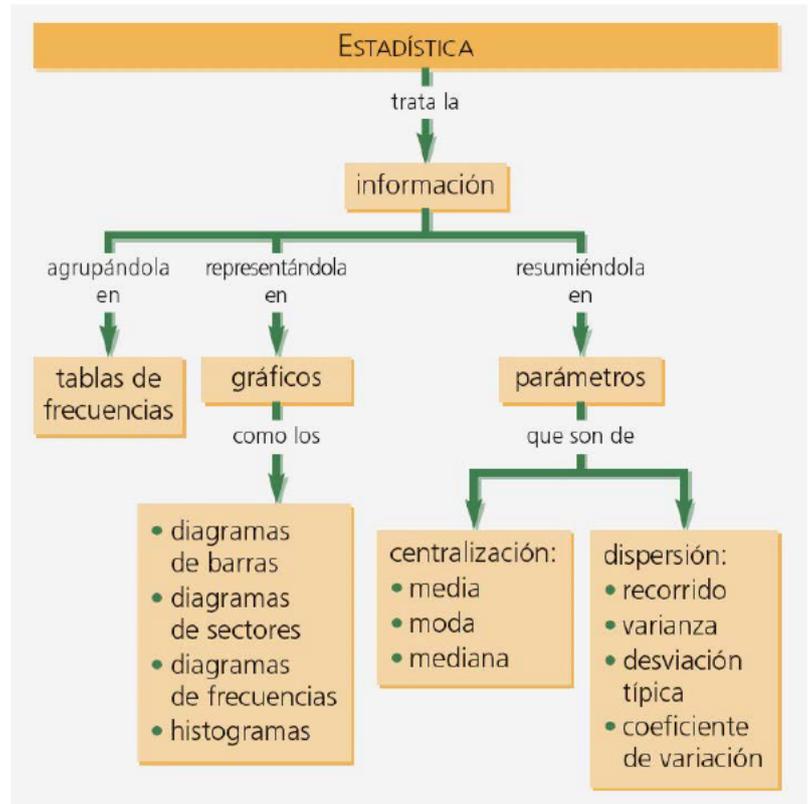
## Estadística: Teoría y Problemas.

Primero veremos la clasificación de los caracteres estadísticos, las tablas de frecuencias y los gráficos.

Después los parámetros de centralización, que indican alrededor de qué valores se distribuyen los datos, estudiando la forma de calcular la media, la mediana y la moda.

Y por último los parámetros de dispersión, que indican la concentración o dispersión de los datos en torno a la media. Estudiaremos el rango, la varianza, la desviación típica y el coeficiente de variación.

La Estadística es un método efectivo para describir e interpretar con exactitud los valores de los datos económicos, políticos, sociales, psicológicos, biológicos y físicos, y una herramienta para relacionar, analizar y obtener conclusiones de dichos datos.



### CONCEPTOS.

**Estadística:** Estudia los métodos para recoger, organizar y analizar información, con la finalidad de describir un fenómeno que se está estudiando y obtener conclusiones.

**Población:** Una población es el conjunto de elementos que son objeto de estudio. Pueden ser personas, animales, plantas o cosas.

**Muestra:** Una muestra es una parte de la población sobre la que se hace el estudio estadístico.

## 1. Caracteres estadísticos

### 1.1. Carácter estadístico

Es una propiedad que se estudia en los individuos de una población. Puede ser:

**Cualitativo:** es el que indica una cualidad. No se puede contar ni medir.

**Cuantitativo:** es el que indica una cantidad. Se puede contar o medir. Se clasifica en:

- a) **Discreto:** sus valores son el resultado de un recuento. Solo puede tomar ciertos valores aislados.
- b) **Continuo:** sus valores son el resultado de una medida. Puede tomar cualquier valor dentro de un intervalo.

### 1.2. Tablas de frecuencias

Los datos estadísticos se organizan en tablas en las que aparecen los valores de la variable y las frecuencias de dichos valores.

**Frecuencia absoluta:** es el número de individuos de la población para los que la variable toma un valor determinado. Se representa por  **$n_i$** .

La suma de todas las frecuencias absolutas,  **$n_i$** , es igual al total de los individuos,  **$N$** .

**Frecuencia relativa:** es el cociente entre la frecuencia absoluta y el número total de individuos. Se representa por  **$f_i$**  y se puede expresar:

$$\text{en tanto por uno } f_i = \frac{n_i}{N}, \text{ o en tanto por ciento } \frac{n_i}{N} \cdot 100$$

La suma de todas las frecuencias relativas es 1.

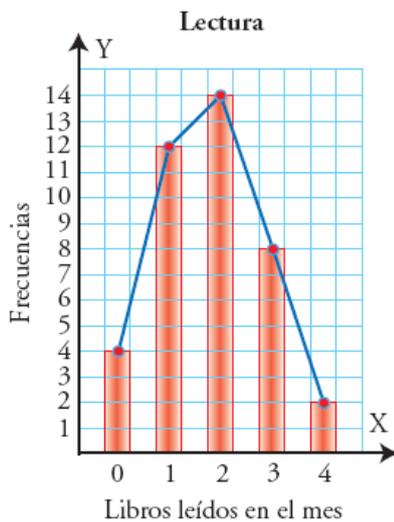
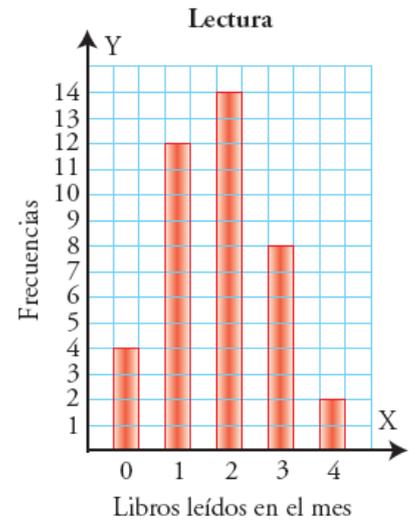
**Frecuencia acumulada de un valor:** es la suma de las frecuencias de los valores menores o iguales a él. Las frecuencias acumuladas pueden ser:

- Absolutas, que se representan por  **$N_i$** .
- Relativas, que se representan por  **$F_i$** .

## Estadística: Teoría y Problemas.

### 1.3. Diagrama de barras

Es un gráfico formado por barras de altura proporcional a la frecuencia de cada valor. Se utiliza con datos cualitativos y cuantitativos discretos.



### 1.4. Polígono de frecuencias o lineal

Es una representación que se realiza uniendo con una línea poligonal los extremos superiores de las barras en un diagrama de barras. Se utiliza con datos cualitativos y cuantitativos discretos.

## 2. Caracteres continuos. Datos agrupados

### 2.1. Tablas de datos agrupados en intervalos

Cuando el número de datos de un carácter cuantitativo discreto es muy grande o los datos son de un carácter cuantitativo continuo, éstos se agrupan en intervalos de igual longitud.

La **marca de clase de un intervalo** es el valor medio de dicho intervalo, y se toma como representante de éste cuando hay que realizar algún cálculo. Se representa por  $x_i$

#### Formación de intervalos

Para formar los intervalos se sigue el procedimiento:

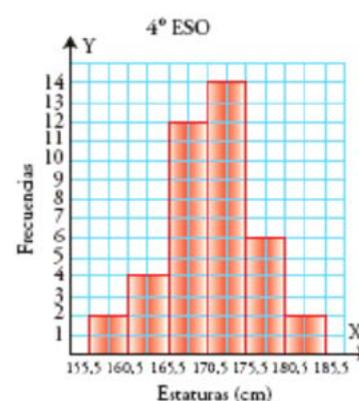
- c) Se determina el valor mayor,  $M$ , y el menor,  $m$ , de todos los datos y se halla el recorrido, que es la diferencia entre el mayor,  $M$ , y el menor,  $m$
- d) El número de intervalos se suele tomar redondeando la raíz cuadrada del número de datos.
- e) La longitud de cada intervalo se calcula dividiendo una longitud que sea múltiplo del número de intervalos, e igual o mayor que el recorrido entre el número de intervalos.

#### CONVENIO:

Se suele establecer el convenio siguiente: en un intervalo el extremo inferior pertenece al intervalo, y el extremo superior no pertenece al intervalo.

### 2.2. Histograma

Es una representación gráfica mediante rectángulos adosados, de altura proporcional a la frecuencia absoluta. Se utiliza cuando los datos son cuantitativos continuos o cuantitativos discretos y toma muchos valores diferentes.



### 2.3. Diagrama de sectores

Es una representación gráfica que consiste en un círculo que se divide en sectores de amplitud proporcional a la frecuencia del valor. El diagrama de sectores se puede utilizar con cualquier tipo de datos, pero se utiliza especialmente en datos cualitativos.

### 3. Parámetros de centralización

#### 3.1. Media

La **media o media aritmética** de un conjunto de datos es el resultado que se obtiene al dividir la suma de todos los datos entre el número total de ellos. Se representa por  $\bar{x}$

La media solo se puede calcular si los datos son cuantitativos.

**Significado de la media:** La media es un parámetro que indica el valor central sobre el cual se distribuyen los valores del carácter estudiado.

**Cálculo de la media:**

$$\bar{x} = \frac{\sum x_i \cdot n_i}{N} = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \dots + x_k \cdot n_k}{N}$$

#### 3.2. Mediana

La **mediana** de una distribución es el valor que está en el centro al ordenar los datos, es decir, el número de datos menores que él es igual al número de datos mayores que él. Para poder calcular la mediana, los datos se tienen que poder ordenar.

Si los datos estudiados son un número par, la mediana es la media de los dos valores centrales.

Si se tienen los datos en una tabla de frecuencias, la mediana es el 1er valor en el que la frecuencia acumulada supera la mitad de los datos.

Si los datos vienen agrupados en intervalos, se llama intervalo mediano.

#### 3.3. Moda

La moda de un conjunto de datos es el valor que tiene mayor frecuencia.

La **moda** se puede calcular siempre, tanto en datos cualitativos como en cuantitativos. Para calcular la moda, solo se debe observar qué valor tiene mayor frecuencia.

## 4. Parámetros de dispersión

### 4.1. Parámetros de dispersión

Son unos valores que indican si los datos de la distribución están más o menos cercanos a los parámetros centrales.

El **recorrido** es la diferencia entre el valor mayor y el menor de la distribución.

La **varianza** es la media de las desviaciones al cuadrado.

$$V = \frac{\sum x_i^2 \cdot n_i}{N} - \bar{x}^2$$

La varianza no se puede comparar con la media, ya que en la varianza se tienen unidades cuadradas, y en la media, unidades lineales. Para poder hacer esta comparación se define la desviación típica.

La **desviación típica** es la raíz cuadrada de la varianza. Se representa con el símbolo  $\sigma$ .

$$\sigma = \sqrt{V}$$

### 4.2. Coeficiente de variación

Si se conocen la media y la desviación típica de una distribución, se sabe, con la media, dónde está su centro y, con la desviación típica, si los datos están agrupados o dispersos.

Para comparar dos distribuciones se define el coeficiente de variación.

El **coeficiente de variación** es la comparación entre la desviación típica y la media aritmética, es decir:

Este coeficiente se suele multiplicar por cien  $CV = \frac{\sigma}{\bar{x}}$  para expresarlo en porcentaje.

Cuanto mayor sea el coeficiente de variación, mayor será la dispersión de los datos.

## Estadística: Teoría y Problemas.

### 5. Problemas

1º). Analiza los siguientes estudios estadísticos e indica en cada caso cuál es la población y el tipo de carácter estadístico que se estudia.

- La profesión de los integrantes de un conjunto coral.
- El porcentaje de fumadores que han dejado el tabaco en cada localidad después de una campaña del Ministerio de Sanidad aplicada en 20 grandes ciudades.
- El número de horas que dedican al estudio los alumnos de un centro escolar durante una semana. ¿En cuál de ellos será necesario seleccionar una muestra para su ejecución?

a) Población: integrantes del conjunto coral Carácter: profesión

No será necesaria una muestra.

b) Población: Personas fumadoras en las veinte localidades Carácter: porcentaje de fumadores que han dejado de fumar Sí será necesario elegir una muestra.

c) Población: alumnos del centro escolar Carácter: horas de estudio

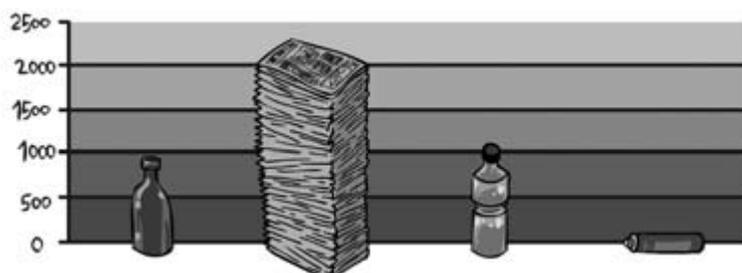
Posiblemente, sería conveniente elegir una muestra.

2º). Se quiere realizar un estudio estadístico sobre ciertos aspectos de los alumnos de un centro escolar: su edad, el tiempo que tardan en llegar al centro, el idioma extranjero que estudian, el lugar preferido de vacaciones y el deporte que suelen practicar.

¿De qué tipo es cada carácter? Indica los posibles valores de cada uno.

- Edad: de 14 a 18 años. Carácter cuantitativo
- Tiempo que se tarda en llegar al centro: entre 0 y 60 minutos. Carácter cuantitativo
- Idiomas extranjeros: inglés, francés, italiano, alemán, etc. Carácter cualitativo
- Lugares preferidos de vacaciones: mar, montaña, ciudad, etc. Carácter cualitativo
- Deportes que practican: fútbol, baloncesto, natación, atletismo, etc. Carácter cualitativo

3º). El gráfico representa las toneladas de diferentes residuos que se reciclan en una ciudad. Indica qué tipo de carácter estadístico se estudia, así como sus frecuencias absolutas y relativas.



## Estadística: Teoría y Problemas.

Se trata de un carácter estadístico cuantitativo.

	$f_i$	$h_i$
Vidrio	900	0,21
Papel	2250	0,52
Plástico	1100	0,25
Pila	100	0,02
Total	4350	1

4º). Se ha preguntado a los 25 alumnos de una clase de 4.º de ESO por el número de veces que se han conectado a internet durante un fin de semana y se han obtenido los siguientes resultados.

2 3 2 1 1 1 3 2 1 2

2 0 1 3 1 4 1 3 0 2

2 4 0 3 1

Ordena los datos en una tabla de frecuencias simple.

- ¿Cuántos alumnos se han conectado dos veces durante el fin de semana? ¿Qué porcentaje representan?
- ¿Cuántos alumnos se han conectado como mucho dos veces durante el fin de semana? ¿Qué porcentaje representan?

$X_i$	$f_i$	$h_i$	$F_i$	$H_i$
0	3	0,12	3	0,12
1	8	0,32	11	0,44
2	7	0,28	18	0,72
3	5	0,20	23	0,92
4	2	0,08	25	1

- Siete alumnos se conectaron dos veces, lo cual representa el 28% del total de alumnos.
- Dieciocho alumnos se conectaron, como mucho, dos veces. Esto representa el 72% del total de alumnos.

5º). Los posibles valores de una variable estadística discreta de una población son 3, 5, 7 y 8. Se ha tomado una muestra de 30 personas de dicha población y se han obtenido los siguientes resultados.

7 3 7 3 7 7 7 3 7 3

5 7 7 3 8 5 7 3 8 5

7 3 8 5 7 3 5 7 3 5

## Estadística: Teoría y Problemas.

- Ordena los datos en una tabla de frecuencias.
- ¿Cuántas personas presentaron un valor de 7 y qué porcentaje del total representaban?
- ¿Cuántas personas presentaron un valor impar y qué porcentaje del total representaban?

a)

$X_i$	$f$	$h$	$F$	$H$
3	9	0,3	9	0,3
5	6	0,2	15	0,5
7	12	0,4	27	0,9
8	3	0,1	30	1
Total	30	1		

- 12 personas presentaron el valor 7. Representaban el 40% del total.
- 27 personas presentaron el valor impar. Representaban el 90% del total.

6º). Los datos de una lista son 50 números comprendidos entre 0 y 40. Se quieren agrupar en 5 clases de igual longitud.

- ¿Cuáles deben ser dichas clases?
- ¿Cuáles son las marcas?

- $0 \leq X < 8$     $8 \leq X < 16$     $16 \leq x < 24$     $24 \leq x < 32$    y    $32 \leq x < 40$
- Las marcas son: 4, 12, 20, 28 y 36.

7º). Los posibles valores de una variable estadística continua son los correspondientes al intervalo [20, 40]. Se ha tomado una muestra de 20 personas de dicha población y se han obtenido los siguientes resultados.

- Ordena los datos agrupados en cinco intervalos de igual longitud y construye una tabla de frecuencias.
- ¿Cuántas personas presentaron un valor de la característica menor de 36?
- ¿Qué porcentaje representaban?

## Estadística: Teoría y Problemas.

a)

Intervalos	Marcas	$f$	$h_i$	$F_i$	$H$
$20 \leq X < 24$	22	3	0,15	3	0,15
$24 \leq X < 28$	26	5	0,25	8	0,40
$28 \leq X < 32$	30	7	0,35	15	0,75
$32 \leq X < 36$	34	4	0,20	19	0,95
$36 \leq X < 40$	38	1	0,05	20	1
Total		20	1		

b) 19 personas presentaron un valor menor que 36.

c) Representaban el 95% del total.

8º). La siguiente tabla representa el número de problemas bien resueltos por un grupo de alumnos en la primera fase de una olimpiada matemática.

Problemas bien resueltos	0	1	2	3	4	5
N.º de alumnos	1	3	10	12	4	2

a) Elabora una tabla de frecuencias.

b) ¿Qué porcentaje de alumnos resolvieron más de tres problemas de forma correcta?

c) ¿Cuántos alumnos resolvieron entre dos y cuatro problemas de forma correcta? ¿Qué porcentaje del total representan?

d) Se quiere que pasen a la siguiente fase un número cercano al 25% de los alumnos. Realiza los cálculos que consideres necesarios para determinar qué alumnos deberán realizar dicha segunda fase.

a)

$X_i$	$f_i$	$h$	$F_i$	$H_i$
0	1	0,031	1	0,031
1	3	0,094	4	0,125
2	10	0,312	14	0,438
3	12	0,375	26	0,813
4	4	0,125	30	0,938
5	2	0,063	32	1
Total	32	1		

b) El 18,75% de los alumnos resolvió más de tres problemas de forma correcta.

c) 26 alumnos resolvieron correctamente entre 2 y 4 problemas. Representan el 81,25% del total.

d) El 25% de los alumnos es 8. Deberían pasar los 6 alumnos que han resuelto 4 ó 5 problemas de forma correcta. Representan el 18,75%.

## Estadística: Teoría y Problemas.

9º). La siguiente tabla muestra la masa, en gramos, de las 125 bolas de billar fabricadas por una máquina en una hora.

Masa (g)	$120 \leq X < 140$	$140 \leq X < 160$	$160 \leq X < 180$
N.º de bolas	15	75	35

- Elabora una tabla de frecuencias que incluya las marcas de clase.
- ¿Qué porcentaje de bolas tienen una masa comprendida entre 140 y 160 gramos?
- ¿Qué porcentaje de bolas tienen una masa inferior a 160 gramos?
- Elabora una nueva tabla del mismo tipo que la del enunciado y en la que se exprese el número de bolas que se espera estén incluidas en cada intervalo si se considera el conjunto de las 1000 bolas fabricadas durante una jornada completa.

a)

Intervalos	Marcas	$f_i$	$h_i$	$F_i$	$H_i$
$120 \leq X < 140$	130	15	0,12	15	0,12
$140 \leq X < 160$	150	75	0,60	90	0,72
$160 \leq X < 180$	170	35	0,28	125	1
Total		125	1		

- El 60% de las bolas tiene entre 140 y 160 g de masa.
  - El 72% de las bolas tiene menos de 160 g de masa.
- d)

Intervalos	$f_i$
$120 \leq X < 140$	120
$140 \leq X < 160$	600
$160 \leq X < 180$	280
Total	1000

10º). La siguiente tabla muestra los valores de las frecuencias absolutas de las cuatro modalidades de un carácter cualitativo.

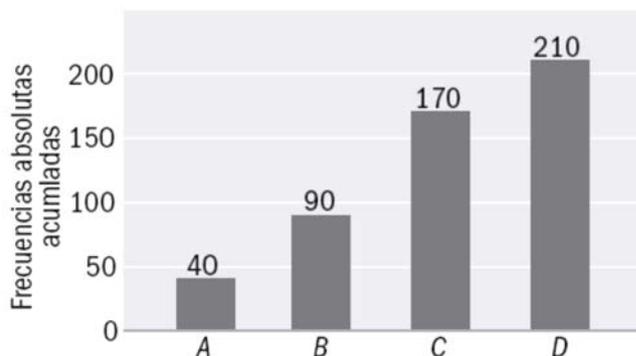
Modalidades	A	B	C	D
F. absolutas	40	50	80	40

Representa las frecuencias absolutas acumuladas mediante un diagrama de barras.

## Estadística: Teoría y Problemas.

Se completa la tabla con la columna de frecuencias absolutas acumuladas

Datos	$f_i$	$F_i$
A	40	40
B	50	90
C	80	170
D	40	210



11º). Se ha lanzado un dado 50 veces. En ocho de ellas se ha obtenido 1; en siete, 2; en nueve, 3; en ocho, 4; en seis, 5, y en doce, 6. Dibuja el diagrama de sectores para representar los datos y el diagrama de barras para las frecuencias acumuladas.

Diagrama de sectores

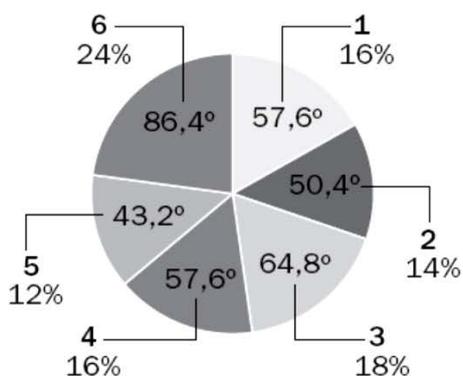
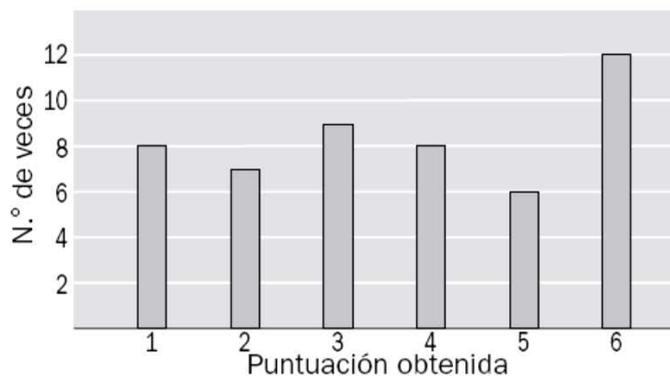


Diagrama de barras



12º). La siguiente tabla representa la edad en años de un grupo de 100 personas.

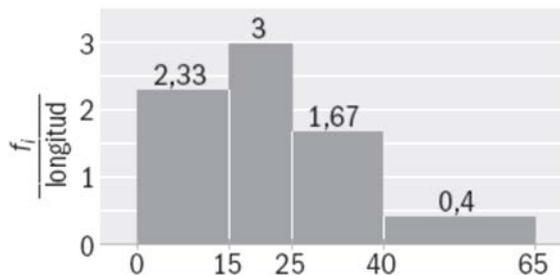
$0 \leq X < 15$	$15 \leq X < 25$	$25 \leq X < 40$	$40 \leq X < 65$
35	30	25	10

Representa los datos en un histograma.

Como las bases de los rectángulos tienen distinta longitud, la medida de sus alturas debe ser el cociente de las frecuencias absolutas correspondientes entre esas longitudes.

## Estadística: Teoría y Problemas.

Clases	Longitudes	$f_j$	$F$ longitud
$0 \leq X < 15$	15	35	2,33
$15 \leq X < 25$	10	30	3
$25 \leq X < 40$	15	25	1,67
$40 \leq X < 65$	25	10	0,4

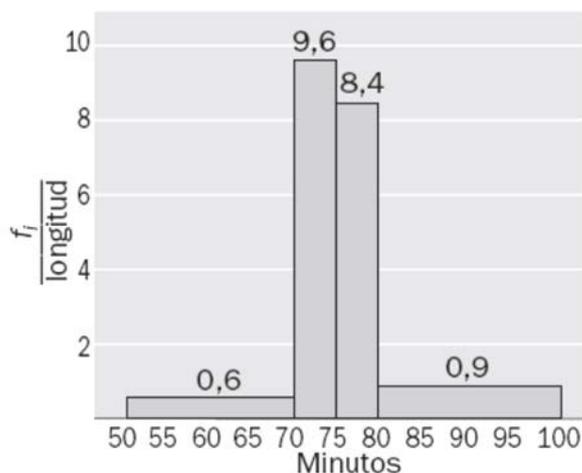


13º). La tabla muestra el tiempo que 120 personas dedican diariamente a las tareas domésticas.

De 50 a 70 min.	De 70 a 75 min.	De 75 a 80 min.	De 80 a 100 min.
12	48	42	18

Dibuja el correspondiente histograma.

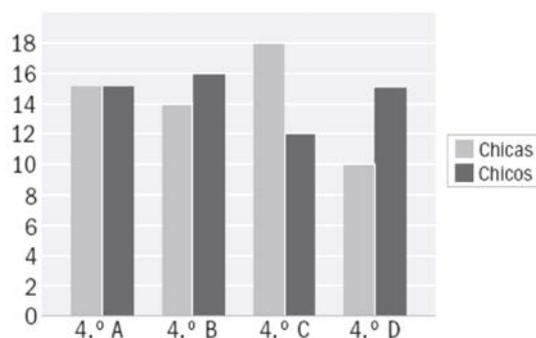
Intervalos	Longitudes	$f_j$	$f$ longitud
$50 \leq X < 70$	20	12	0,6
$70 \leq X < 75$	5	48	9,6
$75 \leq X < 80$	5	42	8,4
$80 \leq X < 100$	20	18	0,9



14º). El diagrama de barras representa el número de chicos y chicas que integran los cuatro grupos de 4.º de ESO de un centro escolar.

Elabora la tabla de frecuencias correspondiente a partir de la información del gráfico.

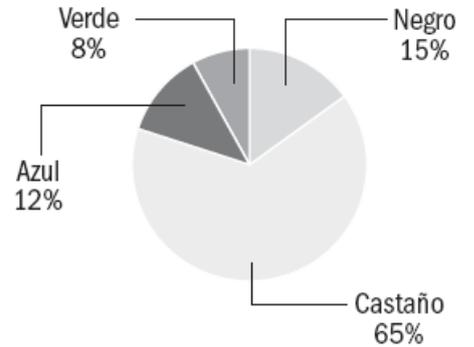
	4.º A	4.º B	4.º C	4.º D
Chicas	15	14	18	10
Chicos	15	16	12	15
Total	30	30	30	25



## Estadística: Teoría y Problemas.

15º). El diagrama de sectores muestra la distribución de 500 personas según el color de sus ojos.

- Elabora una tabla con las frecuencias absolutas de cada modalidad.
- Construye el diagrama de barras correspondiente.



a)

	F. absoluta
Castaño	325
Azul	60
Verde	40
Negro	75
Total	500

b)

