1.- CONCEPTOS ESTADÍSTICOS. TABLAS Y GRÁFICOS ESTADÍSTICOS

Población	Muestra
Es el conjunto formado por todos los elementos de los que se quiere estudiar alguna característica. Por ejemplo, si vamos a estudiar las aficiones de los jóvenes de 15 años nacidos en la capital de Granada y hay 5 000 jóvenes, la población sería el conjunto formado por los 5 000 jóvenes	Es una parte de la población que elegimos para estudiarla. Se toma una muestra cuando la población sea muy numerosa. El proceso de elegir una muestra se llama muestreo.

Muestreo proporcional

Consiste en dividir la población en grupos y tomar aleatoriamente en cada grupo una muestra proporcional al nº de elementos del grupo.
Se utiliza cuando la población se puede dividir en grupos (por ejemplo, menores de edad, jubilados y el resto).

Ejemplo

En una aldea de 2 000 habitantes hay personas de pelo negro, rubio o castaño.
Se ha seleccionado, mediante muestreo aleatorio estratificado proporcional, una muestra constituida por 28 personas de pelo negro, 32 de pelo rubio y 20 de pelo castaño.
Determina cuál es la composición, según el color del pelo, de esa aldea.

Solución:

	pelo negro	pelo rubio	pelo castaño	total
población	x	y	z	2000
muestra	28	32	20	80

$$\Rightarrow \frac{x}{28} = \frac{y}{32} = \frac{z}{20} = \frac{2000}{80} = 25 \Rightarrow \begin{array}{l} x = 28 \cdot 25 = 700 \text{ de pelo negro} \\ y = 32 \cdot 25 = 800 \text{ de pelo rubio} \\ z = 20 \cdot 25 = 500 \text{ de pelo castaño} \end{array}$$

Variable estadística

Una variable estadística es una característica que queremos estudiar de la población.
Hay varios tipos de variables estadísticas:

<table>
<tr><td rowspan="2"><u>Cualitativa</u>
Si los valores son cualidades.
Por ejemplo, partido político preferido, color del pelo, etc.</td><td colspan="2"><u>Cuantitativa</u>
Si los valores son números. Por ejemplo, nº de hermanos, estatura, peso, edad, etc.</td></tr>
<tr><td><u>Discreta</u>
Cuando los valores son aislados.
Por ejemplo, nº de hermanos, edad, etc.</td><td><u>Continua</u>
Cuando entre dos valores, aunque estén muy próximos entre sí, siempre es posible tomar otro valor.
Por ejemplo, la temperatura, el peso, la estatura, etc.</td></tr>
</table>

Tabla de frecuencias

Para organizar y analizar los datos se utilizan unas tablas, llamadas tablas de frecuencias

Ejemplo:

Estas son las edades de un grupo de alumnos:

14 , 15 , 13 , 13 , 14
15 , 15 , 18 , 14 , 13
15 , 13 , 14 , 15 , 16
14 , 15 , 13 , 13 , 15

Tabla de frecuencias

x_i	f_i	F_i	h_i	H_i
13	6	6	30%	30%
14	5	11	25%	55%
15	7	18	35%	90%
16	1	19	5%	95%
18	1	20	5%	100%
Total	20 = n	-	100%	-

x_i son los valores que aparecen en los datos. En la tabla se escriben ordenados de menor a mayor.

f_i se llama **frecuencia absoluta** y es el número de veces que aparece cada valor en los datos.

Por ejemplo, el número 7 de la columna f_i , significa que hay 7 alumnos con 15 años.

La suma de las frecuencias absolutas es igual al número total de datos " **n** " (en este caso n = 20, pues hay 20 datos)

F_i se llama **frecuencia absoluta acumulada** y se calcula sumando uno a uno los valores de la columna f_i.

F_i representa el número de datos que hay menores o iguales al valor x_i correspondiente.

Por ejemplo, el número 11 de la columna F_i , significa que hay 11 alumnos con 14 años o menos.

h_i se llama **frecuencia relativa** y se calcula dividiendo cada valor f_i entre el nº total de datos, n. $\boxed{h_i = \frac{f_i}{n}}$

La frecuencia relativa se suele expresar en forma de % y nos indica el % de datos que hay iguales al valor x_i correspondiente.

Por ejemplo, el 30% de la columna h_i significa que hay un 30% de alumnos con 13 años.

La suma de las frecuencias relativas es siempre igual al 100%

H_i se llama **frecuencia relativa acumulada** y se calcula sumando uno a uno los valores de la columna h_i.

La frecuencia relativa acumulada se suele expresar en forma de % y nos indica el % de datos que hay menores o iguales al valor x_i correspondiente.

Por ejemplo, el 95% de la columna H_i significa que hay un 95% de alumnos que tienen 16 años o menos.

Tabla de frecuencias agrupando los datos

Cuando hay "muchos" valores distintos en los datos, agrupamos los valores en intervalos (llamados clases)

Ejemplo:

Las notas en la materia de inglés de un grupo de alumnos han sido:

3,1 ; 7 ; 2 ; 5,6 ; 6,1   7,3 ; 4,7 ; 5,2 ; 7,1 ; 3   2,8 ; 2,9 ; 4,1 ; 4,9 ; 7,8   6,4 ; 6,2 ; 5,2 ; 5,4 ; 5,3

Nota máxima = 7,8 , Nota mínima = 2   Tomamos, por ejemplo, intervalos de amplitud 1

Tabla de frecuencias

Clases	f_i	F_i	h_i	H_i
[2, 3)	3	3	15%	15%
[3, 4)	2	5	10%	25%
[4, 5)	3	8	15%	40%
[5, 6)	5	13	25%	65%
[6, 7)	3	16	15%	80%
[7, 8)	4	20	20%	100%
Total	20 = n	-	100%	-

Observa, por ejemplo, que:

- Hay 5 alumnos que tienen un 5 ó 5 y "pico"
- Hay 16 alumnos que tienen menos de un 7
- El 10% de los alumnos tienen un 3 o un 3 y "pico"
- El 25% de los alumnos tienen menos de un 4
- El porcentaje de aprobados es 100% – 40% = 60%

Gráficos estadísticos

Los datos obtenidos en un estudio estadístico los podemos representar con diferentes gráficos. Los gráficos nos ayudan a analizar los datos a simple vista. Vamos a ver los gráficos que más se usan:

Diagrama de barras

Se representan los valores x_i en un eje horizontal y para cada valor x_i se dibuja una barra cuya altura sea la frecuencia que se quiera representar, f_i, F_i, h_i ó H_i.

Las barras deben ser de la misma anchura y debemos dibujarlas separadas.

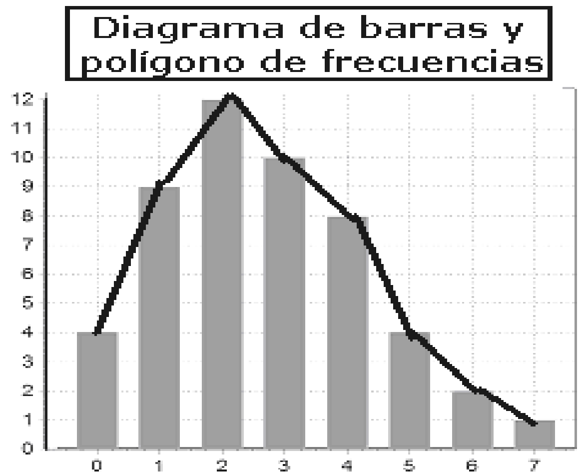
Uniendo los extremos superiores de las barras por su punto medio, se obtiene una línea quebrada llamada polígono de frecuencias

El diagrama de barras se suele utilizar para variables discretas con "pocos" valores o para variables cualitativas

Ejemplo:

Vamos a representar el diagrama de barras para las frecuencias absolutas, fi, de los datos correspondientes al número de hijos de 50 matrimonios.

x_i	f_i
0	4
1	9
2	12
3	10
4	8
5	4
6	2
7	1
Total	50 = n



Histograma

Es similar al diagrama de barras, sólo que la base de cada barra es el intervalo de la tabla de frecuencias y por tanto no hay espacios entre las barras.

Uniendo los extremos superiores de las barras por su punto medio, se obtiene la línea quebrada llamada polígono de frecuencias.

Los histogramas se utilizan cuando los datos los agrupamos en intervalos.

Ejemplo:

Vamos a representar el diagrama de barras para las frecuencias absolutas, f_i, de los datos correspondientes a las notas de en un examen de 20 alumnos:

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

clases	f_i
[2,3)	3
[3,4)	2
[4,5)	3
[5,6)	5
[6,7)	3
[7,8)	4
Total	20 = n

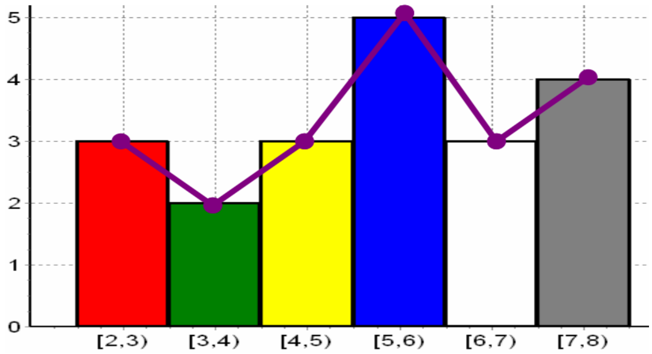


Diagrama de sectores

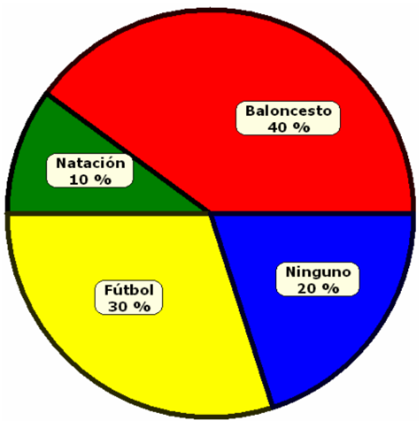
Para el diagrama de sectores se dibuja un círculo y se divide en tantos sectores (quesitos) como valores haya en los datos.

Se utiliza la frecuencia relativa hi, de modo que el ángulo de cada sector sea proporcional al valor de h_i.

Ejemplo:

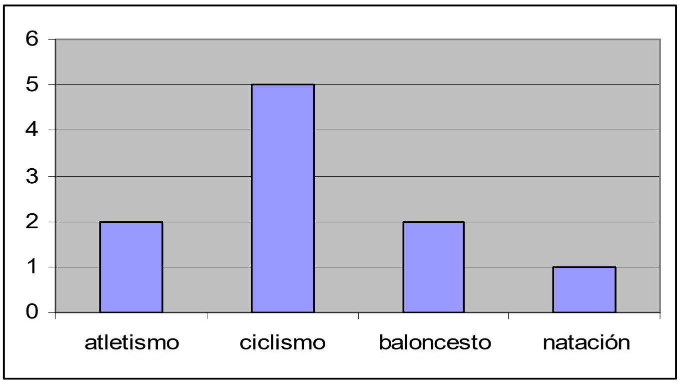
Vamos a representar el diagrama de sectores de los datos correspondientes al deporte elegido como favorito de un grupo de alumnos:

x_i = Deporte	f_i	h_i (en %)	Ángulo del sector
Baloncesto	12	40	40% de 360º = 0,4 . 360 = 144º
Natación	3	10	10% de 360º = 0,1 . 360 = 36º
Fútbol	9	30	30% de 360º = 0,3 . 360 = 108º
Ninguno	6	20	20% de 360º = 0,2 . 360 = 72º
Total	30	100%	360º



ACTIVIDADES

1.- Este es el diagrama de barras de las frecuencias absolutas, que corresponde a los gustos deportivos de un grupo de personas:



a) Completa la tabla:

x_i	f_i	h_i (en %)	Ángulo del sector
atletismo			
ciclismo			
baloncesto			
natación			
Total		100%	360º

b) ¿A cuántas personas no le gusta el atletismo?

c) ¿A qué porcentaje de personas le gusta el baloncesto?

d) Dibuja el diagrama de sectores

2.- El nº de días a la semana que practican deporte un grupo de alumnos de 3º de ESO es:

2 ; 0 ; 2 ; 2 ; 3   3 ; 2 ; 3 ; 1 ; 2   3 ; 2 ; 1 ; 2 ; 0   1 ; 2 ; 3 ; 3 ; 3

a) Completa la tabla de frecuencias:

x_i	f_i	F_i	h_i (en %)	H_i (en %)
Total				

b) Usando la tabla de frecuencias determina:

1) El número de alumnos que practican deporte menos de 3 días a la semana

2) El porcentaje de alumnos que practica deporte más de 1 día a la semana

c) Dibuja el diagrama de barras y el polígono de frecuencias usando las frecuencias relativas

3.- Al preguntar a los profesores del instituto sus edades se han obtenido los datos:

25	44	31	53	28	47	38	26
32	45	30	25	44	46	50	31
40	41	38	33	32	35	31	39
27	48	26	54	29	54	30	27
37	34	33	40	32	28	47	39

a) Agrupa los datos en intervalos de amplitud 5 y construye la tabla de frecuencias.

Clases	f_i	F_i	h_i (en %)	H_i (en %)
Total	n =		100 %	

b) ¿Qué % de profesores tiene menos de 40 años? c) ¿Cuántos profesores tienen menos de 45 años?

d) ¿Qué porcentaje de profesores tiene entre 30 y 49 años (ambas edades incluidas)?

4.- Al observar la marca de coche de un grupo de personas se obtuvieron los siguientes resultados:

S P S R R P P P R R S P R R R P S S R R, donde S = Seat , P = Peugeot , R = Renault.

a) Completa la siguiente tabla:

x_i	f_i	F_i	h_i (en %)	H_i (en %)
Seat				
Peugeot				
Renault				
Total				

b) Completa esta tabla y después dibuja el diagrama de sectores:

x_i	Ángulo del sector
Seat	
Peugeot	
Renault	
Total	

c) Indica cuál es la moda

d) Indica qué porcentaje de personas tienen Peugeot

Actividades del libro: 3, 5 (pág. 273) y 11 (pág. 275)

2.- MEDIDAS ESTADÍSTICAS

La media aritmética

Es la suma de todos los datos dividida entre el número total de datos, n. Se calcula usando la fórmula $\boxed{\overline{x} = \frac{\sum(x_i\ f_i)}{n}}$, donde $\sum$ significa suma.

Ejemplo:

Notas en un examen de un grupo de amigos

x_i	f_i	$x_i\ f_i$
4	1	4
5	2	10
6	4	24
7	3	21
Total	10 = n	59

$$\overline{x} = \frac{\sum(x_i\ f_i)}{n} = \frac{59}{10} = 5,9$$

Si los datos están agrupados en intervalos, se toma como x_i el punto medio del intervalo. Este valor se llama marca de clase

Ejemplo:

Gasto mensual en €, en teléfono móvil, de un grupo de jóvenes

clases	x_i	f_i	$x_i\ f_i$
[10, 11)	10,5	4	42
[11, 12)	11,5	6	69
[12, 13)	12,5	7	87,5
[13, 14)	13,5	3	40,5
Total		20 = n	239

$$\overline{x} = \frac{\sum(x_i\ f_i)}{n} = \frac{239}{20} = 11,95 \text{ €}$$

La moda (Mo)

Es el valor que más se repite en los datos. La moda es el valor xi que tiene mayor frecuencia absoluta.

Si los datos están agrupados en intervalos se toma el intervalo de mayor frecuencia (intervalo o clase modal).

Puede haber más de una moda o puede que no haya moda porque todos los valores tengan la misma frecuencia absoluta.

Ejemplo:

x_i = Equipo de fútbol preferido	Nº de personas
Madrid	12
Granada	7
Barcelona	12
Málaga	6

Hay dos modas, Madrid y Barcelona.

La mediana (Me)

Es el dato que está justamente en medio, cuando tenemos todos los datos ordenados de menor a mayor.

Cálculo de la mediana cuando hay "pocos" datos

- Si el **nº de datos es impar** , la mediana es el dato central

Ejemplo:

Edades de 9 personas: 15 , 12 , 17 , 15 , 14 , 14 , 17 , 15 , 15

Ordenando los datos: 12, 14, 14, 15, $\boxed{15}$, 15, 15, 17, 17 → Me = 15

- Si el **nº de datos es par** , la mediana es la media aritmética de los 2 datos centrales

Ejemplo:

Notas de 12 alumnos: 7 , 4 , 6 , 5 , 7 , 7 , 8 , 5 , 8 , 4 , 4 , 5

Ordenando los datos: 4, 4, 4, 5, 5, $\boxed{5, 6}$, 7, 7, 7, 8, 8 → Me = 5,5

Cálculo de la mediana cuando hay "muchos" datos

En este caso, la mediana es el primer valor x_i cuya H_i es mayor que el **50%**

Ejemplo:

Notas en Inglés de 20 alumnos:

clases	x_i	H_i
[2,3)	2,5	15
[3,4)	3,5	25
[4,5)	4,5	40
[5,6)	5,5	**65**
[6,7)	6,5	80
[7,8)	7,5	100

Me = 5,5

Los cuartiles

Cuando los datos están ordenados de menor a mayor, los cuartiles son tres valores Q_1 , Q_2 , Q_3 que dividen a los datos en 4 partes iguales

El primer cuartil, Q_1 , es el primer valor x_i cuya H_i es mayor que el 25%

El segundo cuartil, Q_2 , es el primer valor x_i cuya H_i es mayor que el 50%. Es decir, Q_2 = Me

El tercer cuartil, Q_3 , es el primer valor x_i cuya H_i es mayor que el 75%

El ***rango intercuartílico*** (RI) es la distancia entre Q_1 y Q_3 → RI: $Q_3 - Q_1$

Si los datos estuviesen agrupados en intervalos se toma como x_i la marca de clase

Ejemplo:

Notas en Inglés de 20 alumnos:

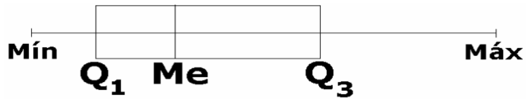
clases	x_i	H_i
[2,3)	2,5	15
[3,4)	3,5	25
[4,5)	4,5	40
[5,6)	5,5	65
[6,7)	6,5	80
[7,8)	7,5	100

$Q_1 = 4,5$ $Q_2 = Me = 5,5$ $Q_3 = 6,5$ RI = 6,5 – 4,5 = 2

Diagrama de caja y bigotes

Los cuartiles se suelen representar en un diagrama, llamado diagrama de caja y bigotes

Para dibujar el diagrama de caja, se calculan los valores mínimo y máximo de x_i así como los cuartiles. Después se dibuja una caja, cuyos extremos son Q_1 y Q_3, que indica donde se concentran el 50% de los datos y una línea central que marca la mediana.

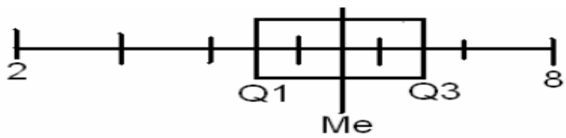


El diagrama de caja nos permite saber qué si la distribución de datos es simétrica respecto de la mediana: si la caja está desplazada hacía la izquierda o hacía la derecha respecto de la mediana significa que la distribución de datos no es simétrica respecto a la mediana.

Ejemplo:

Para las notas en Inglés, Valor mínimo: 2 Valor máximo: 8 Q1 = 4,5 Q2 = Me = 5,5 Q3 = 6,5

El diagrama de caja y bigotes sería:



Recorrido o rango: El *recorrido* o rango, R, es la diferencia entre el mayor y el menor valor de x_i

Varianza: La *varianza*, s^2, se calcula con la fórmula: $\boxed{s^2 = \frac{\sum x_i^2 . f_i}{n} - \overline{x}^2}$

Desviación típica: La *desviación típica*, s, es la raíz cuadrada de la varianza: $\boxed{s = \sqrt{s^2}}$

Coeficiente de variación: El *coeficiente de variación*, CV, se calcula con la fórmula: $\boxed{CV = \frac{s}{\overline{x}}}$

Ejemplo:

Se han agrupado los hoteles de una región por el número de habitaciones, obteniéndose la siguiente tabla:

Habitaciones	número de hoteles
0 – 100	20
100 – 200	50
200 – 300	60
300 – 400	30
400 – 500	40

⇒

clases	x_i	f_i	$x_i\ f_i$	$x_i^2 f_i$
[0,100)	50	20	1 000	50 000
[100,200)	150	50	7 500	1 125 000
[200,300)	250	60	15 000	3 750 000
[300,400)	350	30	10 500	3 675 000
[400,500)	450	40	18 000	8 100 000
Total		n = 200	52 000	16 700 000

Rango: 500 – 0 = 500 $\overline{x} = \frac{52\,000}{200} = 260$

$s^2 = \frac{16\,700\,000}{200} - 260^2 = 15\,900$ $s = \sqrt{15\,900} \cong 126{,}0952$ $C.V. = \frac{126{,}0952}{260} = 0{,}485$

ACTIVIDADES

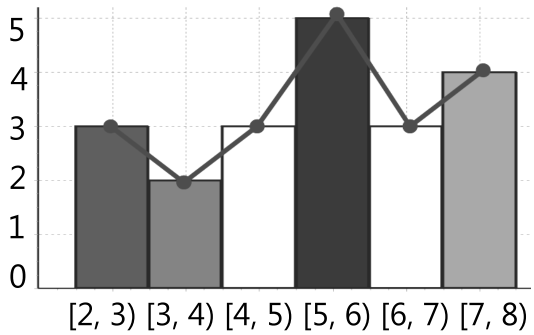
1.- Calcula la mediana de los siguientes datos:
a) Notas de 12 alumnos: 7 , 4 , 6 , 5 , 7 , 7 , 8 , 5 , 8 , 4 , 4 , 5
b) Edades de alumnos: 15, 12, 17, 15, 14, 14, 17, 15 , 15
c) puntuación de un test que se valora de 1 a 5: 4, 1, 2, 4, 2, 5, 3, 4, 4, 1
2.- La profesora de Lengua hizo un examen a sus alumnos.
a) Completa la siguiente tabla

(intervalos de notas) Clases	(marca de clase) x_i	nº de alumnos f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[1,3)		2		
[3,5)		6		
[5,7)		8		
[7,9)		4		
	Total	20		

b) Calcula: 1) La media aritmética 2) La varianza 3) La desviación típica 4) El coeficiente de variación
3.- Completa la siguiente tabla y después halla los cuartiles y dibuja el diagrama de caja correspondiente a las notas en inglés de 20 alumnos:

notas: x_i	nº de alumnos: f_i	F_i	h_i (en %)	H_i (en %)
2	3		15	
3	2		10	
4	3		15	
5	5		25	
6	3		15	
7	4		20	
Total	20		100	

4.- Este es el histograma de frecuencias absolutas de las notas de un grupo de alumnos en Matemáticas

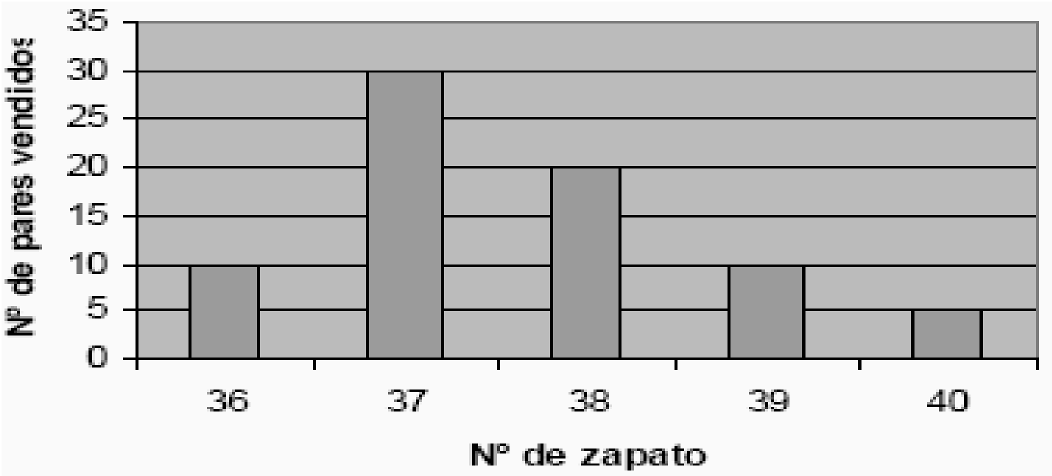


a) Completa la siguiente tabla:

intervalos de notas, clases	marca de clase, x_i	nº de alumnos, f_i	F_i	h_i (en %)	H_i (en %)	$x_i \cdot f_i$	$x_i^2 \cdot f_i$

b) Calcula: 1) El rango 2) La media aritmética 3) La varianza 4) La desviación típica
5) El coeficiente de variación 6) Los cuartiles 7) El rango intercuartílico
c) Dibuja el diagrama de caja

5.- Este es el diagrama de barras de frecuencias absolutas de los pares de zapatos vendidos en una tienda



a) Completa la siguiente tabla:

x_i	f_i	F_i	h_i (en %)	H_i (en %)

b) Completa esta tabla y después dibuja el diagrama de sectores:

x_i	Ángulo del sector

c) Indica cuál es la moda

d) Indica qué porcentaje de personas compran el nº 37

Actividades del libro: 36 (pág. 283) y 65 (pág. 290)