

## Página 215

## PRACTICA

## Media y desviación típica

- 1 El número de faltas de ortografía que cometieron un grupo de estudiantes en un dictado fue:

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

- a) Di cuál es la variable y de qué tipo es.  
 b) Haz una tabla de frecuencias y representa los datos en un diagrama adecuado.  
 c) Calcula la media y la desviación típica.

- a) Variable: “Número de faltas de ortografía”

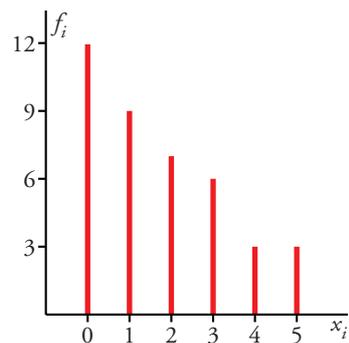
Es una variable cuantitativa discreta.

Llamamos  $x_i$  a dicha variable y sus valores son 0, 1, 2, 3, 4 y 5.

- b) Tabla de frecuencias:

$x_i$	$f_i$	$f_i x_i$	$f_i x_i^2$
0	12	0	0
1	9	9	9
2	7	14	28
3	6	18	54
4	3	12	48
5	3	15	75
	40	68	214

Diagrama de barras:



Nº DE FALTAS DE ORTOGRAFÍA

c) MEDIA:  $\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{68}{40} = 1,7$

$$\sigma^2 = \frac{\sum f_i x_i^2}{\sum f_i} - \bar{x}^2 = \frac{214}{40} - 1,7^2 = 2,46$$

DESVIACIÓN TÍPICA:  $\sigma = \sqrt{2,46} = 1,57$

- 2 A un grupo de 30 personas se les ha tomado el número de pulsaciones por minuto (ritmo cardíaco) obteniéndose los siguientes resultados:

87 85 61 51 64 75 80 70 69 82

80 79 82 74 90 76 72 73 63 65

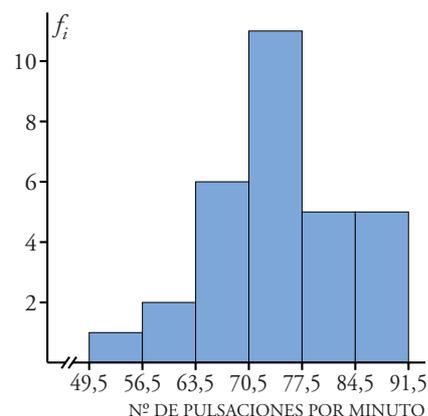
67 71 88 76 68 73 70 76 71 86

- a) Representa gráficamente esta distribución agrupando los datos en 6 intervalos.  
b) Calcula la media y la desviación típica.

- a) • Localizamos los valores extremos: 51 y 90 → recorrido = 39  
• Buscamos un múltiplo de 6 (nº de intervalos) algo mayor que 39, por ejemplo  $r' = 42$ . Así, cada intervalo tendrá una longitud de  $\frac{42}{6} = 7$ .

INTERVALOS	MARCAS DE CLASE ( $x_i$ )	$f_i$	$f_i x_i$	$f_i x_i^2$
49,5 - 56,5	53	1	53	2 809
56,5 - 63,5	60	2	120	7 200
63,5 - 70,5	67	6	402	26 934
70,5 - 77,5	74	11	814	60 236
77,5 - 84,5	81	5	405	32 805
84,5 - 91,5	88	5	440	38 720
		30	2 234	168 704

Puesto que los intervalos son de la misma longitud, la altura de cada barra en este histograma coincide con la frecuencia ( $f_i$ ).

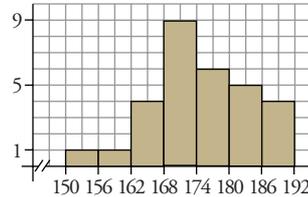


b) MEDIA:  $\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{2 234}{30} = 74,47$

$$\sigma^2 = \frac{\sum f_i x_i^2}{\sum f_i} - \bar{x}^2 = \frac{168 704}{30} - 74,47^2 = 77,69$$

DESVIACIÓN TÍPICA =  $\sigma = \sqrt{77,69} = 8,81$

- 3 Este gráfico muestra las alturas de los árboles de un parque. Haz la tabla de frecuencias correspondiente y calcula  $\bar{x}$  y  $\sigma$ .



INTERVALOS	MARCA DE CLASE ( $x_i$ )	$f_i$	$f_i x_i$	$f_i x_i^2$
150 - 156	153	1	153	23 409
156 - 162	159	1	159	25 281
162 - 168	165	4	660	108 900
168 - 174	171	9	1 539	263 169
174 - 180	177	6	1 062	187 974
180 - 186	183	5	915	167 445
186 - 192	189	4	756	142 884
		30	5 244	919 062

$$\text{MEDIA: } \bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{5\,244}{30} = 174,8$$

$$\sigma^2 = \frac{\sum f_i x_i^2}{\sum f_i} - \bar{x}^2 = \frac{919\,062}{30} - 174,8^2 = 80,36$$

$$\text{DESVIACIÓN TÍPICA: } \sigma = \sqrt{80,36} = 8,96$$

- 4 En una maternidad se han tomado los pesos (en kilogramos) de 50 recién nacidos:

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

- a) Construye una tabla con los datos agrupados en 6 intervalos de amplitud 0,4 kg.  
 b) Representa gráficamente esta distribución.  
 c) Calcula la media y la desviación típica.

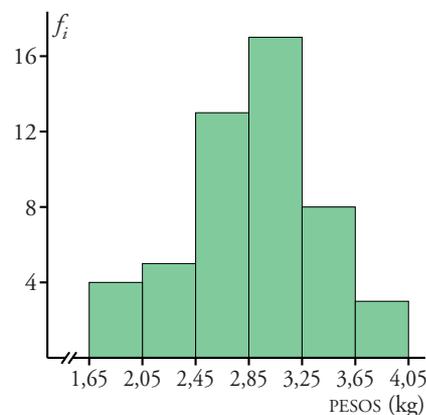
Localizamos los valores extremos: 1,8 y 3,9.

$$\text{Recorrido} = 3,9 - 1,8 = 2,1$$

a)

INTERVALOS	MARCA DE CLASE ( $x_i$ )	$f_i$	$f_i x_i$	$f_i x_i^2$
1,65 - 2,05	1,85	4	7,4	13,69
2,05 - 2,45	2,25	5	11,25	25,31
2,45 - 2,85	2,65	13	34,45	91,29
2,85 - 3,25	3,05	17	51,85	158,14
3,25 - 3,65	3,45	8	27,6	95,22
3,65 - 4,05	3,85	3	11,55	44,47
		50	144,1	428,12

b) Representamos los datos en un histograma; al ser los intervalos de la misma amplitud, la altura de cada barra corresponde a la frecuencia ( $f_i$ ) de cada intervalo.



c)  $\bar{x} = \frac{144,1}{50} = 2,9 \text{ kg}$

$$\sigma^2 = \frac{428,12}{50} - 2,9^2 = 0,1524$$

$$\sigma = \sqrt{0,1524} = 0,39 \text{ kg}$$

**5** El número de personas que acudieron a las clases de natación de una piscina municipal fueron:

38	32	54	47	50	58	46
47	55	60	43	60	45	48
40	53	59	48	39	48	56
52	48	55	60	53	43	52
46	55	56	54	48	39	50

a) Haz una tabla de frecuencias agrupando los datos en intervalos.

b) Representa gráficamente la distribución.

c) Halla  $\bar{x}$  y  $\sigma$ .

Localizamos los valores extremos: 32 y 60.

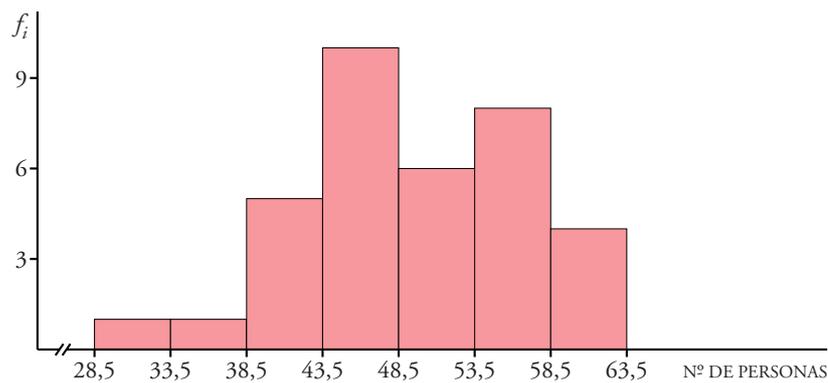
$$\text{Recorrido} = 60 - 32 = 28$$

Agrupamos los datos en 7 intervalos. Con el fin de que los extremos de los intervalos no coincidan con ninguno de los datos, tomamos cada intervalo de longitud 5, en vez de 4.

a)

INTERVALOS	MARCA DE CLASE ( $x_i$ )	$f_i$	$f_i x_i$	$f_i x_i^2$
28,5 - 33,5	31	1	31	961
33,5 - 38,5	36	1	36	1 296
38,5 - 43,5	41	5	205	8 405
43,5 - 48,5	46	10	460	21 160
48,5 - 53,5	51	6	306	15 606
53,5 - 58,5	56	8	448	25 088
58,5 - 63,5	61	4	244	14 884
		35	1 730	87 400

b) Representamos los datos en un histograma. La altura de cada rectángulo coincidirá con la frecuencia absoluta, por ser los intervalos de igual amplitud.



c) MEDIA:  $\bar{x} = \frac{1730}{35} \approx 49,43$

$$\sigma^2 = \frac{87400}{35} - 49,43^2 = 53,82$$

DESVIACIÓN TÍPICA:  $\sigma = \sqrt{53,82} \approx 7,34$

### Mediana, cuartiles y percentiles

6 Las urgencias atendidas durante un mes en un centro de salud fueron:

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

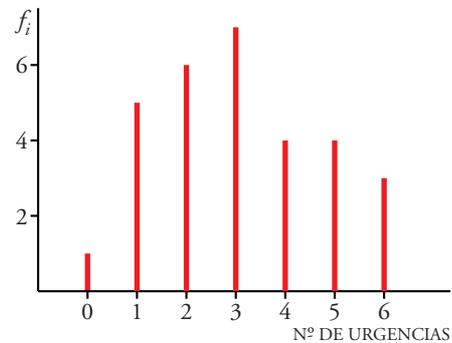
a) Haz una tabla de frecuencias y representa los datos.

b) Haz la tabla de frecuencias acumuladas y di cuál es la mediana.

a)

$x_i =$ urgencias atendidas	$f_i$	$F_i$	en %
0	1	1	3,33
1	5	6	20
2	6	12	40
3	7	19	63,33
4	4	23	76,67
5	4	27	90
6	3	30	100

Representamos los datos en un diagrama de barras:



b)  $Me = p_{50} = 3$  (para  $x_i = 3$ ,  $F_i$  supera el 50%)

**7** La altura, en centímetros, de un grupo de alumnos y alumnas de una misma clase es:

150, 169, 171, 172, 172, 175, 181

182, 183, 177, 179, 176, 184, 158

Calcula la mediana y los cuartiles y explica el significado de estos parámetros.

Colocamos los datos en orden creciente:

150 - 158 - 169 - 171 - 172 - 172 - 175 - 176 - 177 - 179 - 181 - 182 - 183 - 184

Hay 14 datos:

$\frac{14}{2} = 7 \rightarrow$  Mediana: valor intermedio de los dos centrales situados en séptima y octava posición:

$$Me = \frac{175 + 176}{2} = 175,5 \text{ cm}$$

Significa que la mitad de los alumnos tiene una estatura inferior a 175,5 cm.

$\frac{14}{4} = 3,5 \rightarrow$   $Q_1 = 171 \text{ cm}$  (4º lugar)

El 25% de los alumnos mide menos de 171 cm de altura.

$14 \cdot \frac{3}{4} = 10,5 \rightarrow$   $Q_3 = 181 \text{ cm}$  (posición 11)

El 75% de los alumnos tiene una estatura inferior a 181 cm.

**8** Calcula la mediana y los cuartiles de la siguiente distribución:

$x_i$	0	1	2	3	4	5
$f_i$	12	9	7	6	3	3

Completamos la tabla con las frecuencias acumuladas:

$x_i$	$f_i$	$F_i$	en %
0	12	12	30
1	9	21	52,5
2	7	28	70
3	6	34	85
4	3	37	92,5
5	3	40	100

- $Me = 1$ , porque para  $x_i = 1$  la  $F_i$  supera el 50%
- $Q_1 = 0$ , porque para  $F_i$  supera el 25% para  $x_i = 0$
- $Q_3 = 3$ , porque  $F_i$  supera el 75% para  $x_i = 3$

9 Halla la mediana, los cuartiles y el percentil 60 en cada una de las siguientes distribuciones, correspondientes a las notas obtenidas en un test que han hecho dos grupos de estudiantes:

$$\text{A: } 25 - 22 - 27 - 30 - 23 - 22 - 31 - 18$$

$$24 - 25 - 32 - 35 - 20 - 28 - 30$$

$$\text{B: } 27 - 32 - 19 - 22 - 25 - 30 - 21$$

$$29 - 23 - 31 - 21 - 20 - 18 - 27$$

Colocamos en orden creciente los datos:

$$\text{A } 18 - 20 - 22 - 22 - 23 - 24 - 25 - 25 - 27 - 28 - 30 - 30 - 31 - 32 - 35$$

Hay 15 datos:

- La mediana es el valor central (posición 8)  $\rightarrow Me = 25$
- $\frac{15}{4} = 3,75 \rightarrow Q_1 = 22$  (4ª posición)
- $15 \cdot \frac{3}{4} = 11,25 \rightarrow Q_3 = 30$  (12ª posición)
- $15 - \frac{60}{100} = 9 \rightarrow p_{60}$  será el valor intermedio de los datos situados en 9ª y 10ª posición, es decir:

$$p_{60} = \frac{27 + 28}{2} \rightarrow p_{60} = 27,5$$

$$\text{B } 18 - 19 - 20 - 21 - 21 - 22 - 23 - 25 - 27 - 27 - 29 - 30 - 31 - 32$$

Hay 14 datos:

- Los dos valores centrales son 23 y 25  $\rightarrow Me = \frac{23 + 25}{2} = 24$
- $\frac{14}{4} = 3,5 \rightarrow Q_1 = 21$  (4ª posición)
- $14 \cdot \frac{3}{4} = 10,5 \rightarrow Q_3 = 29$  (11ª posición)
- $14 \cdot \frac{60}{100} = 8,4 \rightarrow p_{60} = 27$  (9ª posición)

**Página 216****Muestras**

**10** Se quieren realizar los siguientes estudios:

- a) Tipo de transporte que utilizan los vecinos de un barrio para acudir a su trabajo.
- b) Estudios que piensan realizar los alumnos y alumnas de un centro escolar al terminar la ESO.
- c) Edad de las personas que han visto una obra de teatro en una ciudad.
- d) Número de horas diarias que ven la televisión los niños y niñas españoles con edades comprendidas entre 5 y 10 años.

Di en cada uno de estos casos cuál es la población.

¿En cuáles de ellos es necesario recurrir a una muestra? ¿Por qué?

- a) Población: todos los vecinos del barrio.

Es necesario recurrir a una muestra porque la población es grande y difícil de controlar.

- b) Población: todos los alumnos y alumnas del centro escolar.

Es necesario recurrir a una muestra. La población es grande.

- c) Población: personas que han visto la obra de teatro.

Es necesario recurrir a una muestra: la población es difícil de controlar.

- d) Población: niños y niñas españoles con edades entre 5 y 10 años.

Es necesario recurrir a una muestra. La población es muy grande.

**11** Queremos seleccionar una muestra de 50 alumnos de 3º de ESO.

En cada uno de los siguientes casos debes decidir si el muestreo debe ser aleatorio simple o estratificado por sexos (chicos-chicas) para estudiar las variables indicadas:

- a) Uso de pendientes en las orejas.

- b) Tiempo que emplean los alumnos en ir de su casa al colegio.

- c) Incidencia de caries dental.

- d) Práctica de fútbol.

- a) Estratificado por sexos.

- b) Aleatorio simple.

- c) Aleatorio simple.

- d) Estratificado por sexos.

- 12** Para hacer un estudio sobre los hábitos ecológicos de las familias de una ciudad, se han seleccionado por sorteo las direcciones, calle y número, que serán visitadas. Si en un portal vive más de una familia, se sorteará entre ellas la que será seleccionada. ¿Obtendremos con este procedimiento una muestra aleatoria?

 (Piensa si tiene la misma probabilidad de ser incluida en la muestra una familia que vive en una vivienda unifamiliar que otra que vive en un bloque de 32 viviendas).

No obtendremos una muestra aleatoria.

- 13** En un club deportivo que tiene 500 socios, se quiere elegir una muestra de 120 individuos. Los socios están agrupados por edades en tres grupos de 95, 125 y 280 individuos. ¿Cuántos se deben tomar de cada grupo de forma que el muestreo sea estratificado con reparto proporcional?

$$\text{Grupo de 95 individuos: } \frac{120}{500} = \frac{n_1}{95} \rightarrow n_1 \approx 23$$

$$\text{Grupo de 125 individuos: } \frac{120}{500} = \frac{n_2}{125} \rightarrow n_2 = 30$$

$$\text{Grupo de 280 individuos: } \frac{120}{500} = \frac{n_3}{280} \rightarrow n_3 \approx 67$$

- 14** Los empleados de una empresa están clasificados como indica esta tabla:

CATEGORÍA	A	B	C	D
Nº DE EMPLEADOS	400	300	200	100

Para hacer una consulta sobre la modificación del horario laboral, se elige por sorteo a 50 empleados de la categoría A, 40 de la B, 30 de la C y 20 de la D.

¿Es éste un modelo de muestreo aleatorio estratificado? ¿Es proporcional? ¿Por qué?

Sí es un modelo aleatorio estratificado, pero no es proporcional porque:

$$\frac{50}{400} \neq \frac{40}{300} \neq \frac{30}{200} \neq \frac{20}{100}$$

## PIENSA Y RESUELVE

- 15** Deseamos hacer una tabla con datos agrupados a partir de 384 datos, cuyos valores extremos son 19 y 187.

- Si queremos que sean 10 intervalos de amplitud 17, ¿cuáles serán esos intervalos?
- Haz otra distribución en 12 intervalos de la amplitud que creas conveniente.

Recorrido:  $r = 187 - 19 = 168$

- a) Buscamos un número algo mayor que el recorrido y que sea múltiplo de 10. Por ejemplo,  $r' = 170$ . De este modo, cada intervalo tendrá una longitud de 17.

Los intervalos son:

$$[18, 35); [35, 52); [52, 69); [69, 86); [86, 103); [103, 120) \\ [120, 137); [137, 154); [154, 171); [171, 188)$$

- b) Buscamos ahora un número que sea múltiplo de 12, que es el número de intervalos en este caso.

$$168 = 12 \cdot 14 \rightarrow \text{la amplitud de cada intervalo será } 14.$$

Los intervalos son:

$$[19, 33); [33, 47); [47, 61); [61, 75); [75, 89); [89, 103) \\ [103, 117); [117, 131); [131, 145); [145, 159); [159, 173); [173, 187)$$

- 16** Los gastos mensuales de una empresa A tienen una media de 100 000 euros y una desviación típica de 12 500 euros. En otra empresa B la media es 15 000 euros y la desviación típica 2 500 euros.

Calcula el coeficiente de variación y di cuál de las dos tiene mayor variación relativa.

$$\text{Empresa A: } \left. \begin{array}{l} \bar{x} = 100\,000 \text{ €} \\ \sigma = 12\,500 \text{ €} \end{array} \right\} \text{C.V.} = \frac{\sigma}{\bar{x}} = \frac{12\,500}{100\,000} = 0,125 \text{ o bien } 12,5\%$$

$$\text{Empresa B: } \left. \begin{array}{l} \bar{x} = 15\,000 \text{ €} \\ \sigma = 2\,500 \text{ €} \end{array} \right\} \text{C.V.} = \frac{2\,500}{15\,000} = 0,1\bar{6} \text{ o bien } 16,67\%$$

Tiene mayor variación relativa la empresa B.

- 17** El peso medio de los alumnos de una clase es 58,2 kg y su desviación típica 3,1 kg. El de las alumnas de esa clase es 52,4 kg y su desviación típica es 5,1 kg. Calcula el coeficiente de variación y compara la dispersión de ambos grupos.

$$\text{Alumnos } \left\{ \begin{array}{l} \bar{x} = 58,2 \text{ kg} \\ \sigma = 3,1 \text{ kg} \end{array} \right\} \text{C.V.} = \frac{3,1}{58,2} = 0,053 \rightarrow 5,3\%$$

$$\text{Alumnas } \left\{ \begin{array}{l} \bar{x} = 52,4 \text{ kg} \\ \sigma = 5,1 \text{ kg} \end{array} \right\} \text{C.V.} = \frac{5,1}{52,4} = 0,097 \rightarrow 9,7\%$$

El peso medio de las alumnas es más variable que el peso de los alumnos.

- 18** Se han medido los pesos y las alturas de 6 personas, obteniéndose los siguientes datos:

Calcula el coeficiente de variación y di si están más dispersos los pesos o las alturas.

PESOS ( $x_i$ )	$f_i$	$f_i x_i$	$f_i x_i^2$
60	1	60	3 600
63	2	126	7 938
65	1	65	4 225
68	2	136	9 248
	6	387	25 011

PESOS (kg)	ALTURAS (m)
65	1,7
60	1,5
63	1,7
63	1,7
68	1,75
68	1,8

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{387}{6} = 64,5 \text{ kg}$$

$$\sigma^2 = \frac{\sum f_i x_i^2}{\sum f_i} - \bar{x}^2 = \frac{25\,011}{6} - 64,5^2 = 8,25 \rightarrow \sigma = \sqrt{8,25} = 2,87 \text{ kg}$$

$$\text{C.V.} = \frac{\sigma}{\bar{x}} = \frac{2,87}{64,5} = 0,044 \text{ o bien } 4,4\%$$

ALTURAS ( $y_i$ )	$f_i$	$f_i y_i$	$f_i y_i^2$
1,5	1	1,5	2,25
1,7	3	5,1	8,67
1,75	1	1,75	3,06
1,8	1	1,8	3,24
	6	10,15	17,22

$$\bar{y} = \frac{\sum f_i y_i}{\sum f_i} = \frac{10,15}{6} = 1,69 \text{ m}$$

$$\sigma^2 = \frac{\sum f_i y_i^2}{\sum f_i} - \bar{y}^2 = \frac{17,22}{6} - 1,69^2 = 0,0139 \rightarrow \sigma = \sqrt{0,0139} = 0,12 \text{ m}$$

$$\text{C.V.} = \frac{\sigma}{\bar{y}} = \frac{0,12}{1,69} = 0,071 \text{ o bien } 7,1\%$$

Están más dispersas las alturas que los pesos.

### Página 217

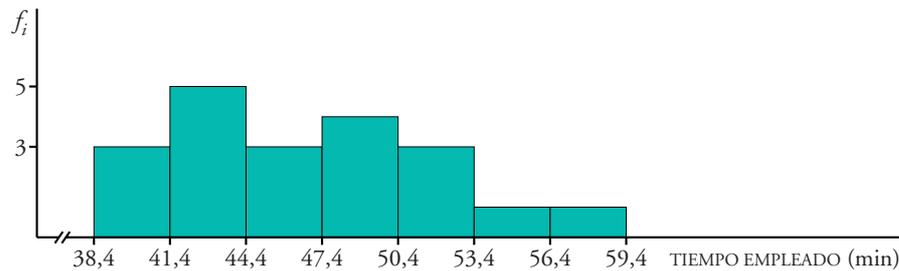
- 19** En una regata de veleros, los tiempos, en minutos, empleados por los participantes en hacer el primer recorrido han sido:

43,7	52	49,5	47,3	42,5
51,6	50,2	48,4	39,8	40,6
41,2	41,8	44	54	45,2
46,4	42,8	49	50,8	58

- a) Representa gráficamente los datos.  
 b) Calcula  $\bar{x}$  y  $\sigma$ .  
 c) Ordena los datos y calcula la mediana y los cuartiles.
- a) El número de valores distintos que hay es grande; luego, es adecuado agruparlos en intervalos.

$$\text{Recorrido} = 58 - 39,8 = 18,2$$

Tomamos 7 intervalos de amplitud 3.



INTERVALO	f <sub>i</sub>	MARCA (x <sub>i</sub> )	f <sub>i</sub> x <sub>i</sub>	f <sub>i</sub> x <sub>i</sub> <sup>2</sup>
38,4 - 41,4	3	39,9	119,7	4 776,03
41,4 - 44,4	5	42,9	214,5	9 202,05
44,4 - 47,4	3	45,9	137,7	6 320,43
47,4 - 50,4	4	48,9	195,6	9 564,84
50,4 - 53,4	3	51,9	155,7	8 080,83
53,4 - 56,4	1	54,9	54,9	3 014,01
56,4 - 59,4	1	57,9	57,9	3 352,41
	20		936	44 310,6

$$\text{MEDIA} \rightarrow \bar{x} = \frac{936}{20} = 46,8 \text{ minutos}$$

$$\sigma^2 = \frac{44\,310,6}{20} - 46,8^2 = 25,29$$

$$\text{DESVIACIÓN TÍPICA} \rightarrow \sigma = \sqrt{25,29} = 5,03 \text{ minutos}$$

- c) Ordenamos los datos:

39,8 - 40,6 - 41,2 - 41,8 - 42,5 - 42,8 - 43,7 - 44 - 45,2 - 46,4

47,3 - 48,4 - 49 - 49,5 - 50,2 - 50,8 - 51,6 - 52 - 54 - 58

Hay 20 datos:

$\frac{20}{2} = 10 \rightarrow$  La mediana es el valor intermedio de los valores situados en las posiciones 10 y 11:

$$Me = \frac{46,4 + 47,3}{2} = 46,85$$

$\frac{20}{4} = 5 \rightarrow Q_1$  es la media aritmética de 42,5 y 42,8, valores situados en 5ª y 6ª posición:

$$Q_1 = \frac{42,5 + 42,8}{2} = 42,65$$

$20 \cdot \frac{3}{4} = 15 \rightarrow Q_3$  es el valor intermedio entre 50,2 y 50,8, valores que ocupan la posición 15 y 16, respectivamente:

$$Q_3 = \frac{50,2 + 50,8}{2} = 50,5$$

**20** El número de errores cometidos en un test por un grupo de personas viene reflejado en la siguiente tabla:

Nº DE ERRORES	0	1	2	3	4	5	6
Nº DE PERSONAS	10	12	8	7	5	4	3

- a) Halla la mediana y los cuartiles inferior y superior, y explica su significado.  
b) ¿Cuál es el número medio de errores por persona?

Construimos la tabla de frecuencias acumuladas:

Nº DE ERRORES ( $x_i$ )	Nº DE PERSONAS ( $f_i$ )	$x_i f_i$	$F_i$	EN %
0	12	0	12	23,53
1	12	12	24	47,06
2	8	16	32	62,75
3	7	21	39	76,47
4	5	20	44	86,27
5	4	20	48	94,12
6	3	18	51	100
	51	107		

- a)  $Me = 2$ . Significa que el 50% de las personas cometen 0, 1 ó 2 errores.  
 $Q_1 = 1$ . El 25% de las personas comete 1 error o ninguno.  
 $Q_3 = 3$ . El 75% de las personas comente 3 errores o menos de 3 errores.

$$b) \bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{107}{51} \approx 2,1$$

El número medio de errores por persona es ligeramente superior a 2.

**21** Al preguntar a un grupo de personas cuánto tiempo dedicaron a ver televisión durante un fin de semana, se obtuvieron estos resultados:

Dibuja el histograma correspondiente y halla la media y la desviación típica.

Como los intervalos no son de la misma longitud, para representar la distribución mediante un histograma pondremos en cada barra una altura tal que el área sea proporcional a la frecuencia:

TIEMPO (en horas)	NÚMERO DE PERSONAS
[0; 0,5)	10
[0,5; 1,5)	10
[1,5; 2,5)	18
[2,5; 4)	12
[4; 8)	12

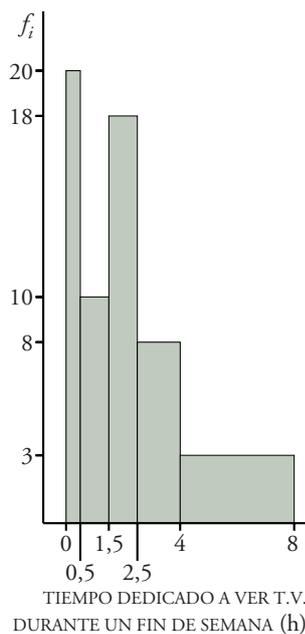
$$[0; 0,5) \rightarrow a_1 = 0,5 \quad f_1 = 10 \rightarrow h_1 = \frac{10}{0,5} = 20$$

$$[0,5; 1,5) \rightarrow a_2 = 1 \quad f_2 = 10 \rightarrow h_2 = 10$$

$$[1,5; 2,5) \rightarrow a_3 = 1 \quad f_3 = 18 \rightarrow h_3 = 18$$

$$[2,5; 4) \rightarrow a_4 = 1,5 \quad f_4 = 12 \rightarrow h_4 = \frac{12}{1,5} = 8$$

$$[4; 8) \rightarrow a_5 = 4 \quad f_5 = 12 \rightarrow h_5 = \frac{12}{4} = 3$$



TIEMPO	MARCA ( $x_i$ )	$f_i$	$f_i x_i$	$f_i x_i^2$
[0; 0,5)	0,25	10	2,5	0,625
[0,5; 1,5)	1	10	10	10
[1,5; 2,5)	2	18	36	72
[2,5; 4)	3,25	12	39	126,75
[4; 8)	6	12	72	432
		62	159,5	641,375

$$\bar{x} = \frac{159,5}{62} = 2,57$$

$$\sigma^2 = \frac{641,375}{62} - 2,57^2 = 3,74 \rightarrow$$

$$\rightarrow \sigma = \sqrt{3,74} = 1,93$$

**22** Un dentista observa el número de caries en cada uno de los 100 niños de un colegio y obtiene los resultados resumidos en esta tabla:

NÚMERO DE CARIES	0	1	2	3	4
FRECUENCIA ABSOLUTA	25	20	$y$	15	$x$
FRECUENCIA RELATIVA	0,25	0,2	$z$	0,15	0,05

a) Completa la tabla obteniendo  $x$ ,  $y$ ,  $z$ .

b) Calcula el número medio de caries.

- a) La frecuencia relativa es el cociente entre la frecuencia absoluta y el número total de individuos (100, en nuestro caso).

$$0,05 = \frac{x}{100} \rightarrow x = 5$$

$$25 + 20 + y + 15 + 5 = 100 \rightarrow y = 35$$

$$z = \frac{y}{100} = \frac{35}{100} \rightarrow z = 0,35$$

b)

Nº DE CARIES ( $x_i$ )	$f_i$	$f_i x_i$
0	25	0
1	20	20
2	35	70
3	15	45
4	5	20
	100	155

$$\bar{x} = \frac{155}{100} = 1,55$$

El número medio de caries es de 1,55.

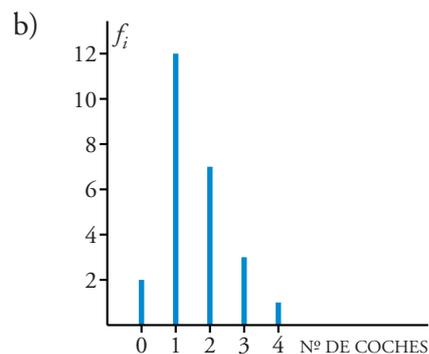
- 23** En una población de 25 familias se ha observado la variable  $X =$  “número de coches que tiene la familia” y se han obtenido los siguientes datos:

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

- a) Construye la tabla de frecuencias de la distribución  $X$ .  
 b) Haz el diagrama de barras.  
 c) Calcula la media y la desviación típica.  
 d) Halla la mediana y los cuartiles.

a)

$x_i$	$f_i$	$f_i x_i$	$f_i x_i^2$	$F_i$	en %
0	2	0	0	2	8
1	12	12	12	14	56
2	7	14	28	21	84
3	3	9	27	24	96
4	1	4	16	25	100
	25	39	83		

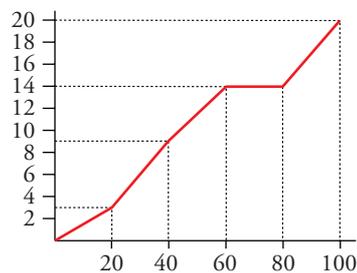


$$c) \bar{x} = \frac{39}{25} = 1,56$$

$$\sigma^2 = \frac{83}{25} - 1,56^2 = 0,89 \rightarrow \sigma = \sqrt{0,89} \approx 0,94$$

$$d) Me = 1, Q_1 = 1 \text{ y } Q_3 = 2$$

**24** Este es el polígono de frecuencias acumuladas correspondiente a una distribución de datos agrupados en intervalos:



a) Escribe la tabla de frecuencias absolutas.

b) Calcula la media y la desviación típica de la distribución.

a)

INTERVALOS	$F_i$	$f_i$	MARCA DE CLASE ( $x_i$ )	$f_i x_i$	$f_i x_i^2$
[0, 20)	3	3	10	30	300
[20, 40)	9	6	30	180	5 400
[40, 60)	14	5	50	250	12 500
[60, 80)	14	0	70	0	0
[80, 100)	20	6	90	540	48 600
		20		1 000	66 800

$$b) \text{ MEDIA } \rightarrow \bar{x} = \frac{1\,000}{20} = 50$$

$$\text{VARIANZA } \rightarrow \sigma^2 = \frac{66\,800}{20} - 50^2 = 840$$

$$\text{DESVIACIÓN TÍPICA } \rightarrow \sigma = \sqrt{840} \approx 28,98$$

**25** Completa la siguiente tabla estadística, donde  $f$ ,  $F$  y  $f_r$  representan, respectivamente, la frecuencia absoluta, la frecuencia absoluta acumulada y la frecuencia relativa.

 Recuerda que  $f_r = f/n$  y calcula  $n$ .

De la primera fila se obtiene fácilmente el valor de  $n$ :

$$fr = \frac{f}{n} \rightarrow 0,08 = \frac{4}{n} \rightarrow n = 50$$

$x$	$f$	$F$	$fr$
1	4	4	0,08
2	4	8	0,08
3	8	16	0,16
4	7	23	0,14
5	5	28	0,1
6	10	38	0,2
7	7	45	0,14
8	5	50	0,1

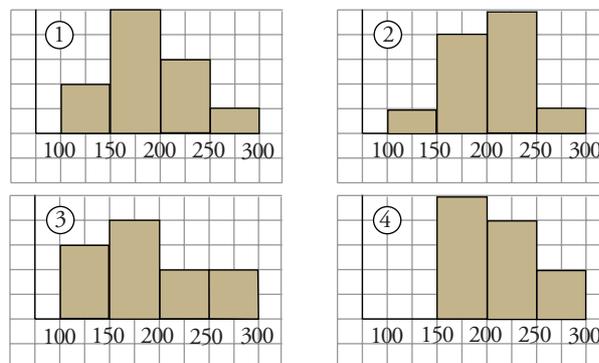
### Página 218

**26** Se ha medido el nivel de colesterol en cuatro grupos de personas sometidas a diferentes dietas.

Las medias y las desviaciones típicas son las que figuran en esta tabla:

DIETA	A	B	C	D
$\bar{x}$	211,3	188,6	202,2	185
$\sigma$	37,4	52,6	39,1	43,6

Las gráficas son, no respectivamente:



Asocia a cada dieta la gráfica que le corresponde.

Fijándonos en las gráficas, se observa que los grupos 1 y 3 tienen una media inferior a 200, mientras que las medias de 2 y 4 son superiores a ese número. Luego podemos asociar:

$$A \text{ y } C \rightarrow 2 \text{ y } 4$$

$$B \text{ y } D \rightarrow 1 \text{ y } 3$$

Por otra parte, las personas de 2 tienen el nivel de colesterol más disperso que las de 4. Según esto, su desviación típica será mayor, por lo que  $C \leftrightarrow 2$  y  $A \leftrightarrow 4$ . Análogamente,  $B \leftrightarrow 3$  y  $D \leftrightarrow 1$ .

## REFLEXIONA SOBRE LA TEORÍA

**27** Justifica que la suma de las frecuencias relativas es siempre igual a 1.

Supongamos que tenemos  $n$  datos:

$$fr_1 + fr_2 + \dots + fr_n = \frac{f_1}{n} + \frac{f_2}{n} + \dots + \frac{f_n}{n} = \frac{f_1 + f_2 + \dots + f_n}{n} = \frac{n}{n} = 1$$

Siendo  $f_i$  la frecuencia absoluta del dato  $x_i$ .

**28** En la distribución de las notas de un examen, el primer cuartil fue 4. ¿Qué significa esto?

$Q_1 = 4$  significa que el 25% de los alumnos obtuvieron una nota igual o inferior a 4.

**29** Completa la tabla de esta distribución en la que sabemos que su media es 2,7.

$x_i$	1	2	3	4
$f_i$	3	...	7	5

Llamamos  $z$  a la frecuencia absoluta del dato  $x_i = 2$ .

Aplicamos la definición de la media:

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} \rightarrow 2,7 = \frac{3 + 2z + 21 + 20}{15 + z}$$

$$2,7 \cdot (15 + z) = 44 + 2z$$

$$40,5 + 2,7z = 44 + 2z \rightarrow 0,7z = 3,5 \rightarrow z = 5$$

**30** Si a todos los datos de una distribución le sumamos un mismo número, ¿qué le ocurre a la media? ¿Y a la desviación típica? ¿Y si multiplicamos todos los datos por un mismo número?

Llamamos  $a$  al valor sumado a cada dato de la distribución:

• MEDIA

$$\begin{aligned} & \frac{(x_1 + a)f_1 + (x_2 + a)f_2 + \dots + (x_n + a)f_n}{n} = \\ & = \frac{x_1 f_1 + x_2 f_2 + \dots + x_n f_n + a(f_1 + f_2 + \dots + f_n)}{n} = \\ & = \frac{\sum f_i x_i}{n} + a \frac{\sum f_i}{n} = \bar{x} + a, \text{ puesto que } \frac{\sum f_i}{n} = \frac{n}{n} = 1 \end{aligned}$$

La nueva media es el valor de la media original más el valor que hemos sumado a cada dato.

- DESVIACIÓN TÍPICA:

$$\begin{aligned} \frac{\sum f_i (x_i + a)^2}{\sum f_i} - (\bar{x} + a)^2 &= \frac{\sum f_i x_i^2 + \sum f_i a^2 + \sum f_i 2x_i a}{\sum f_i} - \bar{x}^2 - a^2 - 2\bar{x}a = \\ &= \frac{\sum f_i x_i^2}{\sum f_i} + a^2 + 2a\bar{x} - \bar{x}^2 - a^2 - 2\bar{x}a = \\ &= \frac{\sum f_i x_i^2}{\sum f_i} - \bar{x}^2 \end{aligned}$$

La desviación típica no se ve alterada al sumar a todos los datos de la distribución un mismo número.

Supongamos ahora que todos los datos se multiplican por un mismo valor  $a$ :

- MEDIA =  $\frac{ax_1f_1 + ax_2f_2 + \dots + ax_nf_n}{n} = a\bar{x} \rightarrow$  la media queda multiplicada por dicho valor.

- DESVIACIÓN TÍPICA:

$$\frac{\sum f_i (x_i a)^2}{\sum f_i} - (\bar{x}a)^2 = \frac{a^2 \sum f_i x_i^2}{\sum f_i} - a^2 \bar{x}^2 = a^2 \left( \frac{\sum f_i x_i^2}{\sum f_i} - \bar{x}^2 \right)$$

La varianza quedaría multiplicada por  $a^2$ , luego la desviación típica queda multiplicada por  $a$ .

### 31 Dos distribuciones estadísticas, $A$ y $B$ , tienen la misma desviación típica.

- Si la media de  $A$  es mayor que la de  $B$ , ¿cuál tiene mayor coeficiente de variación?
- Si la media de  $A$  es doble que la de  $B$ , ¿cómo serán sus coeficientes de variación?

$$\text{C.V.} = \frac{\sigma}{\bar{x}}$$

$$\text{a) Si } \bar{x}_A > \bar{x}_B \rightarrow \frac{1}{\bar{x}_A} < \frac{1}{\bar{x}_B} \xrightarrow{\sigma > 0} \frac{\sigma}{\bar{x}_A} < \frac{\sigma}{\bar{x}_B} \rightarrow B \text{ tiene mayor coeficiente}$$

de variación.

$$\text{b) Si } \bar{x}_A = 2\bar{x}_B$$

$$\text{C.V. de } A \rightarrow \frac{\sigma}{\bar{x}_A} = \frac{\sigma}{2\bar{x}_B}$$

$$\text{C.V. de } B \rightarrow \frac{\sigma}{\bar{x}_B}$$

El coeficiente de variación de  $A$  es la mitad que el de  $B$ .

**32** Observa esta demostración:

$$\begin{aligned} \frac{\sum (x_i - \bar{x})^2}{n} &= \frac{\sum (x_i^2 - 2x_i\bar{x} + \bar{x}^2)}{n} \quad \textcircled{1} \\ &= \frac{\sum x_i^2}{n} - 2\bar{x}^2 + \frac{\sum \bar{x}^2}{n} \quad \textcircled{2} \\ &= \frac{\sum x_i^2}{n} - 2\bar{x}^2 + \bar{x}^2 = \frac{\sum \bar{x}^2}{n} - \bar{x}^2 \end{aligned}$$

Justifica las igualdades ① y ②.

• Justificación de ①  $\rightarrow \frac{\sum (x_i^2 - 2x_i\bar{x} + \bar{x}^2)}{n} = \frac{\sum x_i^2}{n} - \frac{2\bar{x}\sum x_i}{n} + \frac{\sum \bar{x}^2}{n}$

Como  $\frac{\sum x_i}{n} = \bar{x}$ , entonces se obtiene:  $\frac{\sum x_i^2}{n} - 2\bar{x}^2 + \frac{\sum \bar{x}^2}{n}$

• Justificación de ②  $\rightarrow \frac{\sum \bar{x}^2}{n} = \frac{n\bar{x}^2}{n} = \bar{x}^2$  (por ser  $n$  datos con frecuencia 1)

## PROFUNDIZA

**33** En una fábrica se ha medido la longitud de 1000 piezas de las mismas características y se han obtenido estos datos:

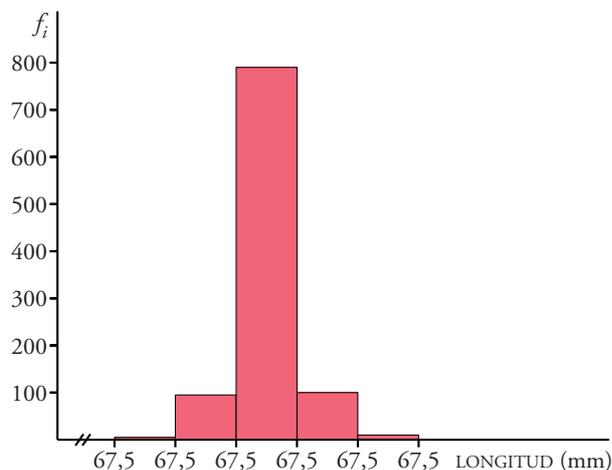
LONGITUD (en mm)	NÚMERO DE PIEZAS
67,5-72,5	5
72,5-77,5	95
77,5-82,5	790
82,5-87,5	100
87,5-92,5	10

a) Representa el histograma correspondiente.

b) Se consideran aceptables las piezas cuya longitud está en el intervalo  $[75, 86]$ . ¿Cuál es el porcentaje de piezas defectuosas?

 Del segundo intervalo habrá que rechazar las que midan entre 72,5 y 75. Calcula qué tanto por ciento de la amplitud representa la diferencia 75-72,5 y halla el porcentaje de la frecuencia correspondiente. Proceda análogamente en el cuarto intervalo.

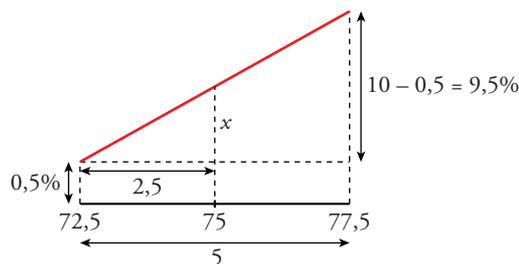
a) Por tener todos los intervalos la misma longitud, la altura de cada una de las barras coincidirá con la frecuencia de cada intervalo.



b) Construimos la tabla de frecuencias absolutas acumuladas:

INTERVALO	$f_i$	$F_i$	en %
67,5 - 72,5	5	5	0,5
72,5 - 77,5	95	100	10
77,5 - 82,5	790	890	89
82,5 - 87,5	100	990	99
87,5 - 92,5	10	1 000	100

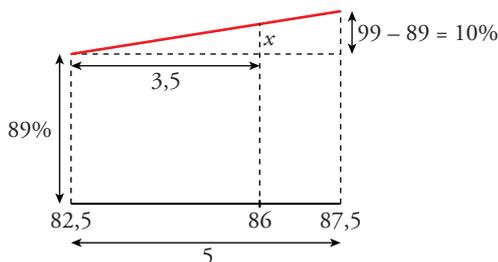
• Calculamos el porcentaje de piezas que hay por debajo de 75 mm:



$$\frac{9,5}{5} = \frac{x}{2,5} \rightarrow x = 4,75$$

Por debajo de 75 mm están el  
4,75 + 0,5 = 5,25% de las piezas.

• Calculamos el porcentaje de piezas que están por debajo de 86 mm:



$$\frac{10}{5} = \frac{x}{3,5} \rightarrow x = 7$$

Por debajo de 86 mm están el  
89 + 7 = 96% de las piezas.

El porcentaje de piezas que hay en el intervalo [75, 86] es:

$$96 - 5,25 = 90,75\%$$

Por tanto, el  $100 - 90,75 = 9,25\%$  de las piezas serán defectuosas.

**34** En un sondeo de opinión entre los jóvenes españoles de 15 a 24 años, una de las preguntas era: *¿Justificas que alguien acepte un soborno en su trabajo?*

Respuesta  $\left\{ \begin{array}{l} \text{Nunca} \rightarrow 1 \quad \text{A veces} \rightarrow 5 \\ \text{Siempre} \rightarrow 10 \end{array} \right.$

En una muestra de 2 000 individuos se obtuvo una puntuación de 2,63.

a) Expresa este resultado sabiendo que en la ficha técnica se dice que el error máximo es de  $\pm 1,22$  con un nivel de confianza del 95%.

b) Si el error máximo fuera  $\pm 0,6$ , ¿el nivel de confianza sería mayor o menor?

a) La puntuación media estará en el intervalo:

$$(2,63 - 1,22; 2,63 + 1,22) = (1,41; 3,85)$$

con un nivel de confianza del 95%.

- b) Con el nuevo error, el intervalo será (2,03; 3,23). Por tanto, el nivel de confianza será menor que el 95% anterior.

### Página 219

- 35** a) Para estimar la estatura media de los 934 soldados de un regimiento, extraemos una muestra de 53 de ellos. La media de la muestra es 172,6 cm. Expresa este resultado sabiendo que en la ficha técnica se dice que el error máximo es de  $\pm 1,8$  cm, con una probabilidad de 0,90.
- b) Si con el mismo estudio anterior admitimos que se cometa un error de  $\pm 2,6$  cm, el nivel de confianza ¿será superior o inferior al 90%?
- c) ¿Cómo podríamos aumentar el nivel de confianza manteniendo la cota de error en  $\pm 1,8$  cm?
- a) La estatura media de los soldados de ese regimiento estará en el intervalo
- $$(172,6 - 1,8; 172,6 + 1,8) = (170,8; 174,4)$$
- con una probabilidad de 0,90.
- b) Ahora el intervalo es (170; 175,2). Como es mayor que el anterior, el nivel de confianza aumenta.
- c) Aumentando la muestra de 53 soldados.

- 36** De una muestra de 75 pilas eléctricas, se han obtenido estos datos sobre su duración:

TIEMPO (en horas)	NÚMERO DE PILAS
[25, 30)	3
[30, 35)	5
[35, 40)	21
[40, 45)	28
[45, 55)	12
[55, 70)	6

- a) Representa los datos gráficamente.
- b) Calcula la media y la desviación típica.
- c) ¿Qué porcentaje de pilas hay en el intervalo  $(\bar{x} - \sigma, \bar{x} + \sigma)$ ?

- a) Representamos los datos en un histograma. Como los intervalos son de distinta longitud, calculamos la altura ( $h_i$ ) de cada barra, sabiendo que el área de cada rectángulo ha de ser proporcional a la frecuencia ( $f_i$ ).

$$[25, 30) \rightarrow h_1 = \frac{3}{5} = 0,6$$

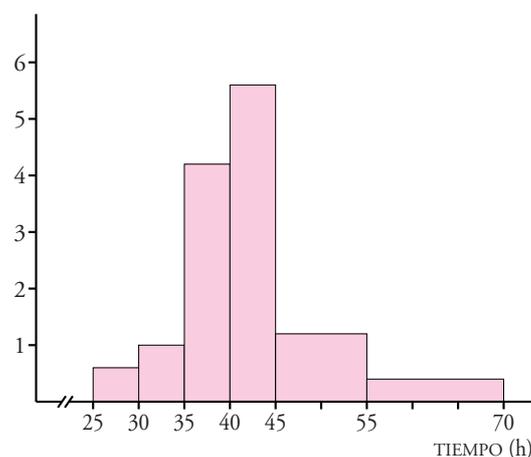
$$[30, 35) \rightarrow h_2 = \frac{5}{5} = 1$$

$$[35, 40) \rightarrow h_3 = \frac{21}{5} = 4,2$$

$$[40, 45) \rightarrow h_4 = \frac{28}{5} = 5,6$$

$$[45, 55) \rightarrow h_5 = \frac{12}{10} = 1,2$$

$$[55, 70) \rightarrow h_6 = \frac{6}{15} = 0,4$$



b)

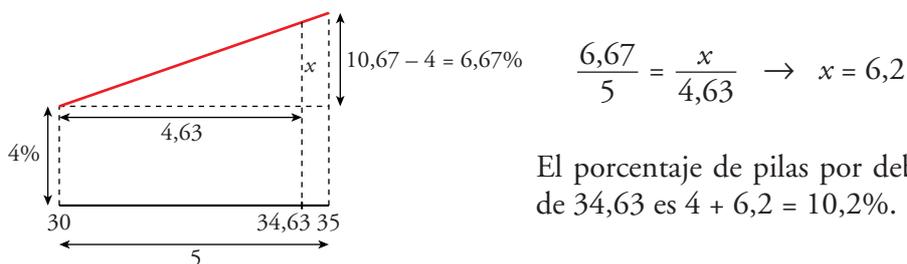
INTERVALO	MARCA ( $x_i$ )	$f_i$	$f_i x_i$	$f_i x_i^2$	$F_i$	EN %
[25, 30)	27,5	3	82,5	2 268,75	3	4
[30, 35)	32,5	5	162,5	5 281,25	8	10,67
[35, 40)	37,5	21	787,5	29 531,25	29	38,67
[40, 45)	42,5	28	1 190	50 575	57	76
[45, 55)	50	12	600	30 000	69	92
[55, 70)	62,5	6	375	23 437,5	75	100
		75	3 197,5	141 093,75		

$$\text{MEDIA} \rightarrow \bar{x} = \frac{3\,197,5}{75} = 42,63$$

$$\sigma^2 = \frac{141\,093,75}{75} - 42,63^2 = 63,93 \rightarrow \text{DESVIACIÓN TÍPICA: } \sigma = \sqrt{63,93} \approx 8$$

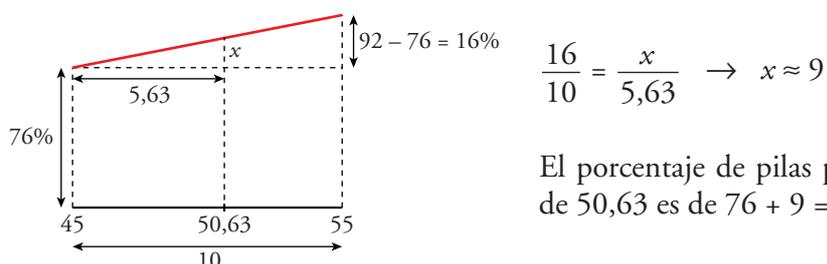
$$c) (\bar{x} - \sigma, \bar{x} + \sigma) = (42,63 - 8; 42,63 + 8) = (34,63; 50,63)$$

- Calculamos el porcentaje de pilas por debajo de 34,63:



El porcentaje de pilas por debajo de 34,63 es  $4 + 6,2 = 10,2\%$ .

- Calculamos el porcentaje de pilas por debajo de 50,63:



El porcentaje de pilas por debajo de 50,63 es de  $76 + 9 = 85\%$

El porcentaje de pilas que hay en el intervalo (34,63; 50,63) es:

$$85 - 10,2 = 74,8\%$$

**37** En el proceso de fabricación de un vino, se le añade un compuesto químico. Se ha comprobado la concentración de este compuesto en una partida de 200 botellas y se han obtenido los datos de la tabla.

- a) Calcula la media y la desviación típica.
- b) Se estima que el vino no se debe consumir si la concentración de ese compuesto es superior a 20,9 mg/l.

Según esto, ¿qué porcentaje de botellas no es adecuado para el consumo?

CONCENT. (mg/l)	NÚMERO DE BOTELLAS
[20; 20,2)	15
[20,2; 20,4)	38
[20,4; 20,6)	76
[20,6; 20,8)	57
[20,8; 21)	14

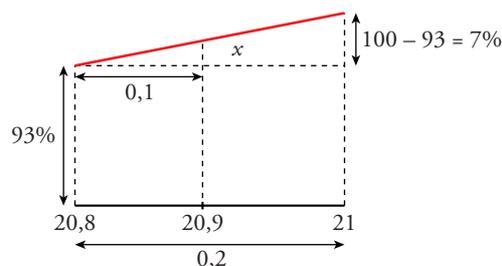
a)

INTERVALO	MARCA ( $x_i$ )	$f_i$	$f_i x_i$	$f_i x_i^2$	$F_i$	EN %
[20; 20,2)	20,1	15	301,5	6 060,15	15	7,5
[20,2; 20,4)	20,3	38	771,4	15 659,42	53	26,5
[20,4; 20,6)	20,5	76	1 558	31 939	129	64,5
[20,6; 20,8)	20,7	57	1 179,9	24 423,93	186	93
[20,8; 21)	20,9	14	292,6	6 115,34	200	100
		200	4 103,4	84 197,84		

$$\text{MEDIA} \rightarrow \bar{x} = \frac{4\,103,4}{200} = 20,52$$

$$\sigma^2 = \frac{84\,197,84}{200} - 20,517^2 = 0,04 \rightarrow \sigma = \sqrt{0,04} = 0,2$$

- b) Calculamos el porcentaje de botellas cuya concentración es inferior a 20,9 mg/l:



$$\frac{7}{0,2} = \frac{x}{0,1} \rightarrow x = 3,5$$

El porcentaje de botellas cuya concentración es inferior a 20,9 mg/l es de  $93 + 3,5 = 96,5\%$ .

No es adecuado para el consumo el  $100 - 96,5 = 3,5\%$  de botellas.

**38** Solo uno de los siguientes procedimientos nos permite obtener una muestra aleatoria. Di cuál es y, en los otros, estudia el sentido del sesgo y su importancia:

- a) Para estudiar las frecuencias relativas de las letras, se toman al azar 20 libros de la biblioteca de un centro escolar y se cuenta las veces que aparece cada letra en la página 20 de los libros seleccionados.
- b) Para conocer la opinión de sus clientes sobre el servicio ofrecido por unos grandes almacenes, se selecciona al azar, entre los que poseen tarjeta de compra, 100 personas entre las que han gastado menos de 1 000 € el últi-

mo año, otras 100 entre las que han gastado entre 1 000 € y 5 000 € y 100 más entre las que han gastado más de 5 000 €.

- c) Para calcular el número medio de personas por cartilla en un Centro de Salud de la Seguridad Social, los médicos toman nota de las cartillas de las personas que acuden a las consultas durante un mes.

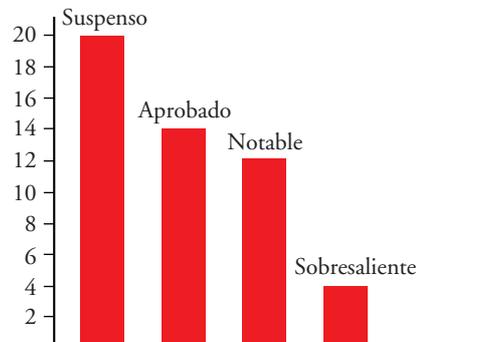
La opción buena es la a).

La b) no es válida, porque solo se tiene en cuenta a los clientes que poseen tarjeta de compra, que no son representativos de todos los clientes.

La c) no es válida, porque existen grupos en la sociedad que acuden más habitualmente al médico y no representan a toda ella.

- 39** Este diagrama de barras muestra las calificaciones obtenidas por un grupo de 50 estudiantes:

- a) Construye el histograma correspondiente a las calificaciones numéricas teniendo en cuenta las siguientes equivalencias: Suspenso, [0, 5); Aprobado, [5, 7); Notable, [7, 9); Sobresaliente, [9, 10).



- b) Calcula la calificación media.

Ten en cuenta que los intervalos no tienen la misma amplitud y que las áreas de los rectángulos deben ser proporcionales a las frecuencias.

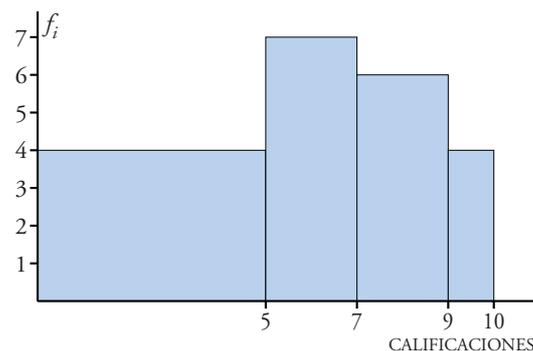
- a) Calculamos las alturas respectivas de cada rectángulo del histograma:

$$[0,5) \rightarrow h_1 = \frac{20}{5} = 4$$

$$[5,7) \rightarrow h_2 = \frac{14}{2} = 7$$

$$[7,9) \rightarrow h_3 = \frac{12}{2} = 6$$

$$[9,10) \rightarrow h_4 = \frac{4}{1} = 4$$



b)

INTERVALO	$f_i$	MARCA ( $x_i$ )	$f_i x_i$
[0, 5)	20	2,5	50
[5, 7)	14	6	84
[7, 9)	12	8	96
[9, 10)	4	9,5	38
	50		268

$$\bar{x} = \frac{268}{50} = 5,36$$