

EJERCICIOS

001 Señala en qué caso es más conveniente estudiar la población o una muestra.

- a) La longitud de los tornillos que fabrica una máquina de manera continua.
b) El peso de un grupo de cinco amigos.
- a) Una muestra, pues la población es excesivamente grande.
b) La población, ya que es pequeña.

002 Determina las variables estadísticas que se estudian en el ejercicio anterior.

En el caso de los tornillos, la variable estadística es la longitud, y en el caso de los amigos, es el peso. En ambos casos son variables cuantitativas continuas.

003 Este es el titular de un periódico:

«EL PESO MEDIO DE LOS ESPAÑOLES ES 69 KG».

- a) ¿Cómo se ha llegado a esta conclusión? ¿Se habrá estudiado a toda la población?
- b) ¿Qué características debería tener la muestra elegida? ¿Podrían ser todos los individuos de la muestra de la misma edad? Y si todos fueran mujeres, ¿sería representativa la muestra?
- a) Se ha llegado a la conclusión a partir del estudio de una muestra.
b) La muestra debe ser representativa de los sectores que forman la población española, respetando la proporción de cada uno de ellos. Si todos los individuos fueran de la misma edad o el mismo sexo no sería representativa, ya que no se mantendría la proporción de los sectores en la población.

004 Las notas en Inglés de 20 alumnos son:

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

Obtén la tabla de frecuencias.

Notas	f_i	h_i	F_i	H_i
1	1	0,05	1	0,05
2	1	0,05	2	0,1
3	1	0,05	3	0,15
4	1	0,05	4	0,2
5	5	0,25	9	0,45
6	3	0,15	12	0,6
7	3	0,15	15	0,75
8	2	0,1	17	0,85
9	2	0,1	19	0,95
10	1	0,05	20	1

005 El número de horas diarias de estudio de 30 alumnos es:

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

Construye una tabla de frecuencias.

Horas	f_i	h_i	F_i	H_i
0	3	0,1	3	0,1
1	8	0,27	11	0,37
2	7	0,23	18	0,6
3	6	0,2	24	0,8
4	3	0,1	27	0,9
5	3	0,1	30	1

006 El color de pelo de 30 personas elegidas al azar es:

$M =$ moreno $M R P M M$
 $R =$ rubio $M M R R P$
 $P =$ pelirrojo $P M M M M$
 $M P R R R$
 $P M M M R$
 $M M M R P$



Construye su tabla de frecuencias.

Color	f_i	h_i
Moreno	16	0,53
Rubio	8	0,27
Pelirrojo	6	0,2

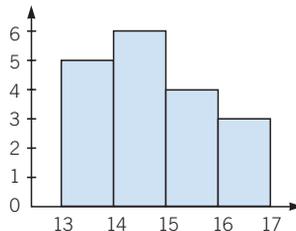
007 Estas son las edades, en años, de 18 alumnos:

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

Realiza un gráfico de sus frecuencias relativas.

Antes de dibujar el gráfico es necesario hacer la tabla de frecuencias. Aunque los valores parecen discretos, al representarlos consideramos que alguien que afirma tener 14 años, realmente tiene más de 14 años y menos de 15 años. Por esta razón utilizamos un histograma.

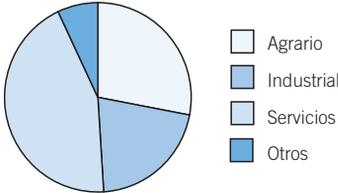
x_i	f_i	h_i
13	5	0,278
14	6	0,333
15	4	0,222
16	3	0,167



Estadística

008 Representa estos datos con el gráfico adecuado.

Sector	Agrario	Industrial	Servicios	Otros
Trabajadores	28%	21%	44%	7%

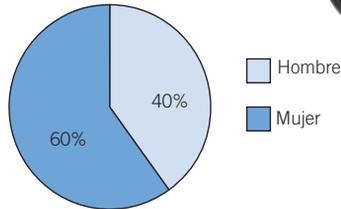


009 El sexo de 20 bebés nacidos en un hospital ha sido:

H M H H M M H H M M
M M M H M M H H M M

Construye la tabla asociada a estos datos, y represéntalos.

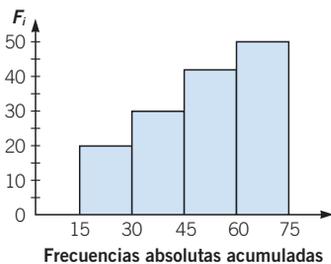
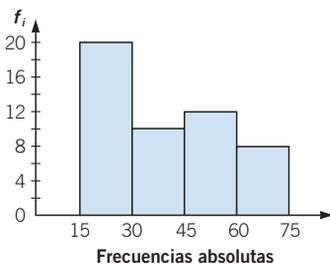
Sexo	f_i	h_i
Hombre	8	0,4
Mujer	12	0,6



010 Completa la tabla de frecuencias y dibuja el histograma de frecuencias absolutas y acumuladas con los datos de esta tabla.

Edad	[15, 30)	[30, 45)	[45, 60)	[60, 75)
N.º de personas	20	10	12	8

Edad	f_i	h_i	F_i	H_i
[15, 30)	20	0,4	20	0,4
[30, 45)	10	0,2	30	0,6
[45, 60)	12	0,24	42	0,84
[60, 75)	8	0,16	50	1

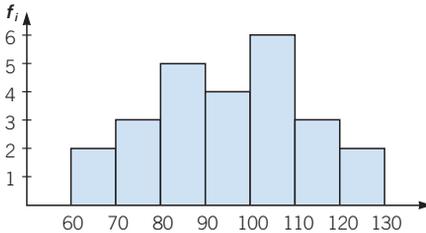


011 Los resultados de un test de inteligencia realizado a 25 personas fueron:

100 80 92 101 65 72 121 68 75 93 101 100
 102 97 89 73 121 114 113 113 106 84 94 83 82

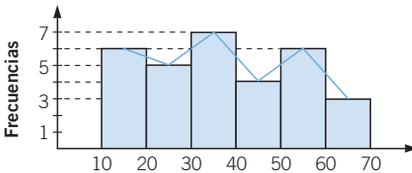
Obtén la tabla de frecuencias y de porcentajes, tomando intervalos de amplitud 10. Representa los datos en un histograma.

Resultados	f_i	%	h_i	F_i	H_i
[60, 70)	2	8	0,08	2	0,08
[70, 80)	3	12	0,12	5	0,2
[80, 90)	5	20	0,2	10	0,4
[90, 100)	4	16	0,16	14	0,56
[100, 110)	6	24	0,24	20	0,8
[110, 120)	3	12	0,12	23	0,92
[120, 130)	2	8	0,08	25	1



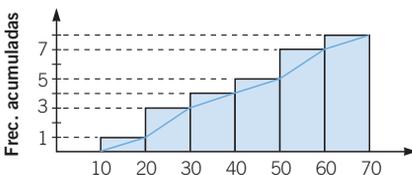
012 Construye las tablas de frecuencias que corresponden a los siguientes gráficos estadísticos, indicando de qué tipo es cada uno.

a) En este histograma están representados las frecuencias absolutas y el polígono de frecuencias. La tabla de frecuencias correspondiente es:



Intervalos	f_i	h_i
[10, 20)	6	0,19
[20, 30)	5	0,16
[30, 40)	7	0,23
[40, 50)	4	0,13
[50, 60)	6	0,19
[60, 70)	3	0,1

b) En este histograma están representados las frecuencias acumuladas y el polígono de frecuencias. La tabla de frecuencias correspondiente es:



Intervalos	F_i	H_i
[10, 20)	1	0,125
[20, 30)	3	0,375
[30, 40)	4	0,5
[40, 50)	5	0,625
[50, 60)	7	0,875
[60, 70)	8	1

Estadística

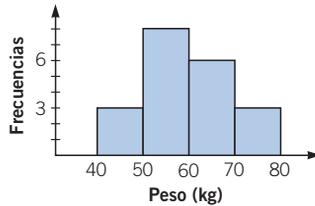
013 Organiza en una tabla de frecuencias estos datos relativos al peso, en kg, de 20 personas.

42 51 56 66 75 47 51 45 63 79
69 59 50 70 59 62 54 60 63 58

- a) Representa los datos mediante el gráfico más adecuado.
b) Calcula sus medidas de centralización.

a)

Peso	x_i	f_i
[40, 50)	45	3
[50, 60)	55	8
[60, 70)	65	6
[70, 80)	75	3



b) La media aritmética es:

$$\bar{x} = \frac{3 \cdot 45 + 8 \cdot 55 + 6 \cdot 65 + 3 \cdot 75}{20} = 59,5$$

La frecuencia mayor es 8, que corresponde al intervalo [50, 60).

Intervalo modal = [50, 60)

Ordenamos los datos: 42, 45, 47, 50, 51, 51, 54, 56, 58, 59, 59, 60, 62, 63, 63, 66, 69, 70, 75, 79

$$Me = \frac{59 + 59}{2} = 59$$

014 Halla las medidas de centralización.

[0, 5)	[5, 10)	[10, 15)	[15, 20)
10	12	37	21

La media aritmética es:

$$\bar{x} = \frac{10 \cdot 2,5 + 12 \cdot 7,5 + 37 \cdot 12,5 + 21 \cdot 17,5}{80} = 11,8$$

La frecuencia mayor es 37, que corresponde al intervalo [10, 15).

Intervalo modal = [10, 15)

Intervalo mediano = [10, 15)

015 Decide qué valores podemos añadir a este conjunto de datos: 18, 8, 7, 9, 12, 15, 21 y 12 para que la mediana siga siendo la misma.

Ordenamos los datos: 7, 8, 9, 12, 12, 15, 18, 21

La mediana es 12.

Si solo añadimos un valor, independientemente de cuál sea, la mediana seguirá siendo 12.

016 Con los datos de la tabla del ejemplo anterior, calcula los siguientes percentiles.

- a) P_{22} b) P_7 c) P_{98} d) P_{66}

Datos	f_i	h_i	F_i	H_i
1	11	0,18	11	0,18
2	27	0,45	38	0,63
3	4	0,07	42	0,7
4	18	0,3	60	1

- a) $P_{22} = 2$ b) $P_7 = 1$ c) $P_{98} = 4$ d) $P_{66} = 3$

017 ¿Qué tipo de frecuencias se utilizan para calcular las medidas de posición?
¿Es la mediana una medida de posición?

Para calcular las medidas de posición se utilizan las frecuencias acumuladas.

La mediana se puede considerar una medida de posición, ya que divide la distribución de los datos en dos partes iguales:

$$Me = Q_2 = P_{50}$$

018 Salen 20 plazas a concurso por oposición y se presentan 200 personas.

Notas	3	4	5	6	7	8	9	10
f_i	6	25	34	42	50	27	13	3

¿Con qué nota se obtiene una de las plazas mediante el concurso por oposición?
¿Qué percentil es la nota 5?

Hay $200 - 20 = 180$ personas que suspenden la oposición. Como 180 es el 90 % de 200, y $P_{10} = 8$, siendo 8 la nota mínima para aprobar.

Ordenados los datos, del 32.º al 65.º tienen de nota 5, luego 5 es el percentil P_{16} , P_{17} , ..., hasta P_{32} .

019 Lidia ha obtenido las siguientes notas en Matemáticas: 7, 5, 6, 10, 9, 7 y 6.
Halla las medidas de dispersión.

$$\bar{x} = 7,14 \quad \sigma^2 = 2,69 \quad \sigma = 1,64 \quad CV = 0,23$$

020 Calcula las medidas de dispersión de estos datos.

N.º de vehículos	0	1	2	3
N.º de familias	115	456	268	161

x_i	f_i	$ x_i - \bar{x} $	$(x_i - \bar{x})^2$	$f_i \cdot (x_i - \bar{x})^2$
0	115	1,475	2,175	250,125
1	456	0,475	0,225	102,6
2	268	0,525	0,275	140,7
3	161	1,525	2,325	374,325
	1.000			867,75

$$\begin{aligned} \bar{x} &= 1,475 \\ \sigma^2 &= 0,867 \\ \sigma &= 0,931 \\ CV &= 0,631 \end{aligned}$$

Estadística

021 Compara la dispersión en estas dos variables: la primera mide el peso de los elefantes, con $\bar{x} = 2.000$ kg y $\sigma = 100$ kg, y la otra mide el peso de los ratones, con $\bar{x} = 0,05$ kg y $\sigma = 0,02$ kg.

$$CV_e = \frac{100}{2.000} = 0,05 \quad CV_r = \frac{0,02}{0,05} = 0,4$$

La dispersión en los ratones es mayor, ya que su coeficiente de variación es mayor.

022 Obtén y comenta las medidas estadísticas de estos datos: 1, 3, 2, 5, 2 y 5.

x_i	f_i	F_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$	$f_i \cdot x_i - \bar{x} $
1	1	1	1	1	2
2	2	3	4	8	2
3	1	4	3	9	0
5	2	6	10	50	4
Total	6		18	68	8

Las medidas de centralización son:

$$\bar{x} = \frac{18}{6} = 3 \quad Me = 2,5 \quad Mo = 2 \text{ y } 5$$

Las medidas de dispersión son:

$$\text{Rango: } R = \text{máximo} - \text{mínimo} = 5 - 1 = 4$$

$$\text{Desviación media: } DM = \frac{\sum f_i \cdot |x_i - \bar{x}|}{N} = \frac{8}{6} = 1,333\dots$$

$$\text{Varianza: } \sigma^2 = \frac{\sum f_i \cdot x_i^2}{N} - \bar{x}^2 = \frac{68}{6} - 9 = \frac{14}{6} = 2,333\dots$$

$$\text{Desviación típica: } \sigma = \sqrt{\sigma^2} = 1,5275$$

$$\text{Coeficiente de variación: } CV = \frac{\sigma}{\bar{x}} = \frac{1,5275}{3} = 0,509 = 50,9\%$$

Los datos presentan una agrupación pequeña respecto de las medidas de centralización.

023 Varía las frecuencias del ejercicio anterior para obtener una conclusión distinta.

Por ejemplo, con 1, 2, 2, 2, 3, 4 se obtendría una agrupación bastante mayor.

024 La tabla muestra el número de accidentes laborales que han tenido lugar en el último año.

N.º de accidentes	[0, 10)	[10, 20)	[20, 30)	[30, 40)
N.º de meses	1	5	4	2

Calcula las medidas de dispersión.

N.º de accidentes	x_i	f_i	F_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$	$f_i \cdot x_i - \bar{x} $
[0, 10)	5	1	1	5	25	15,83
[10, 20)	15	5	6	75	1.125	29,15
[20, 30)	25	4	10	100	2.500	16,68
[30, 40)	35	2	12	70	2.450	28,34
Total		12		250	6.100	90

Las medidas de centralización son:

$$\bar{x} = \frac{250}{12} = 20,83$$

Intervalo mediano = [20, 30)

Intervalo modal = [10, 20)

Las medidas de dispersión son:

Rango: $R = \text{máximo} - \text{mínimo} = 40 - 0 = 40$

$$\text{Desviación media: } DM = \frac{\sum f_i \cdot |x_i - \bar{x}|}{N} = \frac{90}{12} = 7,5$$

$$\text{Varianza: } \sigma^2 = \frac{\sum f_i \cdot x_i^2}{N} - \bar{x}^2 = \frac{6.100}{12} - 433,89 = 74,44$$

Desviación típica: $\sigma = \sqrt{\sigma^2} = 8,63$

$$\text{Coeficiente de variación: } CV = \frac{\sigma}{\bar{x}} = \frac{8,63}{20,83} = 0,414 = 41,4\%$$

ACTIVIDADES

025 Indica el tipo de variable estadística que estudiamos y razona, en cada caso, si sería mejor analizar una muestra o la población.

- a) La talla de los alumnos de un IES.
- b) La temperatura de tu provincia.
- c) La edad de los habitantes de un país.
- d) El sexo de los habitantes de un pueblo.
- e) El dinero gastado a la semana por tus amigos.
- f) Los efectos de un nuevo medicamento en el ser humano.
- g) El color del pelo de tus compañeros de clase.

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

Estadística

026



El número de horas diarias de estudio de 30 alumnos es:

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

- a) Efectúa un recuento y organiza los resultados en una tabla de frecuencias.
b) ¿Qué significan las frecuencias acumuladas que has calculado?

a)

N.º de horas	f_i	h_i	F_i	H_i
0	3	0,1	3	0,1
1	8	0,267	11	0,367
2	7	0,233	18	0,6
3	6	0,2	24	0,8
4	3	0,1	27	0,9
5	3	0,1	30	1

- b) Veamos el significado de las frecuencias acumuladas relativas mediante un ejemplo:
 $F_3 = 24$ quiere decir que hay 24 alumnos que estudian tres horas o menos cada día.
 $H_2 = 0,6$ significa que el 60 % de los alumnos estudia dos horas o menos cada día.

027



Completa la siguiente tabla de frecuencias.

x_i	h_i	F_i	Porcentaje
10	4	4	8
20	5	9	10
30	7	16	14
40	10	26	20
50	15	41	30
60	9	50	18

028



En una evaluación, de los 30 alumnos de una clase, el 10 % aprobó todo, el 20 % suspendió una asignatura, el 50 % suspendió dos asignaturas y el resto más de dos asignaturas.

- a) Realiza la tabla de frecuencias correspondiente.
b) ¿Hay algún tipo de frecuencia que responda a la pregunta de cuántos alumnos suspendieron menos de dos asignaturas? Razona tu respuesta.

a)

N.º de asignaturas suspendas	f_i	h_i	F_i	H_i
0	3	0,1	3	0,1
1	6	0,2	9	0,3
2	15	0,5	24	0,8
Más de 2	6	0,2	30	1

- b) La frecuencia que nos proporciona la respuesta a la pregunta de cuántos alumnos suspendieron menos de dos asignaturas, es la frecuencia acumulada, $F_1 = 9$.

- 029** Explica cómo completarías una tabla de frecuencias conociendo solo las frecuencias absolutas acumuladas. ¿Podrías hacer lo mismo con las frecuencias relativas acumuladas?

Para completar una tabla de frecuencias, conociendo las frecuencias absolutas acumuladas, es necesario tener en cuenta que:

$$F_1 = f_1$$

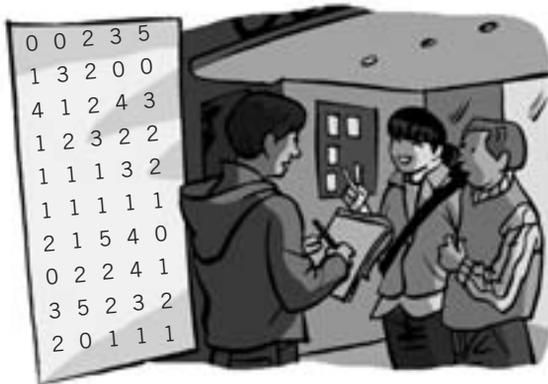
$$F_i = f_{i-1} + f_i \rightarrow f_i = F_i - F_{i-1}, \text{ para } i > 1$$

Conociendo las frecuencias relativas acumuladas, las frecuencias relativas se calculan de forma análoga, ya que se verifica:

$$H_1 = h_1$$

$$H_i = H_{i-1} + h_i \rightarrow h_i = H_i - H_{i-1}, \text{ para } i > 1$$

- 030** Para realizar un estudio hacemos una encuesta entre los jóvenes de un barrio, y les preguntamos por el número de veces que van al cine por semana. Los resultados de la encuesta son:



- a) ¿Cuál y de qué tipo es la variable estadística que estamos estudiando?
 b) Construye una tabla de frecuencias.
 c) ¿Cuántos jóvenes van al cine más de dos veces por semana?
 d) ¿Y cuántos van, al menos, una vez por semana?

a) La variable estadística que estamos estudiando es «el número de veces que los jóvenes van al cine por semana», y es una variable cuantitativa discreta.

b)

x_i	f_i	F_i	h_i	H_i
0	7	7	0,14	0,14
1	16	23	0,32	0,46
2	13	36	0,26	0,72
3	7	43	0,14	0,86
4	4	47	0,08	0,94
5	3	50	0,06	1

- c) 14 jóvenes van al cine más de dos veces por semana.
 d) Al menos una vez por semana van al cine 43 jóvenes.

Estadística

031

De los 30 asistentes a una cena, el 20 % comió ternera, el 40 % cordero y el resto pescado.

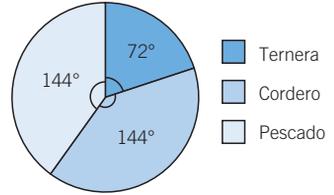


a) Organiza los resultados en una tabla de frecuencias, y representa los datos en un diagrama de sectores.

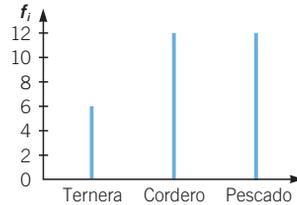
b) Realiza un diagrama de barras y explica cómo lo haces. ¿Cuál de los dos gráficos prefieres? ¿Por qué?

a)

Comida	f_i	F_i	h_i	H_i
Ternera	6	6	0,2	0,2
Cordero	12	18	0,4	0,6
Pescado	12	30	0,4	1



b) Representamos, en el diagrama de barras, las frecuencias absolutas. En este caso es preferible utilizar el gráfico de sectores porque es más fácil de comprender.



032



La siguiente tabla muestra los resultados de lanzar 50 veces un dado.

Cara	1	2	3	4	5	6
N.º de veces	8	12	5	9	6	10

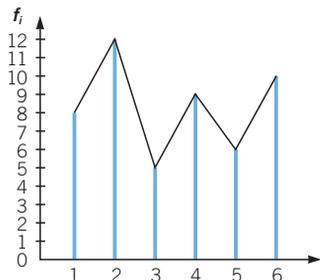
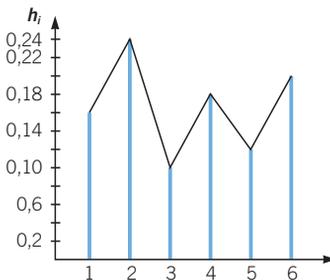
a) Representa los diagramas de barras de frecuencias relativas y absolutas. ¿Qué observas?

b) Sobre los gráficos anteriores, dibuja su polígono de frecuencias.

c) ¿Podrías representar los datos en un histograma? Razona tu respuesta.

a) y b) Al representar los dos diagramas se observa que el gráfico es el mismo, y lo único que cambia es el eje vertical; en un caso son frecuencias absolutas, y el en otro caso, son relativas.

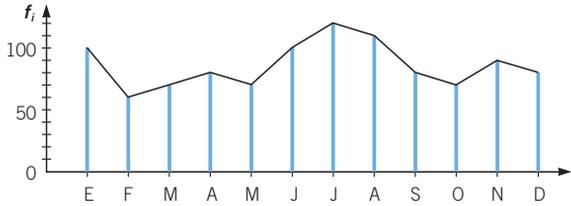
Cara	f_i	F_i	h_i	H_i
1	8	8	0,16	0,16
2	12	20	0,24	0,4
3	5	25	0,1	0,5
4	9	34	0,18	0,68
5	6	40	0,12	0,8
6	10	50	0,2	1



c) Estos datos no pueden representarse mediante un histograma porque la variable no es continua.

033

La venta de turismos durante un año en un concesionario viene representada por el siguiente gráfico.



- a) Halla las frecuencias absolutas y relativas.
- b) Obtén las frecuencias acumuladas.

a) y b)

Mes	f_i	F_i	h_i	H_i
Enero	100	100	0,09	0,09
Febrero	60	160	0,05	0,14
Marzo	70	230	0,06	0,2
Abril	80	310	0,07	0,27
Mayo	70	380	0,06	0,33
Junio	100	480	0,09	0,42
Julio	120	600	0,11	0,52
Agosto	110	710	0,1	0,62
Septiembre	80	790	0,07	0,69
Octubre	70	860	0,06	0,75
Noviembre	90	950	0,08	0,83
Diciembre	80	1.030	0,07	0,9

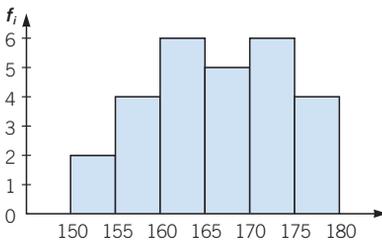
034

Las estaturas, en cm, de 27 jóvenes son las siguientes.

155 178 170 165 173 168 160 166 176
 169 158 170 179 161 164 156 170 171
 167 151 163 158 164 174 176 164 154

Utiliza intervalos de amplitud 5, comenzando con el intervalo [150, 155), forma una tabla, efectúa el recuento y obtén las marcas de clase. Representa los datos en un histograma.

Estatura	x_i	f_i	h_i	F_i	H_i
[150, 155)	152,5	2	0,074	2	0,074
[155, 160)	157,5	4	0,148	6	0,222
[160, 165)	162,5	6	0,222	12	0,444
[165, 170)	167,5	5	0,185	17	0,629
[170, 175)	172,5	6	0,222	23	0,851
[175, 180)	177,5	4	0,148	27	1



Estadística

035



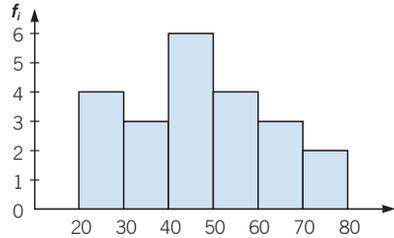
Hemos estudiado el contenido en sales de 22 botellas de agua, y obtenemos los siguientes datos expresados en miligramos.

46 25 27 30 48 40
 27 44 37 62 56 29
 76 75 49 59 33 52
 54 45 66 69



- Clasifica la variable estadística estudiada.
- Justifica el hecho de tomar o no intervalos al hacer una tabla.
- Realiza el gráfico que consideres más adecuado.
 - La variable estadística es cuantitativa continua.
 - Debido al tamaño de la muestra, es conveniente tomar intervalos.
 - Para realizar cualquier representación gráfica es necesario construir primero la tabla de frecuencias. Construimos el histograma.

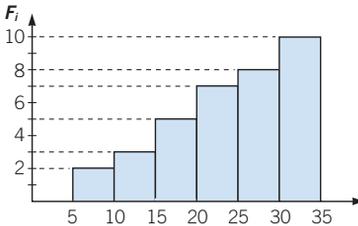
Sales	x_i	f_i
[20, 30)	25	4
[30, 40)	35	3
[40, 50)	45	6
[50, 60)	55	4
[60, 70)	65	3
[70, 80)	75	2



036



Reconstruye la tabla de frecuencias asociada a este gráfico de frecuencias acumuladas.



Intervalos	x_i	f_i	h_i	F_i	H_i
[5, 10)	7,5	2	0,2	2	0,2
[10, 15)	12,5	1	0,1	3	0,3
[15, 20)	17,5	2	0,2	5	0,5
[20, 25)	22,5	2	0,2	7	0,7
[25, 30)	27,5	1	0,1	8	0,8
[30, 35)	32,5	2	0,2	10	1

037 HAZLO ASÍ

¿CÓMO SE DIBUJA UN DIAGRAMA DE CAJAS?

Un diagrama de cajas es un gráfico en el que se dibuja una caja central, que indica el intervalo en el que se concentra el 50 % de los datos (sus extremos son el 1.º y 3.º cuartiles) y una línea central que marca la mediana. A partir de él podemos detectar datos atípicos que se alejan del resto. Con esta información, representa estos datos mediante un diagrama de cajas.

x_i	1	2	3	4	5	6	7	8
f_i	1	7	8	2	1	1	6	4
F_i	1	8	16	18	19	20	26	30

PRIMERO. Se calcula la mediana, Q_1 y Q_3 .

$$Me = 3$$

$$Q_1 = 2$$

$$Q_3 = 7$$

SEGUNDO. Se representan estos datos en una recta.

TERCERO. Se dibuja un rectángulo de anchura comprendida entre Q_1 y Q_3 , una línea que pase por la mediana y rectas exteriores que marcan el rango de los datos.



038 Dibuja un diagrama de cajas para estos datos.

x_i	1	2	3	4	5	6
f_i	3	1	2	1	4	1

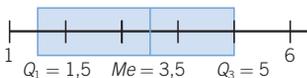
x_i	12	13	14	15
f_i	11	9	8	62

x_i	f_i	h_i	F_i	H_i
1	3	0,25	3	0,25
2	1	0,08	4	0,33
3	2	0,17	6	0,5
4	1	0,08	7	0,58
5	4	0,33	11	0,92
6	1	0,08	12	1

$$Q_1 = 1,5$$

$$Me = 3,5$$

$$Q_3 = 5$$



Estadística

039 Obtén las medidas de centralización de la siguiente serie de datos.

7 3 2 4 5 1 8 6 1 5
 3 2 4 9 8 1 0 2 4 1
 2 5 6 5 4 7 1 3 0 5
 8 6 3 4 0 9 2 5 7 4
 0 2 1 5 6 4 3 5 2 3

x_i	f_i	F_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$	$f_i \cdot x_i - \bar{x} $
0	4	4	0	0	15,44
1	6	10	6	6	17,16
2	7	17	14	28	13,02
3	6	23	18	54	5,16
4	7	30	28	112	0,98
5	8	38	40	200	9,12
6	4	42	24	144	8,56
7	3	45	21	147	9,42
8	3	48	24	192	12,42
9	2	50	18	162	10,28
Total	50		193	1.045	101,56

Las medidas de centralización son: $\bar{x} = \frac{193}{50} = 3,86$ $Me = 4$ $Mo = 5$

Las medidas de dispersión son: Rango: $R = \text{máximo} - \text{mínimo} = 9 - 0 = 9$

Desviación media: $DM = \frac{\sum f_i \cdot |x_i - \bar{x}|}{N} = \frac{101,56}{50} = 2,0312$

Varianza: $\sigma^2 = \frac{\sum f_i \cdot x_i^2}{N} - \bar{x}^2 = \frac{1.045}{50} - 14,9 = 6$

Desviación típica: $\sigma = \sqrt{\sigma^2} = 2,45$

Coefficiente de variación: $CV = \frac{\sigma}{\bar{x}} = \frac{2,45}{3,86} = 0,63 = 63\%$

040 Realiza la actividad anterior, pero agrupa en intervalos de amplitud 2, comenzando en cero. ¿Obtienes los mismos resultados? ¿Por qué?

Intervalos	x_i	f_i	F_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$	$f_i \cdot x_i - \bar{x} $
[0, 2)	1	10	10	10	10	33,6
[2, 4)	3	13	23	39	117	17,68
[4, 6)	5	15	38	75	375	9,6
[6, 8)	7	7	45	49	343	18,48
[8, 10)	9	5	50	45	405	23,2
Total		50		218	1.250	102,56

Las medidas de centralización son:

$\bar{x} = 4,36$ Intervalo mediano = [4, 6) Intervalo modal = [4, 6)

Las medidas de dispersión son: Rango: $R = \text{máximo} - \text{mínimo} = 9 - 0 = 9$

Desviación media: $DM = \frac{\sum f_i \cdot |x_i - \bar{x}|}{N} = \frac{102,56}{50} = 2,05$

$$\text{Varianza: } \sigma^2 = \frac{\sum f_i \cdot x_i^2}{N} - \bar{x}^2 = \frac{1.250}{50} - 19 = 6$$

$$\text{Desviación típica: } \sigma = \sqrt{\sigma^2} = 2,45$$

$$\text{Coeficiente de variación: } CV = \frac{\sigma}{\bar{x}} = \frac{2,45}{4,36} = 0,56 = 56 \%$$

Se observa que no se obtienen los mismos resultados. Hemos trabajado con marcas de clase debido a la agrupación de los datos.

- 041** Una cadena de televisión ha realizado un estudio entre 200 espectadores para determinar el grado de satisfacción de un programa, obteniendo estos resultados.

Opinión	Muy bueno	Bueno	Regular	Malo	Muy malo
Porcentaje	15	25	30	25	5

Calcula e interpreta las medidas de centralización.

$Me = \text{Regular} \rightarrow$ Significa que la mitad de los datos son peores o iguales que Regular y la otra mitad son mejores o iguales que Regular.

$Mo = \text{Regular} \rightarrow$ Significa que Regular es la opinión en la que coinciden más espectadores.

042 HAZLO ASÍ

¿CÓMO SE PUEDEN AÑADIR O SUPRIMIR DATOS PARA OBTENER UNA MEDIA DETERMINADA?

Añade un dato a este conjunto para que:

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

- a) La media no varíe. b) La media sea 6.

PRIMERO. Se calcula la media de los datos.

$$N = 15 \rightarrow \bar{x} = \frac{75}{15} = 5$$

SEGUNDO. Se multiplica la media que se quiere obtener por el nuevo número de datos, $N = 16$.

- a) $16 \cdot 5 = 80$ b) $16 \cdot 6 = 96$

TERCERO. Se resta $\sum_{i=1}^n f_i \cdot x_i$ a este resultado y se obtiene el nuevo resultado.

- a) $80 - 75 = 5$. Hay que añadir un 5. b) $96 - 75 = 21$. Hay que añadir un 21.

- 043** Añade dos datos a este conjunto para que la media cumpla estas condiciones.

5 5 5 5 8 8 8 8 8 10

- a) No varíe. b) Sea 8. c) Sea 4.

$$\bar{x} = 7$$

a) Los dos números deben sumar 14, por ejemplo 7 y 7.

b) $8 \cdot 12 = 96$; $96 - 70 = 26$. La suma de los dos números debe ser 26. Por ejemplo, 10 y 16.

c) $4 \cdot 12 = 48$; $48 - 70 = -22$. La suma de los dos números debe ser -22 . Por ejemplo, -10 y -12 .

044 HAZLO ASÍ

¿CÓMO SE PUEDEN AÑADIR O SUPRIMIR DATOS PARA OBTENER UNA MEDIANA DETERMINADA?

Añade un dato a este conjunto para que la mediana no varíe.

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

PRIMERO. Se calcula la mediana: $Me = 5$

SEGUNDO. Al añadir datos, la mediana se desplazará tantos lugares como datos se añadan.

Al añadir un dato, $N = 16$ y la mediana será la media entre los datos colocados en el 8.º y 9.º lugares (5 y 5). Por tanto, la mediana será siempre 5.

045 Añade dos datos a esta distribución para que la mediana:



8 8 8 8 8 9 9 10 10 10

a) No varíe.

b) Sea 8.

c) Sea 9.

- a) Uno de los datos debe ser mayor o igual a 9 y el otro dato debe ser menor o igual a 8. Por ejemplo, 5 y 11.
 b) Los dos datos deben ser menores o iguales que 8. Por ejemplo, 2 y 6.
 c) Los datos deben ser mayores o iguales a 9. Por ejemplo, 10 y 11.

046 Un corredor entrena, de lunes a viernes, recorriendo las siguientes distancias: 2, 5, 5, 7 y 3 km, respectivamente. Si el sábado también entrena:



a) ¿Cuántos kilómetros debe recorrer para que la media sea la misma?

b) ¿Y para que la mediana no varíe?

c) ¿Y para que la moda no varíe?

- a) La distancia media que recorre, de lunes a viernes, es:

$$\frac{2 + 5 + 5 + 7 + 3}{5} = \frac{22}{5} = 4,4 \text{ km}$$

Suponiendo que el sábado recorre d kilómetros, para que se conserve la media se ha de cumplir que:

$$\frac{2 + 5 + 5 + 7 + 3 + d}{6} = 4,4 \rightarrow \frac{22 + d}{6} = 4,4$$

$$\rightarrow d = 4,4 \cdot 6 - 22 = 44 \text{ km}$$

El sábado deberá recorrer 44 km.

- b) La mediana es $Me = 5$.

2 3 5 5 7

Para que la mediana no varíe, el sábado debe recorrer 5 km o más.

- c) Para que la moda no varíe es necesario que el sábado recorra cualquier distancia excepto 2, 3 o 7 kilómetros. De esta manera, la moda seguirá siendo 5.



- 047** La tabla muestra las notas obtenidas por 120 alumnos en una prueba de 100 preguntas.

Notas	x_i	f_i	F_i
[30, 40)	35	1	1
[40, 50)	45	3	4
[50, 60)	55	11	15
[60, 70)	65	21	36
[70, 80)	75	43	79
[80, 90)	85	32	111
[90, 100)	95	9	120
Total		$N = 120$	

- a) Calcula los tres cuartiles. b) Halla los percentiles P_{23} , P_{46} y P_{90} .**

Para hallar los cuartiles y percentiles es necesario calcular primero las frecuencias acumuladas.

- a) El 25 % de 120 es 30, luego Q_1 debe dejar 30 datos por debajo y el resto por encima. En la columna de frecuencias acumuladas, el primer número mayor o igual que 30 es 36, por lo que $Q_1 = 65$.

Como el 50 % de 120 es 60, repitiendo el proceso, tenemos que $Q_2 = 75$.

Análogamente, como el 75 % de 120 es 90, resulta que $Q_3 = 85$.

- b) El 23 % de 120 es 27,6; y la primera frecuencia acumulada mayor o igual que 27,6 es 36, que se corresponde con la marca de clase: 65, luego $P_{23} = 65$.

Como el 46 % de 120 es 55,2; repitiendo el proceso, resulta que $P_{46} = 75$.

Análogamente, como el 90 % de 120 es 108, tenemos que $P_{90} = 85$.

- 048** Contesta razonadamente a las siguientes cuestiones sobre medidas estadísticas.

- a) ¿Es la media siempre mayor que la desviación típica?
 b) En un estudio estadístico, ¿qué cantidad de datos es mayor que el cuartil primero, Q_1 , y menor que el cuartil tercero, Q_3 ?
 c) ¿Qué significa que el peso de un niño está situado en P_{90} ?

- a) No, por ejemplo, si todos los datos fueran negativos, la media sería negativa, pero la desviación típica es siempre positiva.
 b) Sabemos que, entre dos cuartiles consecutivos cualesquiera, hay un 25 % de los datos, luego entre el cuartil primero, Q_1 , y el cuartil tercero, Q_3 , está el 50 % de los datos.
 c) Si el peso de un niño está situado en el percentil P_{90} , esto quiere decir que el 90 % de los niños pesa menos que lo que este indica.

Estadística

- 049** ●● **Calcula e interpreta las medidas de dispersión de los siguientes datos, que expresan los días de baja por enfermedad de 10 trabajadores de una fábrica.**

0 2 3 4 2 1 1 0 0 3

x_i	f_i	F_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$	$f_i \cdot x_i - \bar{x} $
0	3	3	0	0	4,8
1	2	5	2	2	1,2
2	2	7	4	8	0,8
3	2	9	6	18	2,8
4	1	10	4	16	2,4
Total	10		16	44	12

$$\bar{x} = \frac{16}{10} = 1,6$$

Las medidas de dispersión son:
Rango: $R = \text{máximo} - \text{mínimo} = 4 - 0 = 4$

Desviación media:

$$DM = \frac{\sum f_i \cdot |x_i - \bar{x}|}{N} = \frac{12}{10} = 1,2$$

$$\text{Varianza: } \sigma^2 = \frac{\sum f_i \cdot x_i^2}{N} - \bar{x}^2 = \frac{44}{10} - 2,56 = 1,84$$

$$\text{Desviación típica: } \sigma = \sqrt{\sigma^2} = 1,36$$

$$\text{Coeficiente de variación: } CV = \frac{\sigma}{\bar{x}} = \frac{1,36}{1,6} = 0,85 = 85\%$$

- 050** ●● **Una persona ingresa 6.000 € en un fondo de inversión el 1 de enero de 2002. Las rentabilidades anuales del fondo durante los años siguientes han sido:**

Año	2002	2003	2004	2005
Rentabilidad (%)	5	4	-3	5

Si no ha retirado el capital, ¿cuál ha sido la rentabilidad media de dicho fondo durante estos años?

$$\text{La rentabilidad media ha sido: } \bar{x} = \frac{5 + 4 + (-3) + 5}{4} = 2,75\%$$

- 051** ●● **Realiza el estudio del peso, en kg, de 20 alumnos que se muestra en la tabla.**

Peso	x_i	f_i	F_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$	$f_i \cdot x_i - \bar{x} $
[36, 42)	39	4	4	156	6.084	50,4
[42, 48)	45	4	8	180	8.100	26,4
[48, 54)	51	5	13	255	13.005	3
[54, 60)	57	2	15	114	6.498	10,8
[60, 66)	63	3	18	189	11.907	34,2
[66, 72)	69	2	20	138	9.522	34,8
Total		20		1.032	55.116	159,6

Las medidas de centralización son:

$$\bar{x} = \frac{1.032}{20} = 51,6 \quad \text{Intervalo mediano} = [48, 54) \quad \text{Intervalo modal} = [48, 54)$$

Las medidas de dispersión son:

$$\text{Rango: } R = \text{máximo} - \text{mínimo} = 72 - 36 = 36$$

$$\text{Desviación media: } DM = \frac{\sum f_i \cdot |x_i - \bar{x}|}{N} = \frac{159,6}{20} = 7,98$$

$$\text{Varianza: } \sigma^2 = \frac{\sum f_i \cdot x_i^2}{N} - \bar{x}^2 = \frac{55.116}{20} - 51,6^2 = 93,24$$

$$\text{Desviación típica: } \sigma = \sqrt{\sigma^2} = 9,65$$

$$\text{Coeficiente de variación: } CV = \frac{\sigma}{\bar{x}} = \frac{9,65}{51,6} = 0,19 = 19\%$$

052

Los salarios, en euros, en una empresa son los siguientes.

Mujeres: 1.200, 1.300, 1.000, 900, 900, 1.100, 1.200, 1.100, 1.400, 1.200, 1.000, 1.300, 1.200, 1.100, 1.100

Hombres: 1.200, 1.300, 1.500, 1.300, 1.400, 900, 1.700, 1.600, 1.400, 1.300, 1.500, 1.300, 1.900, 1.700, 1.200

- Calcula la distribución de frecuencias, la media, la mediana y la desviación típica, de cada grupo: hombres y mujeres.
- Calcula sus medidas de dispersión.
- Compara ambos grupos. ¿Cómo lo haces?
- Si consideramos todos los datos en el mismo grupo, ¿qué resultados obtenemos?

a) y b) Mujeres:

x_i	f_i	F_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$
900	2	2	1.800	1.620.000
1.000	2	4	2.000	2.000.000
1.100	4	8	4.400	4.840.000
1.200	4	12	4.800	5.760.000
1.300	2	14	2.600	3.380.000
1.400	1	15	1.400	1.960.000
Total	15		17.000	19.580.000

Las medidas de centralización son:

$$\bar{x} = \frac{17.000}{15} = 1.133 \quad Me = 1.100 \quad Mo = 1.100 \text{ y } 1.200$$

Las medidas de dispersión son:

$$\text{Rango: } R = \text{máximo} - \text{mínimo} = 1.400 - 900 = 500$$

$$\text{Varianza: } \sigma^2 = \frac{\sum f_i \cdot x_i^2}{N} - \bar{x}^2 = \frac{19.580.000}{15} - 1.133^2 = 21.644$$

$$\text{Desviación típica: } \sigma = \sqrt{\sigma^2} = 147$$

$$\text{Coeficiente de variación: } CV = \frac{\sigma}{\bar{x}} = \frac{147}{1.133} = 0,13 = 13\%$$

Estadística

Hombres:

x_i	f_i	F_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$
900	1	1	900	810.000
1.200	2	3	2.400	2.880.000
1.300	4	7	5.200	6.760.000
1.400	2	9	2.800	3.920.000
1.500	2	11	3.000	4.500.000
1.600	1	12	1.600	2.560.000
1.700	2	14	3.400	5.780.000
1.900	1	15	1.900	3.610.000
Total	15		21.200	30.920.000

Las medidas de centralización son:

$$\bar{x} = \frac{21.200}{15} = 1.413 \quad Me = 1.400 \quad Mo = 1.300$$

Las medidas de dispersión son:

$$\text{Rango: } R = \text{máximo} - \text{mínimo} = 1.900 - 900 = 1.000$$

$$\text{Varianza: } \sigma^2 = \frac{\sum f_i \cdot x_i^2}{N} - \bar{x}^2 = \frac{30.920.000}{15} - 1.413^2 = 64.764$$

$$\text{Desviación típica: } \sigma = \sqrt{\sigma^2} = 254$$

$$\text{Coeficiente de variación: } CV = \frac{\sigma}{\bar{x}} = \frac{254}{1.413} = 0,18 = 18\%$$

- c) A la vista de los resultados, podemos afirmar que el salario medio en los hombres es mayor que en las mujeres. En ambos casos, la desviación típica es pequeña con relación a la media. Esto significa que los datos están bastante próximos al respectivo valor medio, estando más próximos en las mujeres que en los hombres, ya que el coeficiente de variación en los hombres es mayor que en las mujeres. En el caso de los hombres, los datos están más dispersos que en el caso de las mujeres.

d)

x_i	f_i	F_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$
900	3	3	2.700	2.430.000
1.000	2	5	2.000	2.000.000
1.100	4	9	4.400	4.840.000
1.200	6	15	7.200	8.640.000
1.300	6	21	7.800	10.140.000
1.400	3	24	4.200	5.880.000
1.500	2	26	3.000	4.500.000
1.600	1	27	1.600	2.560.000
1.700	2	29	3.400	5.780.000
1.900	1	30	1.900	3.610.000
Total	30		38.200	50.380.000

$$\bar{x} = 1.273,33 \quad Me = 1.250 \quad Mo = 1.200 \text{ y } 1.300$$

$$\text{Rango: } R = 1.900 - 900 = 1.000$$

$$\sigma^2 = \frac{50.380.000}{30} - 1.273^2 = 58.804 \quad \sigma = \sqrt{\sigma^2} = 242$$

$$CV = \frac{\sigma}{\bar{x}} = 0,19 = 19\%$$

053

El tiempo, en minutos, que un conjunto de estudiantes dedica a preparar un examen es:

220 500 450 390 550 600
790 200 60 300 400 90

Las calificaciones de ese conjunto de estudiantes son las siguientes.

4 5 6 5 7 6 8 4 1 5 6 2

¿Cuál es la media y la desviación típica de ambos conjuntos? ¿Qué podemos hacer para comparar su variabilidad? ¿En qué conjunto los datos están más dispersos?

	Minutos de estudio	Calificaciones
Media	379	4,92
Varianza	43.037	3,58
Desviación típica	207	1,89
Coefficiente de variación	0,55	0,38

Para comparar la variabilidad nos fijamos en el coeficiente de variación. Están más dispersos los datos de los minutos dedicados al estudio.

054

Dos alumnos realizan 5 pruebas de calificación, obteniendo los siguientes resultados.

Juan: 2 6 5 7 5

Ana: 0 1 9 8 7

Compara sus datos, utilizando la media aritmética y la desviación típica.

La media aritmética y la desviación típica de ambos estudiantes son:

$$\text{Juan: } \bar{x} = 5 \quad \sigma = 1,67 \quad \text{Ana: } \bar{x} = 5 \quad \sigma = 3,74$$

Las dos medias son iguales, pero tienen distinto significado dependiendo de sus desviaciones típicas.

En ambos casos, las medias son iguales y valen 5.

Sin embargo, Juan tiene una desviación típica mucho menor que Ana. Esto significa que Juan es un alumno constante, pues sus notas están próximas a la media. Por el contrario, podemos afirmar que Ana es una alumna bastante irregular, porque alterna notas muy altas y bajas, estando todas excesivamente alejadas de la media.

055

Un grupo de ratones tiene de media de sus pesos $\bar{x} = 70$ g y desviación típica $\sigma = 20$ g. Un conjunto de gatos tiene de media $\bar{x} = 2,5$ kg y desviación típica $\sigma = 20$ g. Compara ambos grupos.



Aunque las desviaciones típicas sean iguales, debido a la diferencia existente entre las medias, podemos decir que en el grupo de ratones hay más dispersión en los datos que en el grupo de gatos: $CV_R = \frac{20}{70} > \frac{20}{2.500} = CV_G$

056



Los diplomados en Informática de gestión tienen un salario medio, en su primer empleo, de 1.080 €, con una desviación típica de 180 €. Los diplomados en Informática de sistemas tienen un salario medio de 960 €, con una desviación típica de 150 €. Si a un diplomado en Informática de gestión le ofrecen un sueldo de 1.200 €, y a un diplomado en Informática de sistemas, un sueldo de 1.140 €, ¿cuál recibe una mejor oferta? ¿Por qué?

Para poder comparar ambas ofertas vamos a medir sus beneficios en unidades de desviación típica.

Sabiendo que un diplomado en Informática de gestión tiene un salario medio de 1.080 €, con una desviación típica de 180 €, podemos decir que la oferta de 1.200 € se desvía por encima de la media:

$$\frac{1.200 - 1.080}{180} = 0,75 \text{ unidades de desviación típica}$$

Sin embargo, una oferta de 1.140 € a un diplomado en Informática de sistemas, con un sueldo medio de 960 € y 150 € de desviación típica, se desvía por encima de la media:

$$\frac{1.140 - 960}{150} = \frac{180}{150} = 1,2 \text{ unidades de desviación típica}$$

Esto indica que el diplomado en Informática de sistemas es quien recibe la mejor oferta.

057



La edad media de los integrantes de una orquesta aumentaría en un año si abandonarían la orquesta 5 músicos de 19 años cada uno, o si se unieran a ella 5 músicos de 17 años cada uno. ¿Es posible esta situación?

Es imposible, ya que si la edad media aumenta quitando 5 músicos de 19 años, esto quiere decir que la media era menor de 19 años, y si aumenta añadiendo 5 músicos de 17 años, significa que la media es inferior a 17, por lo que es imposible.

058



El peso medio de 4 amigos es 90 kg, pero hay un error, pues el peso de Carlos es 79 kg, y no 97 kg. ¿Cuál es realmente el peso medio?

$$\bar{x} = \frac{4 \cdot 90 - 18}{4} = 85,5$$

El peso medio es 85,5 kg.

059



El salario mensual, en euros, de los cinco trabajadores de una empresa es el siguiente.

1.500	1.500	2.000	2.700	11.000
-------	-------	-------	-------	--------

¿Cuál de las tres medidas de centralización describe mejor los sueldos de la empresa?

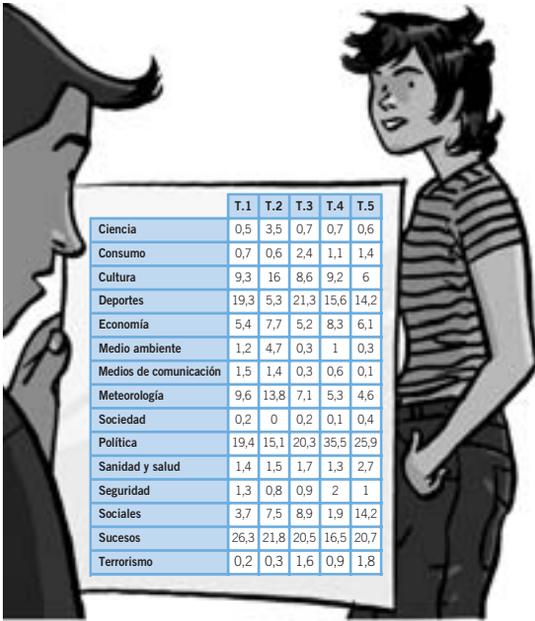
La mediana, ya que la moda nos da el valor mínimo del salario mensual, mientras que la mayoría de la empresa cobra menos de la media.

EN LA VIDA COTIDIANA

060

Para la revista de un centro escolar se hace un estudio estadístico que consiste en analizar el tipo de noticias que ofrecen los informativos de las principales cadenas de televisión.

Después de elegir qué variables se van a estudiar, los alumnos han organizado los datos en una tabla que muestra el porcentaje de noticias de cada tipo que se ha emitido.



Construye un gráfico adecuado y contesta a estas preguntas.

- a) ¿Qué tres tipos de noticias se emiten más en cada cadena?
- b) ¿Cuál de las cinco cadenas reparte más uniformemente sus noticias?
- c) ¿Consideras que alguna noticia interesa mucho más o mucho menos que el resto?

a)

	Noticias más emitidas		
	1.º	2.º	3.º
T.1	Sucesos	Política	Deportes
T.2	Sucesos	Cultura	Política
T.3	Deportes	Sucesos	Política
T.4	Política	Sucesos	Deportes
T.5	Política	Sucesos	Deportes y Sociales

Estadística

b) Calculamos las varianzas.

	T.1	T.2	T.3	T.4	T.5
Varianza	66,9	44,42	57,51	86,93	62,96

La menor varianza la tiene la cadena T.2, por lo que es la cadena que reparte más uniformemente los contenidos.

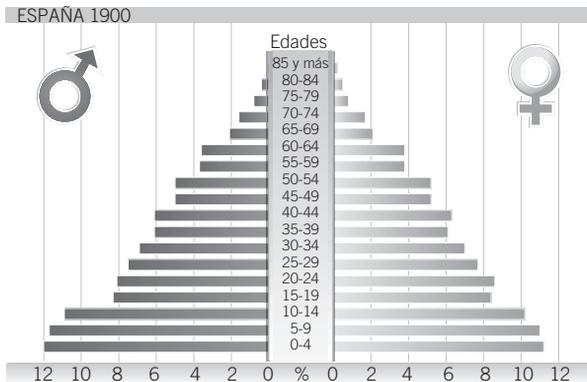
c) Hallamos la media de cada tipo de contenido.

	Media
Ciencia	1,2
Consumo	1,24
Cultura	9,82
Deportes	15,14
Economía	6,54
Medio ambiente	1,5
Medios de comunicación	0,78
Meteorología	8,08
Sociedad	0,18
Política	23,24
Sanidad y salud	1,72
Seguridad	1,2
Sociales	7,24
Sucesos	21,16
Terrorismo	0,96

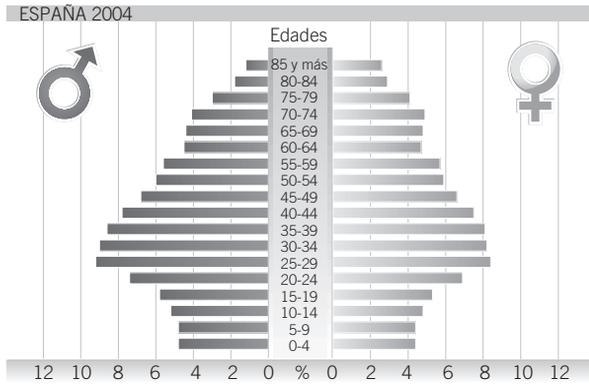
Las noticias que más interesan son las referentes a política, sucesos y deportes.

061

En una revista científica se ha publicado un informe sobre la evolución de la estructura de la población en España durante un siglo.



Los datos se muestran en pirámides de población, es decir, diagramas de barras, donde se representa la estructura de la población por intervalos de edades y porcentaje de cada sexo.



Para interpretar una pirámide de población debemos fijarnos en su forma.



Fíjate en las dos pirámides de población y describe cómo ha cambiado la estructura de la población española de 1900 y la de un siglo después. ¿Qué diferencias puedes destacar?

La pirámide de población en 2004 empieza a ser invertida, y eso quiere decir que la natalidad está disminuyendo cada año, desde hace 25 años, y que la esperanza de vida va aumentando.

Se observa una mayor esperanza de vida en 2004 que en 1900. En 2004 hay más porcentaje de personas de mayor edad.