



1. El número de institutos de enseñanza secundaria en 20 ciudades es:

3 1 4 3 2 1 4 4 3 2  
5 2 3 1 3 2 4 2 3 1

- a) Construye una tabla de frecuencias.
- b) Representa gráficamente los datos mediante un diagrama de barras.

2. La siguiente tabla recoge el número de personas que leen prensa en papel al menos una vez al mes, según el nivel de estudios, sobre una muestra de 150 personas.

NIVEL DE ESTUDIOS	NÚMERO DE PERSONAS
ESO	18
Bachillerato	41
Formación profesional	39
Enseñanza universitaria	52

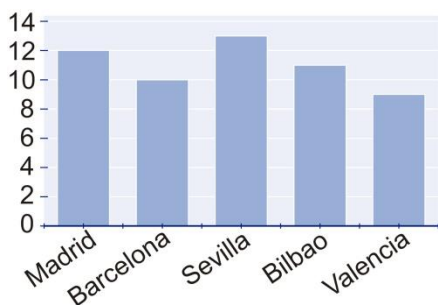
Representa gráficamente los datos en un diagrama de sectores indicando los porcentajes de cada dato.

3. Los puntos obtenidos por un equipo de baloncesto en los 34 partidos de la fase regular de la liga son:

70 90 90 93 90 96 90 81 89 84 79 91 67 92 79 75 83  
88 86 86 99 94 81 87 79 92 80 91 92 78 85 84 81 80

- a) Agrupa los datos en intervalos de igual amplitud.
- b) Construye una tabla de frecuencias.
- c) Representa gráficamente los datos mediante el histograma de frecuencias absolutas.
- d) Representa el polígono de frecuencias del histograma anterior.

4. El siguiente gráfico recoge el número de votos recibido por cada una de las ciudades candidatas a celebrar las olimpiadas de matemáticas.



- a) Construye la tabla de frecuencias.
- b) Representa gráficamente los datos mediante el histograma de frecuencias absolutas y su polígono de frecuencias.

5. Una variable estadística  $X$  tiene la siguiente distribución de frecuencias.

Clase	[0,4)	[4,7)	[7,15)	[15,20)	[20,28)	[28,35)
$f_i$	4	3	6	3	3	1

Construye el histograma de frecuencias absolutas y el polígono de frecuencias.



Para realizar un estudio completo de una variable estadística hay que analizar las **medidas de localización** (media, moda, mediana y cuantiles), que proporcionan una idea de dónde se sitúan las observaciones y de cuál es el centro de la muestra, y las **medidas de dispersión** (recorrido, rango intercuartílico, desviación absoluta media, varianza, desviación típica y coeficiente de variación para variables positivas), que proporcionan información sobre cuánto se alejan los datos respecto a los valores centrales. Fíjate en el estudio de la siguiente variable cuantitativa discreta cuya tabla de frecuencias es:

$x_i$	13	15	17	19
$f_i$	5	6	4	3

Se construye la tabla con todos los valores necesarios para aplicar las fórmulas de las medidas de localización y de dispersión. Asimismo, para calcular los cuantiles, es necesario que los datos estén ordenados de mayor a menor. En este ejemplo vamos a calcular los cuartiles.

$x_i$	$f_i$	$F_i$	$f_i x_i$	$f_i x_i^2$	$f_i  x_i - x $
13	5	5	65	845	13
15	6	11	90	1350	3,6
17	4	15	68	1156	5,6
19	3	18	57	1083	10,2
	18		280	4434	32,4

**Medidas de localización:**

$$\bar{x} = \frac{280}{18} = 15,6, M_o = 15, M = 15, Q_1 = 13, Q_2 = M, Q_3 = 17$$

**Medidas de dispersión:**

$$RIC = 17 - 13 = 4, D_x = \frac{32,4}{18} = 1,8, s^2 = \frac{4434}{18} - (15,6)^2 = 2,97$$

$$s = \sqrt{2,97} = 1,72, CV = \frac{1,72}{15,6} = 0,11$$

- Las pulsaciones en reposo ( $X$ ) de 30 personas que practican *running* habitualmente se presentan agrupadas en la siguiente tabla:

$X$	[40,50)	[50,60)	[60,70)	[70,80)
$f_i$	4	9	12	5

- Calcula la media, la mediana y el intervalo modal.
  - Halla los deciles 3 y 8.
  - Calcula la desviación media y la desviación típica. Compáralas.
- En la siguiente tabla se han agrupado el número de respuestas acertadas de un examen tipo test ( $X$ ) de 34 preguntas realizado por 40 personas.

$X$	[0,4)	[4,7)	[7,15)	[15,20)	[20,28)	[28,35)
$f_i$	2	5	8	18	4	3

- Determina la media, el intervalo modal y los cuantiles.
- Halla los percentiles 35 y 80.
- Calcula el rango intercuartílico, la desviación absoluta media, la desviación típica y el coeficiente de variación. ¿Se puede decir que el conjunto de datos es homogéneo?

- Los goles anotados por dos jugadores de balonmano en 10 partidos son:

Jugador A	15	12	11	9	10	9	11	7	8	10
Jugador B	12	9	9	10	15	16	8	10	11	10

- Halla la media, la moda y la mediana de los goles anotados por cada jugador.
- Calcula la desviación absoluta media, el rango intercuartílico y la desviación típica para cada jugador. ¿Cuál de los jugadores es más variable?

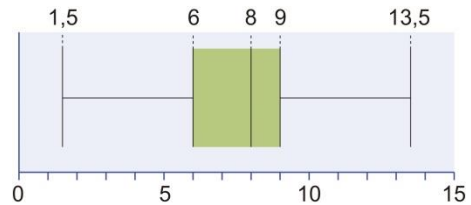


Fíjate en el siguiente ejemplo, en el que se calcula el diagrama de caja y bigotes para la siguiente tabla de distribución de frecuencias de una variable (X) cuantitativa discreta.

<b>X</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
<b>f<sub>i</sub></b>	<b>8</b>	<b>4</b>	<b>2</b>	<b>9</b>	<b>5</b>

	<b>Q<sub>1</sub></b>	<b>M</b>	<b>Q<sub>3</sub></b>	<b>RIC</b>	<b>LI</b>	<b>LS</b>	<b>m</b>	<b>M</b>
<b>X</b>	6	8,5	9	3	1,5	13,5	6	10

DIAGRAMA DE CAJA Y BIGOTES



1. Las siguientes tablas recogen los rendimientos medios (en kilogramos por hectárea) del cultivo en España de trigo (X) y de avena (Y) entre 2000 y 2011.

<b>Año</b>	<b>2000</b>	<b>2001</b>	<b>2002</b>	<b>2003</b>	<b>2004</b>	<b>2005</b>	<b>2006</b>	<b>2007</b>	<b>2008</b>	<b>2009</b>	<b>2010</b>	<b>2011</b>
<b>X</b>	3100	2300	2830	2710	3260	1770	2880	3570	3320	2690	3140	3530

<b>Año</b>	<b>2000</b>	<b>2001</b>	<b>2002</b>	<b>2003</b>	<b>2004</b>	<b>2005</b>	<b>2006</b>	<b>2007</b>	<b>2008</b>	<b>2009</b>	<b>2010</b>	<b>2011</b>
<b>Y</b>	2210	1490	1930	1770	2220	1180	1810	2460	2350	1650	1890	2190

- a) Representa los datos mediante un diagrama de caja y bigotes.
- b) ¿Qué cereal presenta una información más concentrada?

2. El número de libros leídos por un grupo de estudiantes (X) se da en esta tabla:

<b>X</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
<b>f<sub>i</sub></b>	<b>2</b>	<b>13</b>	<b>41</b>	<b>17</b>	<b>10</b>	<b>7</b>

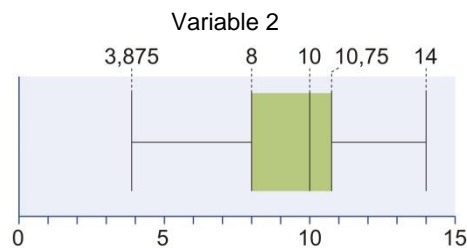
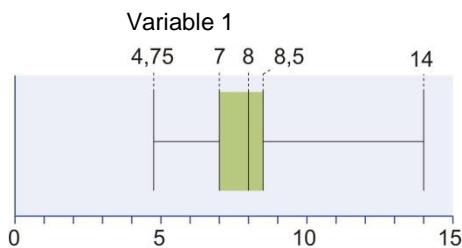
- a) Representa los datos mediante un diagrama de caja y bigotes.
- b) ¿Existen datos atípicos?

3. En la siguiente tabla se han agrupado la edad (X) los empleados de una empresa según su edad.

<b>X</b>	<b>[16,27)</b>	<b>[27,35)</b>	<b>[35,45)</b>	<b>[45,57)</b>	<b>[57,67)</b>
<b>f<sub>i</sub></b>	<b>10</b>	<b>25</b>	<b>22</b>	<b>17</b>	<b>16</b>

- c) Representa los datos mediante un diagrama de caja y bigotes.
- d) Realiza una interpretación del diagrama.

4. Observa los diagramas de caja y bigotes de dos variables.



Realiza un análisis de cada variable respecto a la concentración de los datos.



**Los cuartiles**  $Q_1$ ,  $Q_2$  y  $Q_3$  son los valores que dividen el conjunto de datos en cuatro partes iguales es decir, si  $m$  es el dato menor, y  $M$ , el dato mayor, los intervalos  $[m, Q_1]$   $[Q_1, Q_2]$   $[Q_2, Q_3]$  y  $[Q_3, M]$  contienen el 25% de datos cada uno.

**Los deciles**  $D_1, D_2, \dots, D_9$  son los valores que dividen el conjunto de datos en diez partes iguales es decir, si  $m$  es el dato menor, y  $M$ , el dato mayor, los intervalos  $[m, D_1]$   $[D_1, D_2], \dots, [D_9, M]$  contienen el 10% de datos cada uno.

**Los percentiles**  $P_1, P_2, \dots, P_{99}$  son los valores que dividen el conjunto de datos en cien partes iguales es decir, si  $m$  es el dato menor, y  $M$ , el dato mayor, los intervalos  $[m, P_1]$   $[P_1, P_2], \dots, [P_{99}, M]$  contienen el 1% de datos cada uno.

Para calcular el cuartil  $Q_i$  ( $i = 1, 2, 3$ ) en una distribución de datos agrupados en intervalos se procede de la siguiente forma:

Con la ayuda de la columna de frecuencias relativas acumuladas, se localiza el intervalo  $[E_h, E_{h+1}]$  que sobrepasa, según cada caso, los valores 0,25; 0,50 o 0,75. Este intervalo contendrá el cuartil buscado.

Para obtener el cuartil  $Q_i$  se utiliza la siguiente fórmula de interpolación lineal:

$$Q_i = E_h + \frac{(p - H_{h-1}) \cdot l_h}{h_h} \quad \text{donde} \quad \begin{cases} E_h & \text{Extremo inferior del intervalo } [E_h, E_{h+1}] \\ p = 0,25 \ 0,50 \ 0,75 & \text{según los casos} \\ H_{h-1} & \text{Frecuencia relativa acumulada del intervalo anterior al } [E_h, E_{h+1}] \\ l_h & \text{Longitud del intervalo } [E_h, E_{h+1}] \\ h_h & \text{frecuencia relativa del intervalo } [E_h, E_{h+1}] \end{cases}$$

Para el cálculo de deciles o percentiles, se utiliza la misma fórmula pero cambiando  $p$  por el porcentaje correspondiente (por ejemplo 0,20 para  $D_2$  ó 0,54 para el  $P_{54}$ )

**1. Se quiere estudiar el tiempo de conexión durante el fin de semana a Internet por parte de los alumnos de 1º de bachillerato de un instituto. Para ello se pregunta a 40 estudiantes elegidos al azar y se obtienen los siguientes resultados en minutos.**

90	30	55	40	70
45	45	15	75	65
60	35	75	30	90
115	20	90	50	65

- Calcula los tres cuartiles ordenando dichos datos de menor a mayor.
- Elabora la tabla de frecuencias con los datos agrupados en cuatro intervalos:  
[0, 30) [30, 60) [60, 90) y [90, 120]
- Calcula los tres cuartiles teniendo en cuenta la agrupación hecha en intervalos.
- Compara los resultados de a) y c) y decide si la pérdida de información por haber agrupado los datos ha influido de forma determinante en el resultado de los cuartiles.

**2. Para la tabla de datos agrupados en intervalos del ejercicio anterior:**

- Calcula los deciles  $D_6$  y  $D_9$ .
- Calcula los percentiles  $P_{55}$  y  $P_{95}$ .
- ¿Cuánto tiempo debe estar conectado un alumno para que se encuentre en un lugar superior al percentil  $P_{90}$ ?