

ACTIVIDADES

1. En una revista leemos que el pastor alemán tiene una alzada media de 55 cm. ¿Crees que han medido a todos los pastores alemanes del planeta? Explica cómo crees que han llegado a esta conclusión.

No se ha medido a todos los pastores alemanes.

Para llegar a esa conclusión se ha realizado un estudio estadístico donde se habrá medido una muestra representativa de pastores alemanes.

2. Analiza si tiene sentido elegir una muestra para realizar un estudio estadístico sobre la altura de un grupo de 10 alumnos.

No tiene sentido porque medir a 10 alumnos no resulta muy costoso.

3. Indica el tipo de variable estadística que estamos estudiando en cada uno de los siguientes casos.

- a) El número de libros que lee cada uno de los compañeros de tu clase.
- b) El color de la camiseta de los clubes deportivos de una ciudad.
- c) La distancia al trabajo de los vecinos de un edificio.

a) Cuantitativa discreta.

b) Cualitativa.

c) Cuantitativa continua.

4. En una clase con 30 alumnos se realiza una encuesta sobre el número de aplicaciones para el *smartphone* que han comprado en el último mes, y se han obtenido estas respuestas.

5 2 3 1 6 6 2 0 1 3 0 1 1 2 3
2 0 1 1 3 3 2 1 3 1 5 2 1 4 3

Organiza estos datos en una tabla de frecuencias.

Nº aplicaciones	f_i	h_i	F_i	H_i
6	2	0,067	2	0,067
5	2	0,067	4	0,134
4	1	0,033	5	0,167
3	7	0,233	12	0,4
2	6	0,2	18	0,6
1	9	0,3	27	0,9
0	3	0,1	30	1
Total	30	1		

5. Completa la tabla de frecuencias de la estatura del grupo de personas que aparece a continuación.

Estatura (cm)	[165, 175)	[175, 185)	[185, 195)
N.º de personas	40	85	25

¿Qué porcentaje de personas miden entre 165 y 175 cm?

Estatura	x_i	f_i	h_i	F_i	H_i
[165, 175)	170	40	0,27	40	0,27
[175, 185)	180	85	0,57	125	0,83
[185, 195)	190	25	0,17	150	1
		$N = 150$	$\sum h_i = 1$		

El porcentaje de personas que miden entre 165 cm y 175 cm es del 27 %.

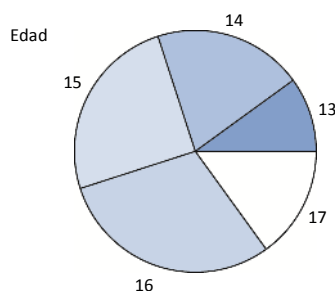
6. En el polideportivo del barrio se ha formado un equipo juvenil de atletismo. Los miembros de este equipo tienen las siguientes edades.

16 17 14 16 13 15 15 17 17 14
 16 16 16 14 15 13 15 16 14 15

Representa estos datos en un gráfico que consideres apropiado utilizando sus frecuencias relativas.

Edad	f_i	h_i
13	2	0,1
14	4	0,2
15	5	0,25
16	6	0,3
17	3	0,15
Total	20	1

Como queremos utilizar las frecuencias relativas, dibujamos un diagrama de sectores.

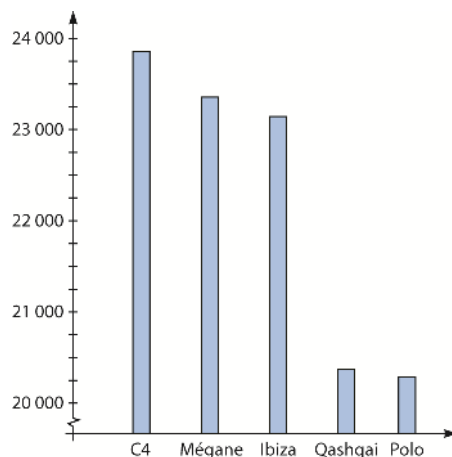


7. Sobre un total de 722 703 coches, los cinco modelos más vendidos en 2013 son los siguientes.

Represéntalos como consideres más adecuado.

Modelo	Unidades
Citroën C4	23837
Renault Mégane	23310
Seat Ibiza	23141
Nissan Qashqai	20372
Volkswagen Polo	20283

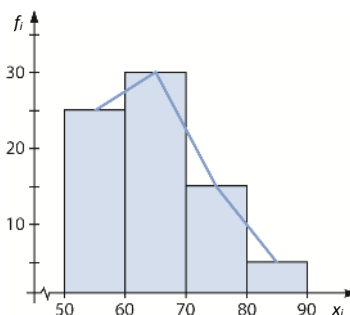
Representamos los datos en un diagrama de barras:



8. Completa la tabla de frecuencias y dibuja el histograma y el polígono de frecuencias.

Peso	[50, 60)	[60, 70)	[70, 80)	[80, 90)
f_i	25	30	15	5

x_i	[50, 60)	[60, 70)	[70, 80)	[80, 90)	Total
f_i	25	30	15	5	75
h_i	0,333	0,4	0,2	0,067	1

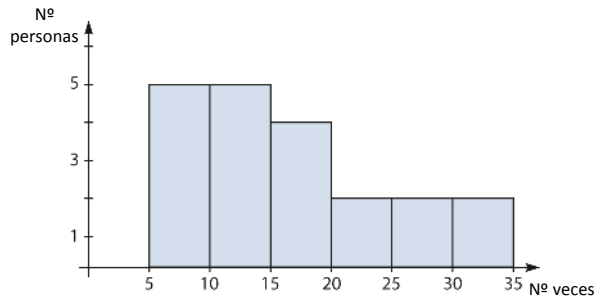


9. El número de veces que asistieron al cine 20 personas fueron:

12 15 33 8 19 7 16 11 6 21
24 25 14 9 31 26 15 9 10 11

Agrupar los datos en intervalos de amplitud 5 y dibuja un histograma.

Nº veces	[5, 10)	[10, 15)	[15, 20)	[20, 25)	[25, 30)	[30, 35)	Total
f_i	5	5	4	2	2	2	20
h_i	0,25	0,25	0,2	0,1	0,1	0,1	1



10. Una encuesta realizada a 10 pilotos, en la que se les pregunta sobre su número de vuelos semanales, muestra los siguientes datos.

N.º de vuelos	0	1	2	3
N.º de pilotos	2	4	3	1

Calcula e interpreta las medidas de centralización.

- $Mo = 1$ → Lo más frecuente es que los pilotos vuelen una vez por semana.
- $Me = 1$ → El valor central es 1 vuelo, es decir, hay tantos pilotos que vuelan una o más veces, como que vuelan una o menos veces.
- $\bar{X} = \frac{0 \cdot 2 + 1 \cdot 4 + 2 \cdot 3 + 3 \cdot 1}{10} = 1,3$ vuelos de media realiza cada piloto a la semana.

11. A partir de estos datos relativos al número de horas de estudio semanales de un grupo de 25 alumnos, construye la tabla de frecuencias sin agrupar los datos en intervalos y calcula las medidas de centralización.

2 2 1 0 1 4 3 2 2 4 0 1 1 3 1
2 2 3 4 5 1 1 1 0 3 4 0 0 2 3

Interpreta los resultados obtenidos.

Horas de estudio	0	1	2	3	4	5	Total
f_i	5	8	7	5	4	1	30
F_i	5	13	20	25	29	30	
h_i	0,167	0,267	0,233	0,167	0,133	0,033	1

Calculamos las medidas de centralización:

- $Mo = 1$ hora \rightarrow Lo más frecuente es que los alumnos estudien una hora por semana.
- $Me = 2$ \rightarrow El valor central es 2 horas, es decir, hay tantos alumnos que estudian 2 o más horas, como que estudian 2 o menos horas.
- $\bar{X} = \frac{0 \cdot 5 + 1 \cdot 8 + 2 \cdot 7 + 3 \cdot 5 + 4 \cdot 4 + 5 \cdot 1}{30} = 1,9\bar{3}$ horas de media estudian los alumnos a la semana.

12. Obtén las medidas de centralización de estos datos.

Datos	[1, 3)	[3, 5)	[5, 7)	[7, 9)	[9, 11)
f_i	3	7	8	4	3

Las marcas de clase son, respectivamente, 2, 4, 6, 8 y 10. Así:

$$Mo = 6 \qquad Me = 6 \qquad \bar{X} = \frac{2 \cdot 3 + 4 \cdot 7 + 6 \cdot 8 + 8 \cdot 4 + 10 \cdot 3}{25} = 5,76$$

13. Halla las medidas de centralización de estos datos.

Datos	[0, 4)	[4, 8)	[8, 12)	[12, 16)	[16, 20)
f_i	5	2	1	4	6

Las marcas de clase son, respectivamente, 2, 6, 10, 14 y 18. Así:

$$Mo = 18 \qquad Me = 14 \qquad \bar{X} = \frac{2 \cdot 5 + 6 \cdot 2 + 10 \cdot 1 + 14 \cdot 4 + 18 \cdot 6}{18} = 10,8$$

14. Calcula e interpreta los cuartiles de las siguientes notas en el examen de Matemáticas.

Notas	1	2	3	4	5	6	7	8	9	10
N.º alumnos	3	6	9	8	11	13	15	11	9	5

Calculamos las frecuencias acumuladas:

$$F_1 = 3 \qquad F_2 = 9 \qquad F_3 = 18 \qquad F_4 = 26 \qquad F_5 = 37$$

$$F_6 = 50 \qquad F_7 = 65 \qquad F_8 = 76 \qquad F_9 = 85 \qquad F_{10} = 90$$

$$25\% \text{ de } 90 = 90 \cdot 0,25 = 22,5 \qquad 50\% \text{ de } 90 = 90 \cdot 0,5 = 45 \qquad 75\% \text{ de } 90 = 90 \cdot 0,75 = 67,5$$

- $Q_1 \rightarrow$ De las frecuencias acumuladas, la primera que es mayor que 22,5 es F_4 , por tanto, $Q_1 = 4$.
- $Q_2 \rightarrow$ De las frecuencias acumuladas, la primera que es mayor que 45 es F_6 , por tanto, $Q_2 = Me = 6$.
- $Q_3 \rightarrow$ De las frecuencias acumuladas, la primera que es mayor que 67,5 es F_8 , por tanto, $Q_3 = 8$.

15. Observa los pesos de bebés al nacer y halla los percentiles P_{27} , P_{50} y P_{90} .

Peso (kg)	[2, 2,5)	[2,5, 3)	[3, 3,5)	[3,5, 4)	[4, 4,5)
N.º bebés	3	6	9	8	11

Completamos la tabla con las frecuencias acumuladas:

Peso (kg)	[2; 2,5)	[2,5; 3)	[3; 3,5)	[3,5; 4)	[4; 4,5)	Total
x_i	2,25	2,75	3,25	3,75	4,25	
f_i	3	6	9	8	11	37
F_i	3	9	18	26	37	

27% de 37 = 9,99

50% de 37 = 18,5

90% de 37 = 33,3

$P_{27} \rightarrow$ De las frecuencias acumuladas, la primera que es mayor que 9,99 es F_3 , por tanto, $P_{27} = 3,25$.

$P_{50} \rightarrow$ De las frecuencias acumuladas, la primera que es mayor que 18,5 es F_4 , por tanto, $P_{50} = 3,75$.

$P_{90} \rightarrow$ De las frecuencias acumuladas, la primera que es mayor que 33,3 es F_5 , por tanto, $P_{90} = 4,25$.

16. Calcula las medidas de dispersión de estos datos.

Datos (x_i)	1	2	3	4	5
Frec. (f_i)	1	5	7	2	1

El número total de datos es $N = 16$.

$$\bar{x} = \frac{1 \cdot 1 + 2 \cdot 5 + 3 \cdot 7 + 4 \cdot 2 + 5 \cdot 1}{16} = 2,8125$$

$Rango = 5 - 1 = 4$

$$DM = \frac{|1 - 2,8125| \cdot 1 + |2 - 2,8125| \cdot 5 + |3 - 2,8125| \cdot 7 + |4 - 2,8125| \cdot 2 + |5 - 2,8125| \cdot 1}{16} = 0,734375$$

$$\sigma^2 = \frac{1^2 \cdot 1 + 2^2 \cdot 5 + 3^2 \cdot 7 + 4^2 \cdot 2 + 5^2 \cdot 1}{16} - 2,8125^2 = 0,9$$

$\sigma = 0,95$

$$CV = \frac{0,95}{2,8125} = 0,34$$

17. Calcula las medidas de dispersión de estos datos.

Datos (x_i)	[0, 5)	[5, 10)	[10, 15)	[15, 20)
Frec. (f_i)	5	6	4	2

Las marcas de clase son:

$x_1 = 2,5$

$x_2 = 7,5$

$x_3 = 12,5$

$x_4 = 17,5$

El número total de datos es $N = 17$.

$$\bar{x} = \frac{2,5 \cdot 5 + 7,5 \cdot 6 + 12,5 \cdot 4 + 17,5 \cdot 2}{17} = 8,38$$

$Rango = 20 - 0 = 20$

$$DM = \frac{|2,5 - 8,38| \cdot 5 + |7,5 - 8,38| \cdot 6 + |12,5 - 8,38| \cdot 4 + |17,5 - 8,38| \cdot 2}{17} = 4,08$$

$$\sigma^2 = \frac{2,5^2 \cdot 5 + 7,5^2 \cdot 6 + 12,5^2 \cdot 4 + 17,5^2 \cdot 2}{17} - 8,38^2 = 24,22$$

$\sigma = 4,92$

$$CV = \frac{4,92}{8,38} = 0,59$$

18. Tras realizar un estudio estadístico se obtienen los siguientes datos.

1 9 8 1 2 9 7 5 10 1

Calcula las medidas de centralización y de dispersión de estos datos. ¿Qué conclusiones puedes extraer al comparar conjuntamente las medidas que has calculado?

Contruimos la tabla de frecuencias:

x_i	1	2	5	7	8	9	10	Total
f_i	3	1	1	1	1	2	1	10
F_i	3	4	5	6	7	9	10	

El número total de datos es $N = 10$.

$$Mo = 1 \qquad Me = 5 \qquad \bar{x} = \frac{1 \cdot 3 + 2 \cdot 1 + 5 \cdot 1 + 7 \cdot 1 + 8 \cdot 1 + 10 \cdot 2 + 10 \cdot 1}{10} = 5,3$$

$$Rango = 10 - 1 = 9$$

$$DM = \frac{|1 - 5,3| \cdot 3 + |2 - 5,3| \cdot 1 + |2 - 5,3| \cdot 1 + |2 - 5,3| \cdot 1 + |5 - 5,3| \cdot 1 + |7 - 5,3| \cdot 1 + |8 - 5,3| \cdot 1 + |9 - 5,3| \cdot 2 + |10 - 5,3| \cdot 1}{10} = 3,3$$

$$\sigma^2 = \frac{1^2 \cdot 3 + 2^2 \cdot 1 + 5^2 \cdot 1 + 7^2 \cdot 1 + 8^2 \cdot 1 + 9^2 \cdot 2 + 10^2 \cdot 1}{10} - 5,3^2 = 12,61 \qquad \sigma = 3,55 \qquad CV = \frac{3,55}{5,3} = 0,67$$

El valor medio de los datos es 5,3 pero podemos deducir que, dado lo elevado de la desviación típica y del coeficiente de variación, los datos están muy dispersos con respecto a la media.

19. La siguiente tabla muestra la antigüedad de los vehículos matriculados en un municipio.

Antigüedad (años)	[0, 4)	[4, 8)	[8, 10)	[10, 12)
N.º de vehículos	346	521	382	151

Halla y analiza las medidas estadísticas.

Completamos la tabla de frecuencias:

Antigüedad (años)	[0, 4)	[4, 8)	[8, 10)	[10, 12)	Total
x_i	2	6	9	11	
f_i	346	521	382	151	1400
F_i	346	867	1249	1400	

El número total de datos es $N = 1400$.

$$Mo = 6 \qquad Me = 6 \qquad \bar{x} = \frac{2 \cdot 346 + 6 \cdot 521 + 9 \cdot 382 + 11 \cdot 151}{1400} = 6,367$$

$$Rango = 12 - 0 = 12$$

$$DM = \frac{|2 - 6,367| \cdot 346 + |6 - 6,367| \cdot 521 + |9 - 6,367| \cdot 382 + |11 - 6,367| \cdot 151}{1400} = 2,43$$

$$\sigma^2 = \frac{2^2 \cdot 346 + 6^2 \cdot 521 + 9^2 \cdot 382 + 11^2 \cdot 151}{1400} - 6,367^2 = 8,97 \qquad \sigma = 2,995 \qquad CV = \frac{2,995}{6,367} = 0,47$$

La antigüedad media de los vehículos del municipio es 6,367 años, pero podemos deducir que, dado el elevado valor de la desviación típica y del coeficiente de variación, los datos están bastante dispersos con respecto a la media.

20. Estudia la dispersión de los datos.

Intervalos	[0, 5)	[5, 15)	[15, 35)	[35, 65)
Porcentaje	22	50	25	3

Los datos son frecuencias relativas dadas en porcentajes. Completamos la tabla de frecuencias:

Antigüedad (años)	[0, 5)	[5, 15)	[15, 35)	[35, 65)
x_i	2,5	10	25	50
h_i	0,22	0,5	0,25	0,03
%	22	50	25	3

Calculamos las medidas estadísticas:

$$\bar{x} = \sum_{i=1}^n h_i \cdot x_i = 2,5 \cdot 0,22 + 10 \cdot 0,5 + 25 \cdot 0,25 + 50 \cdot 0,03 = 13,3$$

$$\sigma^2 = \sum_{i=1}^n x_i^2 \cdot h_i - \bar{x}^2 = (2,5^2 \cdot 0,22 + 10^2 \cdot 0,5 + 25^2 \cdot 0,25 + 50^2 \cdot 0,03) - 13,3^2 = 105,735 \quad \sigma = 10,28$$

$$CV = \frac{10,28}{13,3} = 0,77$$

Podemos deducir que los datos están muy dispersos con respecto a la media, que es 13,3.

21. Di en cuál de las muestras están los datos más concentrados y si existe algún dato atípico.

Muestra A: 8 6 9 6 2 3 6 7

Muestra B: 11 13 15 13 11 12 13 22

▪ Muestra A:

x_i	2	3	6	7	8	9	Total
f_i	1	1	3	1	1	1	8

$$\bar{x}_A = \frac{2 \cdot 1 + 3 \cdot 1 + 6 \cdot 3 + 7 \cdot 1 + 8 \cdot 1 + 9 \cdot 1}{8} = 5,875$$

$$\sigma_A^2 = \frac{2^2 \cdot 1 + 3^2 \cdot 1 + 6^2 \cdot 3 + 7^2 \cdot 1 + 8^2 \cdot 1 + 9^2 \cdot 1}{8} - 5,875^2 = 4,86 \quad \sigma_A = 2,2 \quad CV_A = \frac{2,2}{5,875} = 0,38$$

Analizamos la existencia de datos atípicos. Los datos que no se encuentren en este intervalo podrían considerarse atípicos:

$$(\bar{x}_A - 3\sigma_A, \bar{x}_A + 3\sigma_A) = (-0,725; 15,475) \rightarrow \text{No hay datos atípicos.}$$

▪ Muestra B:

x_i	11	12	13	15	22	Total
f_i	2	1	3	1	1	8

$$\bar{x}_B = \frac{11 \cdot 2 + 12 \cdot 1 + 13 \cdot 3 + 15 \cdot 1 + 22 \cdot 1}{8} = 13,75$$

$$\sigma_B^2 = \frac{11^2 \cdot 2 + 12^2 \cdot 1 + 13^2 \cdot 3 + 15^2 \cdot 1 + 22^2 \cdot 1}{8} - 13,75^2 = 11,1875 \quad \sigma_B = 3,345 \quad CV_B = \frac{3,345}{13,75} = 0,24$$

Analizamos la existencia de datos atípicos. Los datos que no se encuentren en este intervalo podrían considerarse atípicos:

$$(\bar{x}_B - 3\sigma_B, \bar{x}_B + 3\sigma_B) = (3,715; 23,785) \rightarrow \text{No hay datos atípicos.}$$

Analizando las medidas estadísticas obtenidas, podemos concluir que los datos están más concentrados alrededor de la media en la muestra B.

SABER HACER

22. Organiza los datos en una tabla de frecuencias, agrupándolos en intervalos.

61 72 63 80 45 57 66 77 39 57 43 42 61 70 34
69 68 49 39 54 58 69 73 41 53 81 52 40 45 45

Calculamos el número de intervalos que utilizaremos:

$$\sqrt{30} \approx 5,5 \rightarrow 6 \text{ intervalos}$$

Calculamos la amplitud de cada intervalo:

$$\frac{Max - Min}{\sqrt{N}} = \frac{81 - 34}{\sqrt{30}} \approx 8,58 \rightarrow \text{La amplitud de los intervalos será 9.}$$

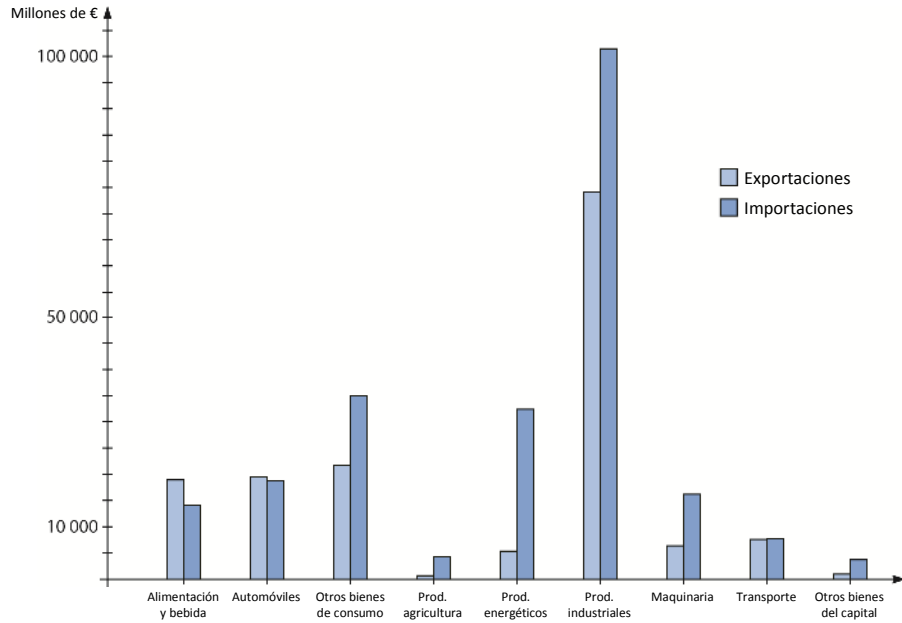
Contruimos la tabla de frecuencias:

Datos	x_i	f_i	F_i	h_i	H_i
[34, 43)	38,5	6	6	0,2	0,2
[43, 52)	47,5	5	11	0,167	0,367
[52, 61)	56,5	6	17	0,2	0,567
[61, 70)	65,5	7	24	0,233	0,8
[70, 79)	74,5	4	28	0,133	0,933
[79, 88)	83,5	2	30	0,067	1
Total		30		1	

23. En un estudio sobre el comercio exterior, la tabla recoge los datos obtenidos, en millones de euros, en importaciones y exportaciones durante un año.

	Exportaciones	Importaciones
Alimentos y bebida	18678	14029,5
Automóviles	19346,5	18491,7
Otros bienes de consumo	21571,6	34466,3
Productos de agricultura	695,5	4317,3
Productos energéticos	5302	31957,9
Productos industriales	73109,8	100808,1
Maquinaria	6297,5	16061,4
Transporte	7540,3	7568,3
Otros bienes de capital	1017,9	3671,2

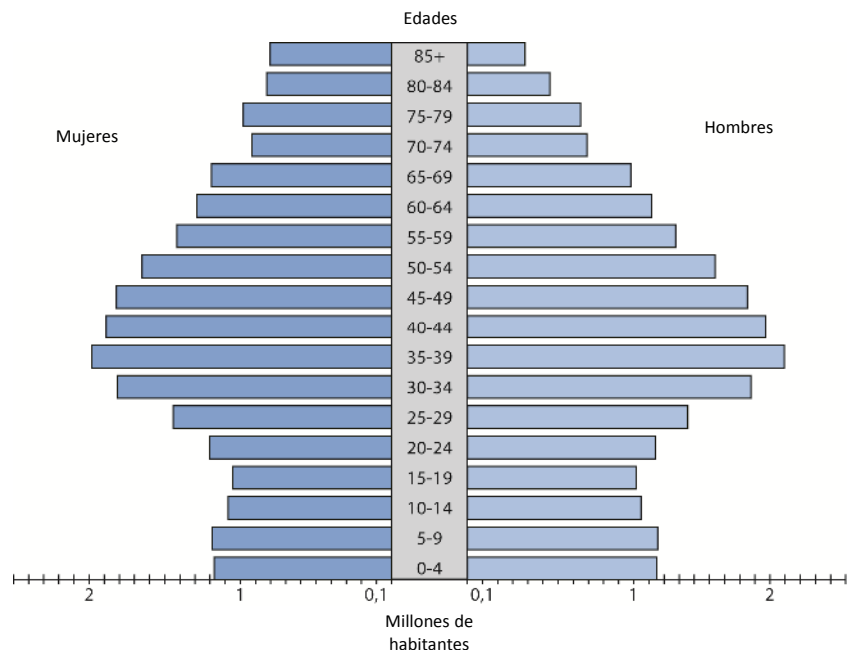
Construye un diagrama de barras adosadas e interprétalo.



Este diagrama contrasta las exportaciones e importaciones realizadas en un país en un año. Las importaciones tienen mucho más peso que las exportaciones, salvo en *Alimentación y bebidas*, donde destacan las exportaciones, y en *Automóviles* y *Transporte*, donde se encuentra parejo. El país posee un déficit comercial.

24. Dibuja y analiza la pirámide de población del año 2013.

Edad	Hombres	Mujeres
0-4	1250261	1172504
5-9	1256481	1184050
10-14	1146098	1080604
15-19	1114480	1051130
20-24	1241695	1201940
25-29	1456221	1443412
30-34	1873020	1811758
35-39	2093321	1983801
40-44	1968657	1886012
45-49	1849194	1818983
50-54	1636317	1648642
55-59	1375659	1419284
60-64	1215561	1286727
65-69	1078145	1190748
70-74	789593	924047
75-79	747102	979003
80-84	544043	825164
85 y más	381844	802321



- La esperanza de vida es bastante alta.
- Las mujeres tienen una esperanza de vida mayor que la de los hombres.
- La mayor parte de la población está entre los 30 y los 54 años y la población menor de 20 años es escasa, lo que augura un envejecimiento de la población.

25. Estudia la simetría de estos datos.

x_i	1	2	3	4	5	6	7	8
f_i	1	1	1	2	3	5	9	10

Completamos la tabla con las frecuencias acumuladas:

x_i	1	2	3	4	5	6	7	8	Total
f_i	1	1	1	2	3	5	9	10	32
F_i	1	2	3	5	8	13	22	32	

Calculamos los cuartiles:

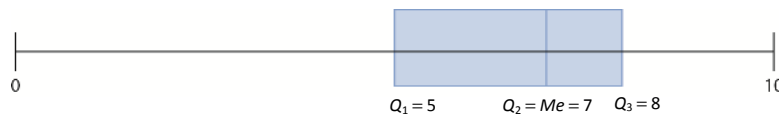
25 % de 32 = 8 50 % de 32 = 16 75 % de 32 = 24

$Q_1 \rightarrow$ De las frecuencias acumuladas, la primera que es mayor que 8 es F_5 , por tanto, $Q_1 = 5$.

$Q_2 \rightarrow$ De las frecuencias acumuladas, la primera que es mayor que 16 es F_7 , por tanto, $Q_2 = Me = 7$.

$Q_3 \rightarrow$ De las frecuencias acumuladas, la primera que es mayor que 24 es F_8 , por tanto, $Q_3 = 8$.

Dibujamos el diagrama de caja:



Los datos mayores que la mediana se agrupan en un intervalo de menor amplitud que los datos inferiores, por lo que los datos presentan una asimetría por la derecha.

26. Determina las medidas estadísticas, utilizando la calculadora, de la tabla que muestra el gasto en agua de los vecinos de un edificio.

N.º de vecinos	3	5	7	10	3	1
Gasto (€)	[20,25)	[25,30)	[30,35)	[35,40)	[40,45)	[45,50)

Las marcas de clase de cada intervalo son:

$x_1 = 22,5$ $x_2 = 27,5$ $x_3 = 32,5$ $x_4 = 37,5$ $x_5 = 42,5$ $x_6 = 47,5$

$\bar{x} = 33,88$ $\sigma^2 = 39,477 \rightarrow \sigma = 6,283$ $CV = \frac{6,283}{33,88} = 0,185$

27. Las medidas estadísticas de las notas de una clase son $\bar{x} = 5,7$; $Me = 4$; $Mo = 4$; $\sigma = 1,70$. Interpretalas.

La nota media de la clase ha sido 5,7, aunque la mayoría ha obtenido un 4 y hay tanta gente que ha obtenido un 4 o más como un 4 o menos.

$CV = \frac{1,70}{5,7} = 0,30$ no es muy alto, lo que indica que las notas se han concentrado sobre el 5,7.

Las tres medidas centrales no están muy próximas, lo que podría significar que ha habido mayoría de suspensos.

28. Los datos muestran las precipitaciones, en ℓ/m^2 , en dos ciudades durante las últimas 6 semanas.

Ciudad A	25	32	22	20	12	19
Ciudad B	2	5	100	0	0	30

¿A y B pertenecen a la misma zona climática? ¿Por qué?

Realizamos un estudio estadístico sobre cada una de las ciudades:

Ciudad A:

Ciudad B:

$$\bar{x}_A = 21,667 \quad \sigma_A = 6,07 \quad CV_A = \frac{6,07}{21,667} = 0,28 \quad \bar{x}_B = 22,833 \quad \sigma_B = 36,057 \quad CV_B = \frac{36,057}{22,833} = 1,58$$

A pesar de que la media de precipitaciones es muy parecida en las dos ciudades, el coeficiente de variación es muchísimo menor en la ciudad A que en la B, lo que significa que las precipitaciones semanales están más cerca de la media o, lo que es lo mismo, son más regulares. Por esto podemos concluir que A y B no pertenecen a la misma zona climática.

29. Completa la tabla y calcula las medidas estadísticas.

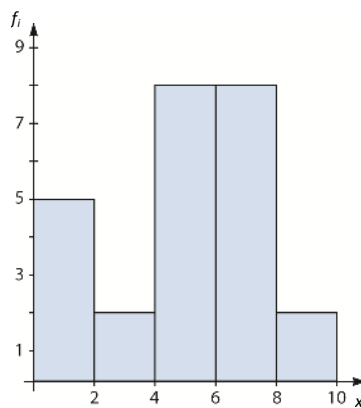
Datos (x_i)	[0, 2)	[2, 4)	[4, 6)	[6, 8)	[8, 10)
Frec. (f_i)	5	2	8	8	2

Datos	x_i	f_i	F_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$
[0, 2)	1	5	5	5	5
[2, 4)	3	2	7	6	18
[4, 6)	5	8	15	40	200
[6, 8)	7	8	23	56	392
[8, 10)	9	2	25	18	162
Total		25		125	777

$$\begin{aligned} \bar{x} &= 5 \\ \sigma^2 &= 6,08 \\ \sigma &= 2,466 \\ CV &= \frac{2,466}{5} = 0,493 \end{aligned}$$

30. Realiza un histograma con estos datos.

Datos (x_i)	[0, 2)	[2, 4)	[4, 6)	[6, 8)	[8, 10)
Frec. (f_i)	5	2	8	8	2



ACTIVIDADES FINALES

31. Queremos hacer un estudio sobre práctica deportiva en los alumnos de la ESO de nuestro instituto preguntándoles cuántos días al mes realizan alguna actividad deportiva.

- a) ¿Qué muestra elegirías para realizar el estudio?
 - b) ¿De qué tamaño es la muestra?
 - c) ¿Cuál es la población?
- a) Se deberían escoger chicos y chicas de los cuatro cursos de la ESO.
 - b) Dependerá de la cantidad de alumnos del instituto.
 - c) Todos los alumnos de la ESO de ese instituto.

32. Indica el tipo de variable estadística que estamos estudiando y razona, en cada caso, qué sería más aconsejable: estudiar una muestra o la población.

- a) El dinero gastado en una semana por tus amigos.
- b) La talla de pantalón de todas las personas de tu ciudad.
- c) El color de ojos de los miembros de tu familia.
- d) La temperatura a las 12 de la mañana de hoy de todos los pueblos de tu provincia.
- e) La estatura de todos los soldados del ejército.
- f) El nivel de colesterol en sangre de todos los españoles.
- g) La edad de todos los alumnos de un colegio.

- a) Cuantitativa continua. Estudiar la población.
- b) Cuantitativa discreta. Estudiar una muestra.
- c) Cualitativa. Estudiar la población.
- d) Cuantitativa continua. Estudiar una muestra.
- e) Cuantitativa continua. Estudiar una muestra.
- f) Cuantitativa continua. Estudiar una muestra.
- g) Cuantitativa discreta. Estudiar una muestra.

33. El entrenador de un club de fútbol ha anotado el número de tarjetas rojas que tuvieron a lo largo de la temporada todos los integrantes de la plantilla obteniendo estos datos.

0 2 4 1 5 0 0 2 1 1 3 4 1 1 0 1 2 0 1 2 4 0 1 2 2

- a) Organiza los resultados en una tabla de frecuencias.
- b) ¿Qué significado tiene en este caso la columna de las frecuencias acumuladas?

a)

Nº tarjetas rojas	f_i	F_i	h_i	H_i
0	6	6	0,24	0,24
1	8	14	0,32	0,56
2	6	20	0,24	0,8
3	1	21	0,04	0,84
4	3	24	0,12	0,96
5	1	25	0,04	1
Total	25		1	

- b) El número de jugadores que tuvieron menos de i tarjetas.

34. Se han perdido algunos datos de esta tabla con las provincias de nacimiento de un grupo de personas. ¿Podrías completarla?

Provincia	f_i	h_i	Porcentaje
Albacete	12	0,2	20%
Ciudad Real	9	0,15	15%
Cuenca	21	0,35	35%
Guadalajara	6	0,1	10%
Toledo	12	0,2	20%
Total	60	1	100%

35. Completa la tabla relativa a la marca de teléfono móvil de un grupo de alumnos.

Marca	f_i	h_i	Porcentaje
Apple	4	0,16	16%
Samsung	12	0,48	48%
LG	3	0,12	12%
Sony	5	0,2	20%
Shen Wai	1	0,04	4%
Total	25	1	100%

36. Determina la moda, la mediana y la media aritmética en las siguientes series de datos.

- a) 7, 9, 7, 9, 7, 9, 7, 9 b) 2, 4, 8, 4, 5, 5, 7 c) 12, 2, 6, 5, 9, 5, 3 d) 3, 5, 9, 5, 6, 6, 9, 5 e) 6, 6, 5, 5, 4, 8, 4, 2

a)

x_i	7	9	Total
f_i	4	4	8
F_i	4	8	

$Mo = \{7, 9\}$ $Me = 8$ $\bar{x} = 8$

b)

x_i	2	4	5	7	8	Total
f_i	1	2	2	1	1	7
F_i	1	3	5	6	7	

$Mo = \{4, 5\}$ $Me = 5$ $\bar{x} = 5$

c)

x_i	2	3	5	6	9	12	Total
f_i	1	1	2	1	1	1	7
F_i	1	2	4	5	6	7	

$Mo = 5$ $Me = 5$ $\bar{x} = 6$

d)

x_i	3	5	6	9	Total
f_i	1	3	2	2	7
F_i	1	4	6	8	

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

e)

x_i	2	4	5	6	8	Total
f_i	1	2	2	2	1	8
F_i	1	3	5	7	8	

$Mo = \{4, 5, 6\}$ $Me = 5$ $\bar{x} = 5$

37. Para una muestra de 572 conductores de autobús se registra el número de accidentes en los que han estado implicados durante 4 años.

N.º de accidentes	0	1	2	3	4	5	6	7	8	9
N.º de conductores	121	156	130	60	57	23	12	9	3	1

- a) Realiza una tabla de frecuencias.
- b) Calcula las medidas de centralización.
- c) Halla los valores de las medidas de dispersión.

a)

x_i	0	1	2	3	4	5	6	7	8	9	Total
f_i	121	156	130	60	57	23	12	9	3	1	572
F_i	121	277	407	467	524	547	559	568	571	572	
h_i	0,212	0,273	0,227	0,105	0,1	0,04	0,021	0,016	0,005	0,001	1
H_i	0,212	0,485	0,712	0,817	0,917	0,957	0,978	0,994	0,999	1	

b) $Mo = 1$ $Me = 2$ $\bar{x} = \frac{1107}{572} = 1,935$

c) $DM = \frac{760,164}{572} = 1,329$ $\sigma^2 = \frac{3849}{572} - 1,935^2 = 2,984$ $\sigma = 1,727$ $CV = \frac{1,727}{1,935} = 0,89$

38. El número de hermanos, excluidos ellos, de un grupo de alumnos de 1.º de Bachillerato se han recogido en la tabla siguiente.

N.º de hermanos	0	1	2	3	4	5	6
f_i	7	12	6	3	1	0	1

Halla las medidas de centralización.

El número total de datos es $N = 30$.

$Mo = 1$ hermano $Me = 1$ hermano $\bar{x} = \frac{43}{30} = 1,4\hat{3}$ hermanos

39. Añade un dato a este conjunto para que la media sea 6. ¿Varía la mediana?

3 3 3 4 4 4 5 5 5 6 6 6 7 7 7

La media de los datos actuales es $\bar{x} = \frac{3 \cdot (3+4+5+6+7)}{15} = 5$

$N = 16 \rightarrow 6 \cdot 16 = 96$ $96 - 3 \cdot (3+4+5+6+7) = 21 \rightarrow$ El nuevo dato debería ser 21.

40. En una encuesta se preguntó por el número de televisores en funcionamiento que hay en los hogares. Las respuestas están recogidas en la tabla. Calcula las medidas de centralización: media, mediana y moda.

N.º de televisores	0	1	2	3
f_i	2	8	12	3

El número total de datos es $N = 25$.

$\bar{x} = \frac{41}{25} = 1,64$ televisores $Me = 2$ televisores $Mo = 2$ televisores

41. En la siguiente tabla aparece información sobre las edades de los socios infantiles y juveniles de un club de natación.

Edad (años)	[5,8)	[8,11)	[11,14)	[14,17)	[17,20)
f_i	36	23	30	27	24

- a) Halla la media aritmética.
- b) Calcula la mediana.
- c) Obtén la moda.

Las marcas de clase de los intervalos son:

$x_1 = 6,5$ $x_2 = 9,5$ $x_3 = 12,5$ $x_4 = 15,5$ $x_5 = 18,5$

El número total de datos es $N = 140$.

a) $\bar{x} = \frac{1690}{140} = 12,071$ años b) $Me = 12,5$ años c) $Mo = 6,5$ años

42. Dados los siguientes datos encuentra la cantidad tal que sumada o restada a cada dato haga que la media sea 1.

x_i	1	2	3	4	5
f_i	2	1	3	2	2

Calculamos la media de los datos: $\bar{x} = \frac{31}{10} = 3,1$

La cantidad que hay que sumar a los datos para que la media sea 1 será: $1 - 3,1 = -2,1$.

Por tanto, hay que restarle 2,1 a todos los datos para que su media sea 1.

43. En un gimnasio han ingresado este año 160 nuevos socios. Sus edades se distribuyen como podemos ver en la siguiente tabla.

Determina los percentiles 30, 40, 60, 80 y 90.

Edad	18	19	20	21	22	23	24
f_i	22	36	40	24	16	12	10

Completamos la tabla con las frecuencias acumuladas:

Edad	18	19	20	21	22	23	24	Total
f_i	22	36	40	24	16	12	10	160
F_i	22	58	98	122	138	150	160	

30% de 160 = 48 → $P_{30} = 19$

80% de 160 = 128 → $P_{80} = 22$

40% de 160 = 64 → $P_{40} = 20$

90% de 160 = 144 → $P_{90} = 23$

60% de 160 = 96 → $P_{60} = 20$

44. El tutor de 1.º A ha hecho una relación de los alumnos del grupo reunidos por el número de suspensos que han tenido en la primera evaluación.

N.º de suspensos	0	1	2	3	5	Total
f_i	12	9	6	4	1	32

Calcula.

a) La media aritmética.

c) La varianza.

b) La desviación media.

d) La desviación típica.

a) $\bar{x} = \frac{38}{32} = 1,1875$

c) $\sigma^2 = \frac{94}{32} - 1,1875^2 = 1,527$

b) $DM = \frac{31,875}{32} = 0,996$

d) $\sigma = 1,236$

45. Las marcas, en centímetros, de los saltadores de altura de un equipo juvenil son las siguientes.

192 178 186 202 199 181 188 185 179 194

Determina.

a) La media aritmética.

c) La varianza.

b) La desviación media.

d) La desviación típica.

a) $\bar{x} = \frac{1884}{10} = 188,4$

c) $\sigma^2 = \frac{355556}{10} - 188,4^2 = 61,04$

b) $DM = \frac{66,8}{10} = 6,68$

d) $\sigma = 7,8128$

46. Un supermercado está haciendo un estudio de mercado para adecuar su oferta con los gustos de los clientes. Hoy ha controlado el número de artículos diferentes que llevaban en su carro los clientes cuando pasaban por la caja. Han resumido los datos de la forma siguiente.

Artículos	Cientes
[0, 5)	28
[5, 10)	44
[10, 15)	30
[15, 20)	7
[20, 25)	3
[25, 30)	2

- a) Halla el número medio de artículos diferentes que lleva cada carro que sale de las cajas del supermercado.
- b) Calcula también sus medidas de dispersión.
- c) Interpreta los resultados obtenidos.

Las marcas de clase son:

$$x_1 = 2,5 \quad x_2 = 7,5 \quad x_3 = 12,5 \quad x_4 = 17,5 \quad x_5 = 22,5 \quad x_6 = 27,5$$

Y el número total de datos es: $N = 114$.

a) $\bar{x} = \frac{1020}{114} = 8,947$ artículos

b) $DM = \frac{488,421}{114} = 4,284 \quad \sigma^2 = \frac{112512,5}{114} - 8,947^2 = 29,7 \quad \sigma = 5,45 \quad CV = \frac{5,45}{8,947} = 0,61$

- c) Cada cliente lleva de media 8,947 artículos, casi 9. Pero el coeficiente de variación indica que los datos no están demasiado concentrados en la media.

47. El seleccionador nacional de baloncesto femenino está valorando la inclusión en el equipo de una jugadora. Ha estudiado sus últimos diez partidos y ha apuntado el número de minutos jugados y los puntos anotados. Los puedes ver en la tabla.

N.º de partidos	Minutos jugados	Puntos anotados
1	32	26
2	30	15
3	34	23
4	27	11
5	12	2
6	26	20
7	30	21
8	34	13
9	32	27
10	33	18

- a) Calcula la media aritmética y la desviación típica de cada una de las variables estudiadas.
- b) Determina el coeficiente de variación en ambas para estimar cuál de las dos variables es más dispersa.

a) La media y la desviación típica de los minutos jugados son:

$$\bar{x}_M = \frac{290}{10} = 29 \qquad \sigma_M^2 = \frac{8\,798}{10} - 29^2 = 38,8 \rightarrow \sigma_M = 6,229$$

La media y la desviación típica de los puntos anotados son:

$$\bar{x}_P = \frac{176}{10} = 17,6$$

$$\sigma_P^2 = \frac{3\,618}{10} - 17,6^2 = 52,04 \rightarrow \sigma_P = 7,214$$

b) El coeficiente de variación media de los minutos jugados es:

$$CV_M = \frac{6,226}{29} = 0,215.$$

El coeficiente de variación media de los puntos anotados es:

$$CV_P = \frac{7,214}{17,6} = 0,41.$$

$CV_P > CV_M \rightarrow$ Los puntos anotados son más dispersos que los minutos jugados.

48. Una empresa realiza un estudio sobre el salario anual de sus trabajadores, cuyos resultados se recogen en la siguiente tabla.

Salario (en miles de €)	[10, 15)	[15, 20)	[20, 25)	[25, 30)	[30, 35)	[35, 40)
N.º de trabajadores	21	32	85	90	25	5

- a) Calcula las medidas de centralización.
- b) Halla las medidas de dispersión.

Completamos la tabla de frecuencias:

Salario (miles de €)	[10, 15)	[15, 20)	[20, 25)	[25, 30)	[30, 35)	[35, 40)	Total
x_i	12,5	17,5	22,5	27,5	32,5	37,5	
f_i	21	32	85	90	25	5	258
F_i	21	53	138	228	253	258	

a) $Mo = 27,5$ $Me = 27,5$

$$\bar{x} = \frac{6\,210}{258} = 24,07$$

b) $DM = \frac{1173,256}{258} = 4,55$

$$\sigma^2 = \frac{157\,612,5}{258} - 24,07^2 = 31,547$$

$$\sigma = 5,617$$

$$cv = \frac{5,617}{24,07} = 0,233$$

49. Analiza, utilizando las medidas de centralización y las de dispersión, los salarios mensuales de estas dos empresas.

Empresa A		Empresa B	
Salarios	N.º de empleados	Salarios	N.º de empleados
[600, 1000)	15	[600, 1000)	10
[1000, 1400)	20	[1000, 1400)	30
[1400, 1800)	30	[1400, 1800)	35
[1800, 2000)	20	[1800, 2000)	24
[2000, 2400)	15	[2000, 2400)	10

Las marcas de clase son, para los dos estudios, las siguientes:

$$x_1 = 800 \quad x_2 = 1\,200 \quad x_3 = 1\,600 \quad x_4 = 1\,900 \quad x_5 = 2\,200$$

▪ Empresa A:

El número total de empleados es $N_A = 100$.

Las medidas de centralización son:

$$Mo_A = 1\,600 \text{ €} \quad Me_A = 1\,600 \text{ €}$$

$$\bar{X}_A = \frac{155\,000}{100} = 1\,550 \text{ €}$$

Las medidas de dispersión son:

$$DM_A = \frac{36\,500}{100} = 365$$

$$\sigma_A^2 = \frac{260\,000\,000}{100} - 1\,550^2 = 197\,500 \quad \sigma_A = 444,41$$

$$CV_A = \frac{444,41}{1\,550} = 0,287$$

▪ Empresa B:

El número total de empleados es $N_B = 109$.

Las medidas de centralización son:

$$Mo_B = 1\,600 \text{ €} \quad Me_B = 1\,600 \text{ €}$$

$$\bar{X}_B = \frac{167\,600}{109} = 1\,537,61 \text{ €}$$

Las medidas de dispersión son:

$$DM_B = \frac{35\,009,17}{109} = 321,185$$

$$\sigma_B^2 = \frac{274\,240\,000}{109} - 1\,537,61^2 = 151\,704,402 \quad \sigma_B = 389,45$$

$$CV_B = \frac{389,45}{1\,537,61} = 0,253$$

50. María trabaja como operadora telefónica en la empresa de telefonía, LEMON. Le han encargado que realice una encuesta a todos los clientes que atienda. Les hace solo tres preguntas:

X) ¿Cuánto tiempo ha pasado desde que marcó nuestro número de atención al cliente y el momento en que hemos resuelto su problema?

Y) Valoración personal del servicio prestado. MB (muy bien), B (bien), R (regular), M (mal) y MM (muy mal).

Z) Número de años que lleva abonado a la empresa de telefonía LEMON.

La operadora ha resumido los datos de los 40 últimos encuestados en una tabla para pasarlos al departamento de marketing.

X	Y	Z	X	Y	Z
12	R	1	1	B	3
5	B	4	12	MB	3
18	M	3	6	MB	2
3	MB	3	7	M	3
2	MB	2	23	MM	4
7	MM	1	15	MB	7
6	M	3	4	M	5
11	B	4	21	R	4
2	MB	5	7	B	4
22	MM	6	2	B	6
4	B	1	6	B	3
7	B	4	13	B	2
9	R	2	9	R	9
10	R	3	18	M	2
18	R	3	10	R	4
5	B	3	11	R	2
12	MB	2	2	B	4
2	MB	4	5	MB	6
7	MB	4	13	B	7
5	B	5	4	B	5

- Toma cada una de las variables y realiza un recuento (si son muchos valores diferentes, agrúpalos en clases).
- Confecciona con cada una su tabla de frecuencias.
- Realiza en cada caso el gráfico que te parezca adecuado.
- Determina, si es posible, sus medidas de centralización.
- Obtén, en cada una, los cuartiles inferior y superior.
- Halla, si es posible, las medidas de dispersión de cada una de las tres series.
- Decide entre la primera y la tercera variable cuál es más dispersa.

a) Variable X:

Agrupamos los datos en intervalos de clase. El número de intervalos será $\sqrt{40} = 6,3 \approx 7$, y su tamaño será $\frac{23-1}{\sqrt{40}} = 3,5 \approx 5$.

Minutos	[1, 6)	[6, 11)	[11, 16)	[16, 21)	[21, 26)	Total
x_i	3,5	8,5	13,5	18,5	23,5	
f_i	14	12	8	3	3	40

Variable Y:

Valoración	MB	B	R	M	MM	Total
f_i	10	14	8	5	3	40

Variable Z:

Años	1	2	3	4	5	6	7	8	9	Total
f_i	3	7	10	10	4	3	2	0	1	40

b) Variable X:

Minutos	[1, 6)	[6, 11)	[11, 16)	[16, 21)	[21, 26)	Total
x_i	3,5	8,5	13,5	18,5	23,5	
f_i	14	12	8	3	3	40
F_i	14	26	34	37	40	
h_i	0,35	0,3	0,2	0,075	0,075	1
H_i	0,35	0,65	0,85	0,925	1	

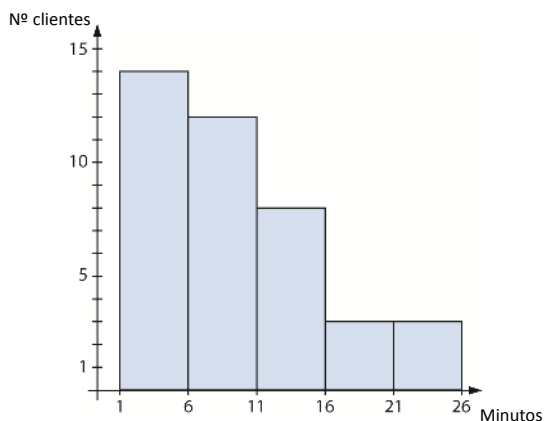
Variable Y:

Valoración	MB	B	R	M	MM	Total
f_i	10	14	8	5	3	40
h_i	0,25	0,35	0,2	0,125	0,075	

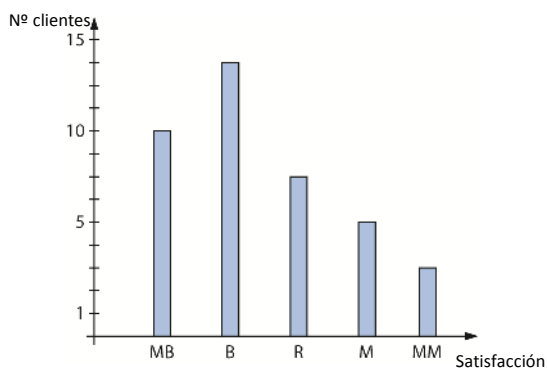
Variable Z:

Años	1	2	3	4	5	6	7	8	9	Total
f_i	3	7	10	10	4	3	2	0	1	40
F_i	3	10	20	30	34	37	39	39	40	
h_i	0,075	0,175	0,25	0,25	0,1	0,075	0,05	0	0,025	1
H_i	0,075	0,25	0,5	0,75	0,85	0,925	0,975	0,975	1	

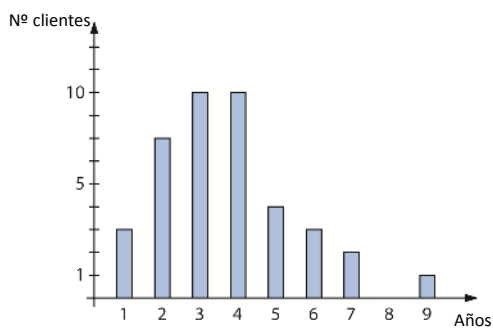
c) Variable X:



Variable Y:



Variable Z:



d) Variable X:

$$Mo_x = 3,5 \text{ minutos} \quad Me_x = 8,5 \text{ minutos} \quad \bar{x}_x = \frac{385}{40} = 9,625 \text{ minutos}$$

Variable Y:

Es una variable cualitativa, por lo que la única medida de centralización que podemos calcular es la moda:

$$Mo_y = B$$

Variable Z:

$$Mo_z = \{3, 4\} \quad Me_z = 3,5 \quad \bar{x}_z = \frac{148}{40} = 3,7$$

e) 25 % de 40 = 10 y 75 % de 40 = 30.

▪ Variable X:

$$Q_{x1} = 3,5 \qquad Q_{x3} = 13,5$$

▪ Variable Y:

Es una variable cualitativa, por lo que no tiene sentido calcular los cuartiles.

▪ Variable Z:

$$Q_{z1} = 2,5 \qquad Q_{z3} = 4,5$$

f) ▪ Variable X:

$$DM_x = \frac{198,5}{40} = 4,96 \qquad \sigma_x^2 = \frac{5\,180}{40} - 9,625^2 = 36,86 \qquad \sigma_x = 6,07 \qquad CV_x = \frac{6,07}{9,625} = 0,63$$

▪ Variable Y:

Es una variable cualitativa, por lo que no se pueden calcular las medidas de dispersión.

▪ Variable Z:

$$DM_z = \frac{54}{40} = 1,35 \qquad \sigma_z^2 = \frac{668}{40} - 3,7^2 = 3,01 \qquad \sigma_z = 1,735 \qquad CV_z = \frac{1,735}{3,7} = 0,469$$

g) $CV_x > CV_z \rightarrow$ La variable X es más dispersa que la Z.

51. El 15% de los científicos de un centro de investigación habla solo en castellano. Un 48% puede hablar en dos idiomas diferentes. Hay un 28% que es capaz de comunicarse en tres lenguas distintas y el resto habla cuatro o más idiomas.

Realiza una tabla de frecuencias con estos datos. ¿Hay algún tipo de frecuencia que responda a la pregunta de cuántos científicos del centro hablan menos de tres lenguas? Razona tu respuesta.

Nº idiomas	1	2	3	4 o +	Total
%	15	48	28	9	100
h_i	0,15	0,48	0,28	0,09	1
H_i	0,15	0,63	0,91	1	

Sí, la frecuencia relativa acumulada $H_2 = 0,63 = 63\%$.

52. Calcula la media aritmética, la desviación típica y el coeficiente de variación de los datos de esta serie.

21 34 28 26 24 25 20 30

- a) Escribe otra serie estadística cuya media sea doble de la anterior y que tenga menor coeficiente de variación.
 b) Escribe otra serie diferente cuya media sea la mitad que la primera y que su coeficiente de variación sea mayor.

$$\bar{x} = \frac{208}{8} = 26 \qquad \sigma^2 = \frac{5\,558}{8} - 26^2 = 18,75 \rightarrow \sigma = 4,33 \qquad CV = \frac{4,33}{26} = 0,167$$

a) Respuesta abierta. Por ejemplo:

Si sumamos la media a todos los datos, la serie obtenida tendrá como media el doble de la inicial y su desviación típica no variará, por tanto, su coeficiente de variación disminuirá:

47 69 54 52 50 51 46 56

$$\bar{x} = \frac{416}{8} = 52 = 2 \cdot 26 \qquad \sigma^2 = \frac{21\,782}{8} - 56^2 = 18,75 \rightarrow \sigma = 4,33 \qquad CV = \frac{4,33}{52} = 0,083$$

b) Respuesta abierta. Por ejemplo:

La media tiene que ser $\frac{26}{2} = 13$. De forma similar a lo realizado en el apartado anterior, si restamos 13 a todos los datos, la serie obtenida tendrá como media la mitad de la inicial y su desviación típica no variará, por lo que su coeficiente de variación aumentará:

$$\begin{array}{cccccccc} & 8 & 21 & 15 & 13 & 11 & 12 & 7 & 17 \\ \bar{x} = \frac{104}{8} = 13 = \frac{26}{2} & & \sigma^2 = \frac{1502}{8} - 13^2 = 18,75 \rightarrow \sigma = 4,33 & & & & CV = \frac{4,33}{13} = 0,333 & & \end{array}$$

53. El seleccionador nacional de baloncesto estudia con mucho cuidado las características de todos sus jugadores. En la tabla siguiente había reflejado el número de faltas personales cometidas por el jugador X en sus últimos 50 partidos.

Faltas personales	5	4	3	2	1	0	Total
Partidos	4	6		14		1	50

Como ves, se han perdido un par de datos. ¿Podrías recuperarlos sabiendo que el número medio de faltas personales del jugador X era de 2,62?

Sean x e y, respectivamente, el primer y el segundo dato desconocido.

Sabemos que la suma del total de partidos es 50, es decir:

$$4 + 6 + x + 14 + y + 1 = 50 \rightarrow x + y = 25$$

También sabemos que la media de faltas personales es 2,62, es decir:

$$\bar{x} = \frac{5 \cdot 4 + 4 \cdot 6 + 3 \cdot x + 2 \cdot 14 + 1 \cdot y + 0 \cdot 1}{50} = 2,62 \rightarrow 3x + y = 59$$

Con las ecuaciones obtenidas planteamos un sistema de ecuaciones que resolvemos:

$$\left. \begin{array}{l} x + y = 25 \\ 3x + y = 59 \end{array} \right\} \rightarrow x = 17, y = 8$$

El jugador cometió 3 faltas personales en 17 partidos y 1 falta en 1 partido.

54. Las notas obtenidas por Daniel en Matemáticas durante este trimestre son 4, 7, 9, 7 y 5.

- a) Calcula la media aritmética, la moda y la mediana de esta serie.
- b) Aún le falta realizar otro examen. ¿Qué nota debería obtener en él para que la nota media resultara ser 6,5?
- c) ¿Qué condiciones tendría que cumplir esa nota que falta para que la nueva serie tuviera la misma mediana?
- d) ¿Qué condiciones tendría que cumplir esa nota que falta para que la nueva serie tuviera la misma moda?

a) $\bar{x} = \frac{32}{5} = 6,4$ $Mo = 7$ $Me = 7$

b) Sea x la nota que tendría que obtener. La nueva media sería:

$$\bar{x} = \frac{32 + x}{6} = 6,5 \rightarrow x = 7 \text{ es la nota que debería obtener Daniel en el próximo examen.}$$

- c) Ser mayor o igual que la mediana inicial.
- d) Ser el valor de la moda o un valor distinto a las otras notas obtenidas.

55. Un entrenador de atletismo ha formado dos grupos de corredores de 100 metros lisos, uno femenino y otro masculino. La media de las marcas del equipo masculino es de 12,42 segundos, con una desviación típica de 1,2 segundos. La media del equipo femenino es de 13,04 segundos, con una desviación típica de 1,05. Al club de atletismo se incorpora una pareja, él tiene una marca de 13,03 segundos y ella de 13,24 segundos. Si solo puede elegir uno de los atletas, ¿cuál de los dos tiene mejor marca en su categoría?

Analizamos cómo de alejadas de las medias marcas están las marcas con respecto a la desviación típica:

$$\text{El atleta: } \frac{|13,03 - \bar{x}|}{\sigma} = \frac{|13,03 - 12,42|}{1,2} = 0,51$$

$$\text{La atleta: } \frac{|13,24 - \bar{x}|}{\sigma} = \frac{|13,24 - 13,04|}{1,05} = 0,19$$

El valor es menor en el caso de la chica, por lo que es, de los dos, quien mejor marca tiene en su categoría.

56. Germán se está preparando para participar en su primera media maratón. Cada día sale a entrenar y anota la distancia recorrida. Hoy, que es viernes, acaba de terminar su entrenamiento y está pensando en lo que debería correr durante el fin de semana. Sus anotaciones han sido las que se muestran en la tabla.

Día de la semana	L	M	X	J	V	S	D
Distancia	12	18	15	20	19		

- ¿Cuáles son la media aritmética y la desviación típica de la serie semanal?
- Añade dos posibles distancias que recorrer el sábado y el domingo de modo que se mantenga la misma media aritmética pero que disminuya su desviación típica.
- ¿Qué dos distancias podría escribir en el sábado y el domingo de modo que, manteniéndose la media aritmética, aumentara su desviación típica?
- ¿Sería posible añadir dos distancias en sábado y domingo y que aumente la media aritmética pero no la desviación típica?
- ¿Podrías hacer lo mismo añadiendo otras dos y que disminuyera la media aritmética sin aumentar la desviación típica?

$$\text{a) } \bar{x} = \frac{84}{5} = 16,8 \text{ km} \quad \sigma^2 = \frac{1454}{5} - 16,8^2 = 8,56 \rightarrow \sigma = 2,926 \text{ km}$$

- b) Para mantener la media, los nuevos datos, x e y , deben cumplir:

$$\bar{x} = \frac{84 + x + y}{7} = 16,8 \rightarrow x + y = 33,6 \rightarrow y = 33,6 - x$$

Si queremos disminuir la desviación típica manteniendo la media, solo tenemos que tomar dos valores muy próximos a la media como, por ejemplo, la propia media:

$$x = 16,8 \text{ km} \rightarrow y = 16,8 \text{ km} \rightarrow \sigma^2 = \frac{2018,48}{7} - 16,8^2 = 6,114 \rightarrow \sigma = 2,473 \text{ km}$$

- c) Partiendo de la misma idea del apartado b), tomamos dos valores muy alejados de la media. Por ejemplo:

$$x = 1 \text{ km} \rightarrow y = 32,6 \text{ km} \rightarrow \sigma^2 = \frac{2517,76}{7} - 16,8^2 = 77,44 \rightarrow \sigma = 8,8 \text{ km}$$

- d) Sí, bastaría con añadir dos valores que estén por encima de la media y a una distancia menor que la desviación típica. Por ejemplo, 17 y 18:

$$\bar{x} = \frac{119}{7} = 17 \text{ km} \quad \sigma^2 = \frac{2067}{7} - 17^2 = 6,286 \rightarrow \sigma = 2,507 \text{ km}$$

- e) Sí, bastaría con añadir dos valores que estén por debajo de la media y a una distancia menor que la desviación típica. Por ejemplo, 15 y 14:

$$\bar{x} = \frac{113}{7} = 16,14 \text{ km} \quad \sigma^2 = \frac{1875}{7} - 16,14^2 = 7,265 \rightarrow \sigma = 2,7 \text{ km}$$

57. Un grupo de personas está participando en el ensayo clínico de un medicamento. Por ello se está vigilando su nivel de colesterol en sangre. Los resultados del último análisis han sido los que se recogen en la tabla.

Nivel de colesterol	f_i
[160, 180)	6
[180, 200)	14
[200, 220)	32
[220, 240)	40
[240, 260)	28

- a) Calcula la media aritmética y la desviación típica correspondiente a estos datos.
 b) El equipo médico desea seleccionar un intervalo de niveles de colesterol que contenga a un 95% de las personas y esté centrado en la media, o sea, que sea del tipo $(\bar{x} - 2\sigma, \bar{x} + 2\sigma)$. ¿Cuáles serán los extremos de dicho intervalo?

a) Las marcas de clase son:

$$x_1 = 170 \qquad x_2 = 190 \qquad x_3 = 210 \qquad x_4 = 230 \qquad x_5 = 250$$

El número total de datos es $N = 120$.

$$\bar{x} = \frac{26600}{120} = 221,6 \qquad \sigma^2 = \frac{5956000}{120} - 221,6^2 = 497,2 \rightarrow \sigma = 22,3$$

b) El intervalo sería:

$$(221,6667 - 2 \cdot 22,3; 221,6667 + 2 \cdot 22,3) = (177,0667; 266,2667)$$

58. Andrés, Beatriz, Carmen y Daniel son cuatro amigos jubilados. Los cuatro trabajaban juntos como estadísticos en el Instituto Nacional de Estadística. Les gusta juntarse y se divierten proponiendo problemas.

–Como sabéis, curiosamente todos tenemos el mismo número de nietos. Hoy le he dado una propina a cada uno de los míos –comienza Daniel–. Como tienen edades muy distintas, le he dado una cantidad diferente a cada uno.

–Sin saber lo que les has dado, pero sabiendo que no eres muy generoso, yo daré cuatro euros más que tú a cada uno de mis nietos –dice Carmen.

–Pues yo a cada uno de mis nietos le voy a dar la cuarta parte de lo que les ha dado Daniel a los suyos, que yo sé que no es un tacaño –afirma Beatriz un poco picada con su amiga.

–Como yo también creo que Daniel es un poco roñoso, voy a darles cuatro veces más que él a cada uno de mis nietos –responde Andrés, que es el más chulo de todos.

–Está bien. Haremos cálculos y la estadística pondrá a cada uno en su lugar –afirma Daniel un poco mosqueado–. Me pregunto que si llamo «m» a la media de las propinas que yo he dado y «s» a la desviación típica, ¿cuáles serán las medias aritméticas y las desviaciones típicas de las cantidades que vais a dar cada uno de vosotros?

¿Puedes poner un poco de paz entre los amigos realizando tú los cálculos?

Pista. Pon una serie ejemplo para Daniel y realiza los cálculos. Luego construye las series de los demás y vuelve a efectuar los cálculos. Haz una hipótesis general y trata de encontrar una demostración para tu hipótesis.

La media, m , y la desviación típica, s , de las propinas de Daniel son:

$$m = \frac{\sum_{i=1}^n f_i \cdot X_i}{N} \qquad s = \sqrt{\frac{\sum_{i=1}^n f_i \cdot X_i^2}{N} - m^2}$$

La media y la desviación típica de las propinas de Carmen serán:

$$\bar{X}_{\text{Carmen}} = \frac{\sum_{i=1}^n f_i \cdot (X_i + 4)}{N} = \frac{\sum_{i=1}^n f_i \cdot X_i}{N} + \frac{\sum_{i=1}^n f_i \cdot 4}{N} = m + 4 \cdot \frac{N}{N} = m + 4$$

$$\begin{aligned} \sigma_{\text{Carmen}}^2 &= \frac{\sum_{i=1}^n f_i \cdot (X_i + 4)^2}{N} - (m + 4)^2 = \frac{\sum_{i=1}^n f_i \cdot (X_i^2 + 4X_i + 16)}{N} - (m + 4)^2 = \frac{\sum_{i=1}^n f_i \cdot (X_i^2 + 4X_i + 16)}{N} - (m + 4)^2 = \\ &= \frac{\sum_{i=1}^n f_i \cdot X_i^2 + \sum_{i=1}^n f_i \cdot 8X_i + \sum_{i=1}^n f_i \cdot 16}{N} - (m + 4)^2 = \frac{\sum_{i=1}^n f_i \cdot X_i^2}{N} + 8m + 16 - (m^2 + 8m + 16) = \frac{\sum_{i=1}^n f_i \cdot X_i^2}{N} - m^2 = s^2 \end{aligned}$$

$$\sigma_{\text{Carmen}} = s$$

La media y la desviación típica de las propinas de Beatriz serán:

$$\bar{X}_{\text{Beatriz}} = \frac{\sum_{i=1}^n f_i \cdot \left(\frac{X_i}{4}\right)}{N} = \frac{1}{4} \cdot \frac{\sum_{i=1}^n f_i \cdot X_i}{N} = \frac{m}{4}$$

$$\sigma_{\text{Beatriz}}^2 = \frac{\sum_{i=1}^n f_i \cdot \left(\frac{X_i}{4}\right)^2}{N} - \left(\frac{m}{4}\right)^2 = \left(\frac{1}{4}\right)^2 \cdot \left(\frac{\sum_{i=1}^n f_i \cdot X_i^2}{N} - m^2\right) = \left(\frac{1}{4}\right)^2 \cdot s^2 = \left(\frac{s}{4}\right)^2 \rightarrow \sigma_{\text{Beatriz}} = \frac{s}{4}$$

La media y la desviación típica de las propinas de Andrés se calculan de forma similar a las de Beatriz:

$$\bar{X}_{\text{Andrés}} = \frac{\sum_{i=1}^n f_i \cdot 4X_i}{N} = 4m \qquad \sigma_{\text{Andrés}}^2 = \frac{\sum_{i=1}^n f_i \cdot 4X_i^2}{N} - 4m^2 = (4s)^2 \rightarrow \sigma_{\text{Andrés}} = 4s$$

59. El tutor de un curso está comparando las notas de dos alumnas, María y Esther. Míralas en la tabla siguiente:

Notas en Matemáticas							Notas en Lengua Española						
María	3	8	9	3	8	8	María	4	6	6	5	4	5
Esther	5	7	8	6	7	6	Esther	7	8	9	7	9	8

- a) Calcula las medias aritméticas y las desviaciones típicas de las dos materias para cada una de las chicas. Explica lo que sucede.
 b) ¿Quién tiene las notas más dispersas? ¿Qué medida has calculado para respaldar tus afirmaciones?

a) ▪ **María:**

$$\text{Matemáticas} \rightarrow \bar{X}_{MMat} = 6,5 \qquad \sigma_{MMat} = 2,5$$

$$\text{Lengua} \rightarrow \bar{X}_{MLen} = 5 \qquad \sigma_{MLen} = 0,82$$

▪ **Esther:**

$$\text{Matemáticas} \rightarrow \bar{X}_{EMat} = 6,8\bar{3} \qquad \sigma_{EMat} = 1,07$$

$$\text{Lengua} \rightarrow \bar{X}_{ELen} = 8 \qquad \sigma_{ELen} = 0,82$$

b) Calculamos los coeficientes de variación para analizar la dispersión de las notas:

▪ María:

Matemáticas $\rightarrow CV_{MMat} = 0,38$

Lengua $\rightarrow CV_{MLen} = 0,16$

▪ Esther:

Matemáticas $\rightarrow CV_{EMat} = 0,16$

Lengua $\rightarrow CV_{ELen} = 0,1$

Los coeficientes de variación mayores son los de María, lo que indica que sus notas son más dispersas.

60. A los participantes en un curso de buceo se les ha medido el tiempo que resisten sin respirar. El resumen de los resultados se presenta en la tabla.

Tiempo (s)	N.º de buceadores
[40, 52)	10
[52, 56)	28
[56, 64)	25
[64, 76)	44
[76, 88)	11
[88, 100)	7

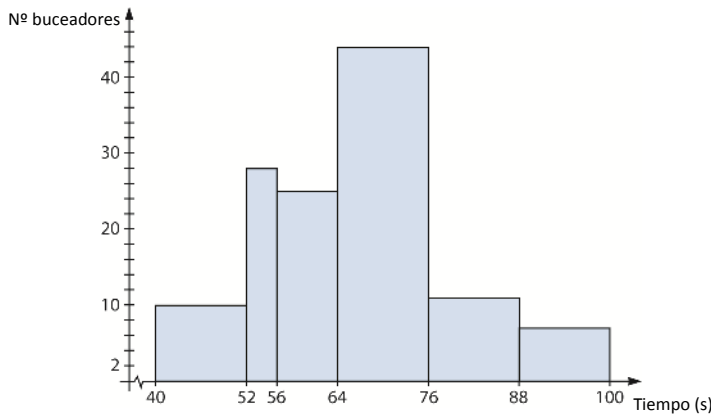


- Completa la tabla de frecuencias añadiendo columnas para la frecuencia acumulada, la frecuencia relativa y el porcentaje.
- Representa los datos empleando el gráfico estadístico que te parezca más adecuado a la naturaleza de los datos.
- Halla las medias de centralización.
- Determina también las medidas de dispersión de la serie.
- En la escuela de buceo afirman que el 80% de las personas aguantan más de un minuto sin respirar. En este grupo, ¿qué porcentaje cumple lo dicho?
- También afirman que un 5% se mantienen durante más de un minuto y medio sin respirar. ¿Es eso cierto en este grupo?

a)

Tiempo (s)	x_i	f_i	F_i	h_i	H_i	%
[40, 52)	46	10	10	0,08	0,08	8
[52, 56)	54	28	38	0,224	0,304	22,4
[56, 64)	60	25	63	0,2	0,504	20
[64, 76)	70	44	107	0,352	0,856	35,2
[76, 88)	82	11	118	0,088	0,944	8,8
[88, 100)	94	7	125	0,056	1	5,6
Total		125		1		100

b)



c) $Mo = 70$ $Me = 60$ $\bar{x} = \frac{8\,112}{125} = 64,896$

d) $DM = \frac{1232,896}{125} = 9,863$ $\sigma^2 = \frac{544\,224}{125} - 64,896^2 = 142,3^2$ $\sigma = 11,929$ $CV = \frac{11,929}{64,896} = 0,184$

e) Como mínimo, aguantan más de un minuto sin respirar $35,2 + 8,8 + 5,6 = 49,6\%$.

Podrían ser más, pero uno de los intervalos es [56, 64) y no podemos saber cuánta personas de este intervalo superaron el minuto.

f) Hay un 5,6% de alumnos que aguantan 88 s; los datos no nos permiten ser más precisos.

61. A un laboratorio han llegado 24 botellas de agua, 12 botellas de 1 litro y 12 botellas de medio litro, para analizar su contenido en sales.

Se han obtenido los siguientes datos, expresados en mg.

■ Botellas de 1 litro:

46 25 27 30 48 40
27 44 37 62 56 29

■ Botellas de medio litro:

76 75 49 59 33 52
54 45 66 69 34 53

a) Clasifica la variable estadística de concentración de sales.

b) ¿Es conveniente tomar intervalos al realizar una tabla?

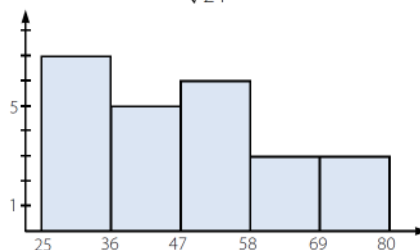
a) La variable es cuantitativa continua.

b) Se pueden agrupar los datos en intervalos para realizar la tabla y facilitar su estudio.

$$\sqrt{24} = 4,9 \rightarrow 5 \text{ intervalos}$$

$$\frac{76 - 25}{\sqrt{24}} = 10,41$$

Contenido (mg)	f_i
[25, 36)	7
[36, 47)	5
[47, 58)	6
[58, 69)	3
[69, 80)	3



PARA PROFUNDIZAR

62. Elige la respuesta adecuada. (Concurso de Primavera)

La media de las edades de todos los miembros de una familia compuesta por padre, madre y varios hijos es 20 años. Si el padre tiene 48 años y la media de las edades de la madre y de todos los hijos es 16, ¿cuántos hijos tienen?	2	3	4	5	6
La suma de 49 números enteros consecutivos es 7^5 . ¿Cuál es su mediana?	7	7^2	7^3	7^4	7^5
En una prueba deportiva de mi instituto, estudiantes de 3.º ESO, 4.º ESO y 1.º Bachillerato han obtenido una puntuación media de 12, 15 y 10, respectivamente. Si participaron el doble de estudiantes de 3.º que de 4.º y el doble de 4.º que de 1.º Bachillerato, ¿cuál es la puntuación media de los participantes?	12	$\frac{37}{3}$	$\frac{88}{7}$	13	145
El peso medio de las patatas que había en una bolsa subió al doble cuando a las cuatro patatas se le añadió una patata inmensa. ¿Cuál es el cociente entre el peso de ese patatón y la suma de los pesos de las cuatro patatas que había?	$\frac{3}{2}$	6	$\frac{8}{3}$	2	1

- Lo primero que se piensa es que la suma de las edades de toda la familia es múltiplo de 20, debido a que su media es 20. Si a es el número de miembros de la familia podemos escribir:

$$48 + \text{edad madre} + \text{edad hijos} = 20a$$

Si no se suma la edad del padre, la suma de las edades de los demás se podrá escribir como:

$$\text{Edad madre} + \text{Edad hijos} = 16(a - 1)$$

Con lo que se tienen que cumplir ambas condiciones, esto es:

$$20 \cdot a = 16(a - 1) + 48 \rightarrow 4a = 32 \rightarrow a = 8$$

Si en total son 8 miembros en la familia, tendrán 6 hijos.

- En una sucesión cualquiera de números enteros consecutivos e impar, la media coincidirá con la mediana, puesto que la mitad de los datos estarán en el punto medio. Por tanto:

$$\bar{x} = \frac{7^5}{49} = \frac{7^5}{7^2} = 7^3$$

- Si llamamos x al número de alumnos de 1.º de Bachillerato, tenemos:

$$2x \rightarrow \text{alumnos de 4.º ESO} \quad 4x \rightarrow \text{alumnos de 3.º ESO}$$

$$\text{La suma de las notas entre el número de alumnos será: } \frac{12(4x) + 15(2x) + 10x}{x + 2x + 4x} = \frac{88x}{7x} = \frac{88}{7}$$

$$\text{Con lo que la puntuación media de los participantes será } \frac{88}{7}$$

- Si llamamos x a la suma del peso de las cuatro patatas, el peso medio por patata será $\frac{x}{4}$. Cuando se añade el patatón, el peso medio pasa a ser $\frac{x}{2}$, con lo que si se renombra su peso como y , queda:

$$\frac{x + y}{5} = \frac{x}{2} \rightarrow y = \frac{5}{2}x - x \rightarrow y = \frac{3}{2}x \rightarrow \frac{\frac{3}{2}x}{\frac{x}{4}} = \frac{12}{2} = 6$$

63. Tenemos 120 datos que hemos clasificado en tres grupos y queremos representarlos mediante un diagrama de sectores. Analizando los distintos grupos hemos llegado a estas conclusiones.

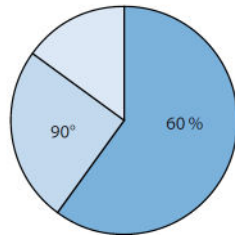
- Sector 1: representa el primer conjunto de datos, y comprende el 60% del diagrama.
- Sector 2: compuesto por el segundo grupo de datos, está representado por un ángulo de 90°.
- Sector 3: representa el tercer grupo de datos.

Construye el diagrama y calcula el número de datos que contiene cada sector.

Si el 60 % corresponde al primer conjunto de datos, como en total son 120 datos, en este grupo se encuentran: $0,6 \cdot 120 = 72$ datos.

Como el segundo sector es de 90° corresponde a una frecuencia relativa de 0,25; por tanto, en el segundo grupo hay: $0,25 \cdot 120 = 30$ datos.

Así, el tercer conjunto de datos está formado por: $120 - 72 - 30 = 18$ datos.



64. Tenemos una variable estadística cuya media aritmética es m y su desviación típica es d . Investiga qué sucede con ambos parámetros si:

- a) Sumamos 4 a todos los números.
- b) Restamos 4 a todos los números.
- c) Multiplicamos por 4 todos los números.
- d) Dividimos entre 4 todos los números.

(Pon una serie de ejemplos y realiza los cálculos. Haz una hipótesis general y trata de encontrar la demostración a tu hipótesis).

a) Si sumamos 4 a todos los valores, la media es $m + 4$ y la desviación típica es d .

Por ejemplo, se considera la serie: 5 6 8 9

$$\bar{x} = \frac{28}{4} = 7 \qquad \sigma = \sqrt{\frac{206}{4} - 7^2} = \sqrt{2,5} = 1,58$$

Al sumar 4 a todos los números: 9 10 12 13

$$\bar{x} = \frac{44}{4} = 11 \qquad \sigma = \sqrt{\frac{494}{4} - 11^2} = \sqrt{2,5} = 1,58$$

b) Si restamos 4 a todos los números, la media es $m - 4$ y la desviación típica es d .

Teniendo en cuenta la serie anterior, al restar 4 a todos los valores: 1 2 4 5

$$\bar{x} = \frac{12}{4} = 3 \qquad \sigma = \sqrt{\frac{46}{4} - 3^2} = \sqrt{2,5} = 1,58$$

c) Si multiplicamos por 4 todos los valores, la media es $4m$ y la desviación típica es $4d$.

La nueva serie es: 20 24 32 36

$$\bar{x} = \frac{112}{4} = 28 \qquad \sigma = \sqrt{\frac{3296}{4} - 28^2} = \sqrt{40} = 6,32$$

d) Si dividimos entre 4 todos los números, la media es $\frac{1}{4}m$ y la desviación típica es $\frac{1}{4}d$.

La nueva serie es: 1,25 1,5 2 2,25

$$\bar{x} = \frac{7}{4} = 1,75 \qquad \sigma = \sqrt{\frac{12,875}{4} - 1,75^2} = \sqrt{0,16} = 0,39$$

65. Se ha hecho un estudio del profesorado de Bachillerato a nivel nacional. Este estudio indica que entre los docentes menores de 40 años hay más mujeres que hombres; en concreto, están en la relación 11 a 10.

Es decir, por cada 11 mujeres que ejercen la docencia en esta etapa educativa, hay 10 hombres que también desarrollan su función educadora en estos niveles.

Si la edad media de las profesoras es de 34 años y la de los profesores es de 32 años, ¿cuál es la media de edad de los docentes menores de 40 años en Bachillerato?

$$\bar{x} = \frac{34 \cdot 11n + 32 \cdot 10n}{11n + 10n} = \frac{694n}{21n} = 33,05 \text{ años}$$

66. Las puntuaciones medias en un concurso de los chicos y las chicas por separado, y los chicos y chicas conjuntamente, de dos centros A y B, sobre una puntuación máxima de 150 puntos, son las que se indican en la siguiente tabla.

	A	B	A y B
Chicos	71	81	79
Chicas	76	90	
Chicos y chicas	74	84	

¿Cuál fue la media de las chicas de los dos centros a la vez?

Si x es el número de chicos del centro A e y es el número de chicas:

$$\frac{71x + 76y}{x + y} = 74 \rightarrow 3x - 2y = 0$$

Análogamente, si m es el número de chicos del centro B y n el de chicas:

$$\frac{81m + 90n}{m + n} = 84 \rightarrow 3m - 6n = 0$$

Y como x es el número de chicos del centro A y m es el número de chicos del centro B:

$$\frac{71x + 81m}{x + m} = 79 \rightarrow 8x - 2m = 0$$

$$\left. \begin{array}{l} 3x - 2y = 0 \\ m - 2n = 0 \\ 4x - m = 0 \end{array} \right\} \rightarrow x = \frac{2}{3}y \rightarrow m = \frac{8}{3}y \rightarrow n = \frac{4}{3}y$$

Así, la media de las chicas de los dos centros es:

$$\frac{76y + 90 \cdot \frac{4}{3}y}{y + \frac{4}{3}y} = \frac{228y + 360y}{3y + 4y} = \frac{588y}{7y} = 84$$

67. La media de un conjunto de 12 datos es 6, y la media de otro conjunto con 13 datos es 5,5. ¿Cuál sería la media si uniéramos todos los datos en un único conjunto de 25 datos?

$$\bar{x} = \frac{6 \cdot 12 + 5,5 \cdot 13}{25} = 5,74$$

MATEMÁTICAS EN TU VIDA

1. ¿Cuál fue el principal objetivo de la encuesta?

El principal objetivo de la encuesta es caracterizar las prácticas y los escenarios de consumo de bienes y servicios culturales de la población de un país.

2. Describe la muestra que se tomó para el estudio.

La muestra estaba formada por 8 275 personas mayores de 4 años tomadas en 2 415 hogares de 14 municipios del país.

3. A la vista de los resultados, ¿crees que una empresa que decidiera vender libros debería tener una orientación hacia el público femenino?

Sí, se puede deducir que las mujeres de la encuesta leen más libros que los hombres.

4. Clasifica las variables estadísticas que aparecen en el texto.

«Número de libros que leen al año la población del país.»

«Número de libros que leen al año las mujeres del país.»

«Número de libros que leen al año los hombres del país.»

5. Elabora una tabla de frecuencias utilizando los datos que están representados en el diagrama que aparece en el texto.

Nº libros	1	2-4	5-7	8-10	11 o +	Total
Hombres (%)	1,9	2,28	5,16	23,69	13,37	46,4
Mujeres (%)	2,5	2,86	5,97	24,87	17,40	53,6
Total (%)	4,40	5,14	11,13	48,56	30,77	100

6. Halla las frecuencias para hombres solo y para mujeres solo.

Nº libros	1	2-4	5-7	8-10	11 o +	Total
Hombres (%)	4,09	4,91	11,12	51,06	28,82	100

Nº libros	1	2-4	5-7	8-10	11 o +	Total
Mujeres (%)	4,66	5,34	11,14	46,4	32,46	100

7. Determina la media de la cantidad de libros leídos por las personas encuestadas.

Aunque sea una variable discreta, los datos están dados en intervalos, por lo que tenemos que calcular las marcas de clase:

$$x_1 = 1 \quad x_2 = 3 \quad x_3 = 6 \quad x_4 = 9 \quad x_5 = 11$$

En el último intervalo consideramos el valor menor, dado que es el único valor seguro:

$$\bar{X} = \frac{1 \cdot 4,4 + 3 \cdot 5,14 + 6 \cdot 11,13 + 9 \cdot 48,56 + 11 \cdot 30,77}{100} = 8,62 \text{ libros}$$