

El valor de las muestras. Un curioso experimento

¿Quiénes crees que obtendrán mejores resultados? Es decir, en general las medidas de las muestras de A, ¿se parecerán más a μ que las de B o al revés?

Como veremos, los del equipo B, que eligen las muestras aleatoriamente, se aproximan (en general) más a μ que los del equipo A.

Resuelve

1. Elige varias muestras de 10 números concienzudamente (es decir, observando los números seleccionados con el fin de que sean representativos de la totalidad).

Selecciona varias muestras aleatoriamente (es decir, diez de estos números elegidos al azar).

Compara los resultados calculando la media de cada muestra y comparándola con la verdadera media: $\bar{x} = 603,4$.

- Posibles muestras de 10 números elegidos concienzudamente:

① 2 200	② 4 553	③ 3 700	
37	2 830	2 211	
193	2 001	122	
4 780	400	418	
3 259	1	4 902	$\mu_1 = 1 336,7$
76	70	72	$\mu_2 = 1 259,1$
1 532	184	290	$\mu_3 = 1 323,3$
490	923	1 187	
789	1 584	317	
11	45	14	

- Posibles muestras de 10 números aleatorios:

① 4 001	② 456	③ 1	
401	188	311	
120	130	111	
4 850	17	8	
146	3 820	2 343	$\mu_1 = 1 541,9$
220	97	215	$\mu_2 = 1 181,7$
131	2 001	147	$\mu_3 = 681,4$ (la más próxima
190	5 000	19	a la verdadera)
370	33	500	
4 990	75	3 159	

2 Tablas de frecuencias

Página 195

- 1. Reparte los cuarenta datos del ejercicio resuelto anterior en 10 intervalos con el mismo recorrido total.**

Tomando $r' = 30$ y siendo 10 el número de intervalos, la longitud de cada intervalo será de $\frac{30}{10} = 3$.

INTERVALOS	MARCA DE CLASE	FRECUENCIAS
148,5 - 151,5	150	2
151,5 - 154,5	153	1
154,5 - 157,5	156	1
157,5 - 160,5	159	6
160,5 - 163,5	162	7
163,5 - 166,5	165	9
166,5 - 169,5	168	6
169,5 - 172,5	171	3
172,5 - 175,5	174	4
175,5 - 178,5	177	1

- 2. Reparte los cuarenta datos del ejercicio resuelto anterior en 8 intervalos. Para ello, toma $r' = 32$.**

Tomando $r' = 32$ y siendo 8 el número de intervalos, la longitud de cada uno de ellos será $\frac{32}{8} = 4$.

INTERVALOS	MARCA DE CLASE	FRECUENCIAS
147,5 - 151,5	149,5	2
151,5 - 155,5	153,5	1
155,5 - 159,5	157,5	4
159,5 - 163,5	161,5	10
163,5 - 167,5	165,5	12
167,5 - 171,5	169,5	6
171,5 - 175,5	173,5	4
175,5 - 179,5	177,5	1

3 Parámetros estadísticos: \bar{x} y σ

Página 197

1. Halla, manualmente y con calculadora, \bar{x} , σ y C.V. en la tabla obtenida en el ejercicio resuelto de la página 195:

x_j	151	156	161	166	171	176
f_j	2	4	11	14	5	4

x_j	f_j	$f_j x_j$	$f_j x_j^2$
151	2	302	45 602
156	4	624	97 344
161	11	1 771	285 131
166	14	2 324	385 784
171	5	855	146 205
176	4	704	123 904
	40	6 580	1 083 970

$$\bar{x} = \frac{6\,580}{40} = 164,5$$

$$\sigma = \sqrt{\frac{1\,083\,970}{40} - 164,5^2} = 6,24$$

$$\text{C.V.} = \frac{6,24}{164,5} = 0,038 \rightarrow 3,8\%$$

2. Halla, manualmente y con calculadora, \bar{x} , σ y C.V. en la distribución de los ejercicios 1 y 2 de la página 195:

Compara los resultados entre sí y con los del ejercicio 1 de esta página.

1.ª distribución

INTERVALOS	x_j	f_j	$f_j x_j$	$f_j x_j^2$
148,5-151,5	150	2	300	45 000
151,5-154,5	153	1	153	23 409
154,5-157,5	156	1	156	24 336
157,5-160,5	159	6	954	151 686
160,5-163,5	162	7	1 134	183 708
163,5-166,5	165	9	1 485	245 025
166,5-169,5	168	6	1 008	169 344
169,5-172,5	171	3	513	87 723
172,5-175,5	174	4	696	121 104
175,5-178,5	177	1	177	31 329
		40	6 576	1 082 664

$$\text{MEDIA: } \bar{x} = \frac{\sum f_j x_j}{\sum f_j} = \frac{6\,576}{40} = 164,4 \text{ cm}$$

$$\text{VAR.: } \frac{\sum f_j x_j^2}{\sum f_j} - \bar{x}^2 = \frac{1\,082\,664}{40} - 164,4^2 = 39,24$$

$$\text{DESVIACIÓN TÍPICA: } \sigma = \sqrt{39,24} = 6,26 \text{ cm}$$

$$\text{C.V.} = \frac{\sigma}{\bar{x}} = \frac{6,26}{164,4} = 0,038 \rightarrow 3,8\%$$

2.ª distribución

INTERVALOS	x_j	f_j	$f_j x_j$	$f_j x_j^2$
147,5-151,5	149,5	2	299	44 700,5
151,5-155,5	153,5	1	153,5	23 562,25
155,5-159,5	157,5	4	630	99 225
159,5-163,5	161,5	10	1 615	260 822,5
163,5-167,5	165,5	12	1 986	328 683
167,5-171,5	169,5	6	1 017	172 381,5
171,5-175,5	173,5	4	694	120 409
175,5-179,5	177,5	1	177,5	31 506,25
		40	6 572	1 081 290

$$\text{MEDIA: } \bar{x} = \frac{\sum f_j x_j}{\sum f_j} = \frac{6\,572}{40} = 164,3 \text{ cm}$$

$$\text{VAR.: } \frac{\sum f_j x_j^2}{\sum f_j} - \bar{x}^2 = \frac{1\,081\,290}{40} - 164,3^2 = 37,76$$

$$\text{DESVIACIÓN TÍPICA: } \sigma = \sqrt{37,76} = 6,14 \text{ cm}$$

$$\text{C.V.} = \frac{\sigma}{\bar{x}} = \frac{6,14}{164,3} = 0,037 \rightarrow 3,7\%$$

Como se puede ver, las diferencias entre unas y otras son inapreciables.

4 Parámetros de posición para datos aislados

Página 198

1. Halla Q_1 , Me , Q_3 y p_{40} en esta distribución:

0	1	1	2	2	2	3	3	4	4	4	4	5	5	5
6	6	7	7	7	7	8	8	8	8	8	9	9	9	10

Hay 30 individuos en la distribución.

$30 : 4 = 7,5$ individuos en cada grupo

$7,5 \rightarrow$ individuo $8.^\circ \rightarrow Q_1 = 3$

$7,5 \cdot 2 = 15 \rightarrow$ individuo entre $15.^\circ$ y $16.^\circ \rightarrow Me = 5,5$

$7,5 \cdot 3 = 22,5 \rightarrow$ individuo $23.^\circ \rightarrow Q_3 = 8$

Para calcular p_{40} :

$30 \cdot \frac{40}{100} = 12 \rightarrow$ individuo entre $12.^\circ$ y $13.^\circ \rightarrow p_{40} = 4,5$

Página 199

2. En la siguiente distribución de notas, halla Me , Q_1 , Q_3 , p_{80} , p_{90} y p_{99} :

NOTAS	1	2	3	4	5	6	7	8	9	10
N.º DE ALUMNOS	6	19	37	45	109	81	39	22	30	12

NOTAS	f_i	F_i	% ACUM.
1	6	6	1,5
2	19	25	6,25
3	37	62	15,5
4	45	107	26,75
5	109	216	54
6	81	297	74,25
7	39	336	84
8	22	358	89,5
9	30	388	97
10	12	400	100

$$Me = p_{50} = 5$$

$$Q_1 = p_{25} = 4$$

$$Q_3 = p_{75} = 7$$

$$p_{80} = 7$$

$$p_{90} = 9$$

$$p_{99} = 10$$

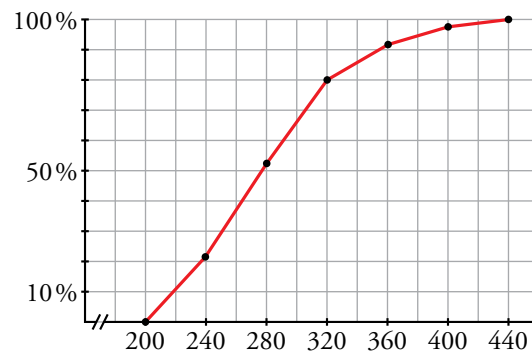
5 Parámetros de posición para datos agrupados

Página 200

1. Obtén la distribución de frecuencias acumuladas y representa el correspondiente polígono, relativo a los datos de la tabla siguiente:

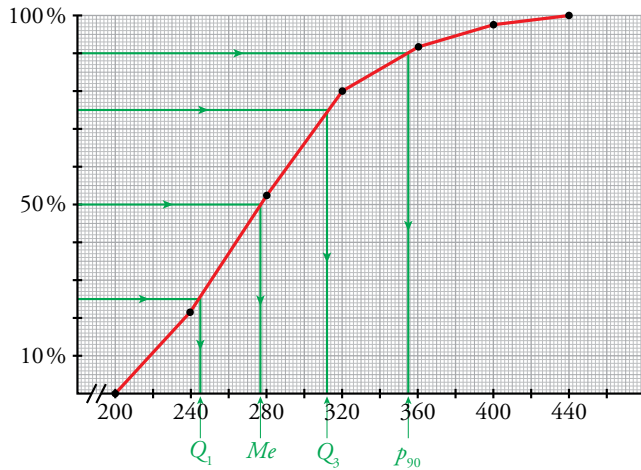
INTERVALOS	200-240	240-280	280-320	320-360	360-400	400-440
FRECUENCIAS	57	82	73	31	15	7

EXTREMOS	F_i	% ACUM.
200	0	0
240	57	21,51 \approx 21,5
280	139	52,45 \approx 52,5
320	212	80
360	243	91,7
400	258	97,36 \approx 97,4
440	265	100



Página 201

2. Halla, gráfica y numéricamente Q_1 , Me , Q_3 y p_{90} en la distribución del ejercicio propuesto de la página anterior. Averigua qué percentil corresponde a un valor de 325.



$Q_1 \approx 245$

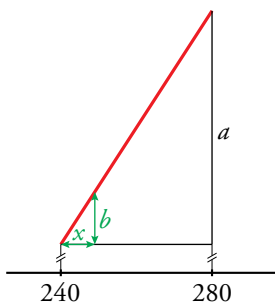
$Me \approx 277$

$Q_3 \approx 312$

$p_{90} \approx 355$

Numéricamente:

Cálculo de Q_1 :



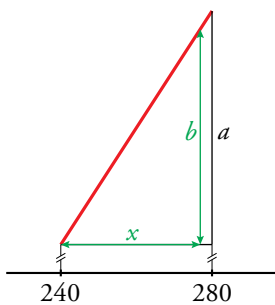
$a = 52,45 - 21,51 = 30,94$

$b = 25 - 21,51 = 3,49$

$\frac{30,94}{3,49} = \frac{40}{x} \rightarrow x = 4,5$

Por tanto: $Q_1 = 240 + 4,5 = 244,5$

Cálculo de Me :



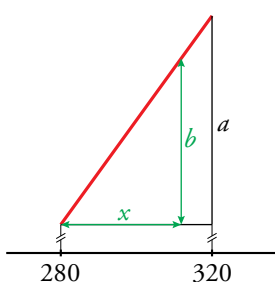
$a = 52,45 - 21,51 = 30,94$

$b = 50 - 21,51 = 28,49$

$\frac{30,94}{28,49} = \frac{40}{x} \rightarrow x = 36,83$

Por tanto: $Me = 240 + 36,83 = 276,83$

Cálculo de Q_3 :



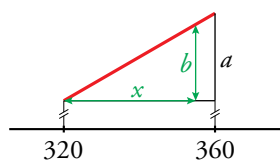
$a = 80 - 52,45 = 27,55$

$b = 75 - 52,45 = 22,55$

$\frac{27,55}{22,55} = \frac{40}{x} \rightarrow x = 32,74$

Por tanto: $Q_3 = 280 + 32,74 = 312,74$

Cálculo de p_{90} :



$$a = 91,70 - 80 = 11,70$$

$$b = 90 - 80 = 10$$

$$\frac{11,70}{10} = \frac{40}{x} \rightarrow x = 34,19$$

Por tanto: $p_{90} = 320 + 34,19 = 354,19$

Para calcular el percentil que corresponde a un valor de 325, utilizamos el esquema anterior. En este caso, debemos buscar x tal que:

$$\frac{11,70}{10} = \frac{x}{325 - 320} \rightarrow x = 5,85 \rightarrow \text{El percentil es } 85,85\% \approx 86\%.$$

6 Diagramas de caja

Página 203

1. Haz el diagrama de caja correspondiente a esta distribución de notas:

x_i	f_i
1	6
2	15
3	22
4	24
5	33
6	53
7	22
8	16
9	8
10	1

x_i	f_i	F_i
1	6	6
2	15	21
3	22	43
4	24	67
5	33	100
6	53	153
7	22	175
8	16	191
9	8	199
10	1	200

Comenzamos hallando Me , Q_1 y Q_3 :

$$n = 200$$

$$\frac{n}{2} = 100 \rightarrow Me = 5,5$$

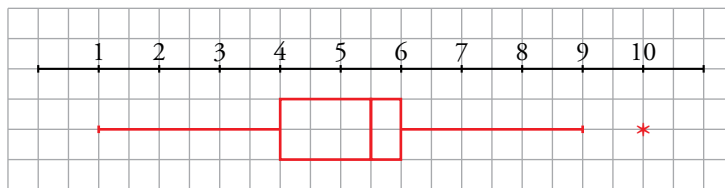
$$\frac{n}{4} = 50 \rightarrow Q_1 = 4$$

$$\frac{3}{4} \cdot n = 150 \rightarrow Q_3 = 6$$

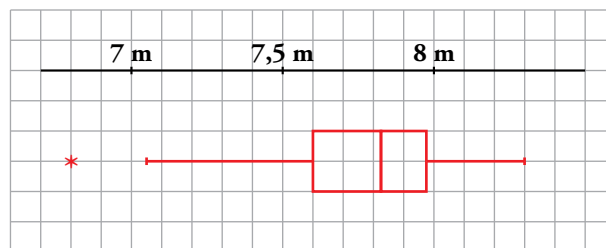
La longitud de la caja será $Q_3 - Q_1 = 6 - 4 = 2$.

$1,5 \cdot 2 = 3 \rightarrow$ Los bigotes llegarán hasta $4 - 3 = 1$ y hasta $6 + 3 = 9$.

Por tanto, el diagrama de caja y bigotes será:



2. Interpreta el siguiente diagrama de caja y bigotes relativo a las marcas de algunos saltadores de longitud:



$$Me = 7,825 \text{ m}; Q_1 = 7,6 \text{ m}; Q_3 = 7,975 \text{ m}$$

Todos saltaron entre 7,05 m y 8,3 m, excepto uno que saltó 6,8 m.

Un 25 % de los saltadores saltó menos de 7,6 m.

Un 25 % saltó entre 7,6 m y 7,825 m.

Un 25 % saltó entre 7,825 m y 7,975 m.

Un 25 % saltó más de 7,975 m.

7 Estadística inferencial

Página 204

- 1. Un fabricante de tornillos desea hacer un control de calidad. Recoge uno de cada 100 tornillos fabricados y lo analiza.**

El conjunto de tornillos analizados, ¿es población o muestra? ¿Por qué?

Los tornillos analizados constituyen una muestra, pues solo se analiza uno de cada cien tornillos fabricados.

- 2. El responsable de calidad de una empresa que fabrica pilas quiere estudiar la energía suministrada por cada pila hasta que se gasta.**

¿Puede hacer el estudio sobre la población o debe recurrir a una muestra? ¿Por qué?

Debe recurrir a una muestra porque el estudio requiere el consumo de las pilas.

- 3. El dueño de un vivero tiene 285 plantas de interior. Para probar la eficacia de un nuevo fertilizante, las mide todas antes y después del semestre que dura el tratamiento.**

El conjunto de esas 285 plantas, ¿es población o muestra? ¿Por qué?

Las 285 plantas sería la población. En este caso, es posible estudiar toda la población, no hace falta trabajar con una muestra.

Página 206

- 4.** Se les ha pasado un test a los 64 individuos de una muestra seleccionada aleatoriamente. Con los resultados obtenidos se ha llegado a la siguiente conclusión:

“La puntuación media que alcanzarían los individuos de toda la población si se les pasara este test estaría entre 42,7 puntos y 44,1 puntos. Y esto lo podemos afirmar con un nivel de confianza del 95%”.

a) Si el intervalo que se diera fuera 42-44,8, el nivel de confianza sería del...

- 90%
- 95%
- 98%

b) Si quisiéramos un nivel de confianza del 99% y un intervalo de la misma amplitud, ¿cómo tendría que ser la muestra?

- De menos de 64
- De 64
- De más de 64

a) 98 %, pues al ampliar la longitud del intervalo, también aumenta el nivel de confianza.

b) De más de 64, pues cuanto mayor es la muestra, mayor es el nivel de confianza.

Página 207

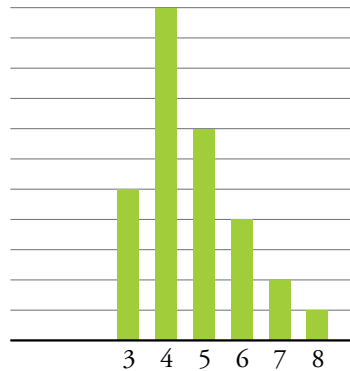
Hazlo tú. Se ha preguntado a los estudiantes de una clase por el número de personas que viven en casa. Los resultados son:

3 5 4 5 8 6 5 6 4 3
3 5 6 4 5 4 4 5 7 4
4 4 3 4 5 3 4 4 6 7

Resuelve los apartados del ejercicio resuelto anterior para los datos de este.

a)

x_i	f_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$
3	5	15	45
4	11	44	176
5	7	35	175
6	4	24	144
7	2	14	98
8	1	8	64
	30	140	702



MEDIA: $\bar{x} = \frac{140}{30} = 4,7$

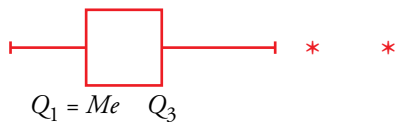
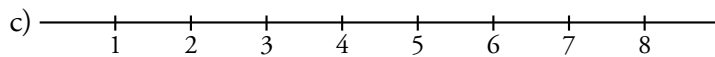
DESVIACIÓN TÍPICA: $\sigma = \sqrt{\frac{702}{30} - 4,7^2} = \sqrt{1,31} = 1,14$

COEFICIENTE DE VARIACIÓN: $C.V. = \frac{1,14}{4,7} = 0,24$

b)

x_i	f_i	F_i	% ACUM.
3	5	5	16,7
4	11	16	53,3
5	7	23	76,7
6	4	27	90
7	2	29	96,7
8	1	30	100

$Q_1 = p_{25} = 4$
 $Me = p_{50} = 4$
 $Q_3 = p_{75} = 5$
 $p_{30} = 4$
 $p_{90} = 6,5$
 $p_{99} = 8$



Página 208

Hazlo tú. Las longitudes (en cm) de una cierta especie de plantas son las siguientes:

13 23 14 16 25 24 15 19
 21 17 18 12 24 20 16 10
 14 17 20 18 15 22 45 14
 16 17 14 15 20 21 22 16

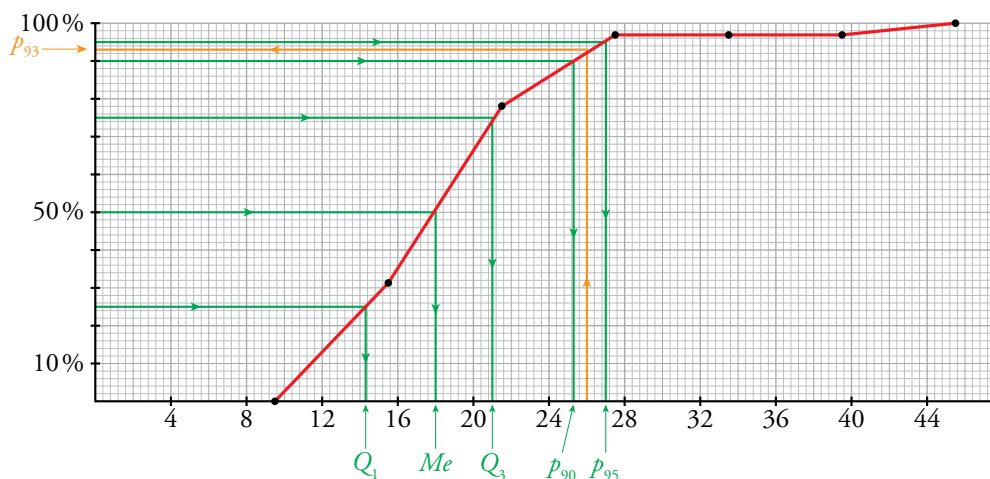
- a) Construye una tabla con datos agrupados en seis intervalos de longitud 6 que empiecen en 9,5.
- b) Representa el polígono de porcentajes acumulados y ayúdate de él para hallar Q_1 , Me , Q_3 , p_{90} y p_{95} . ¿En qué percentil está una planta de 26 cm de longitud?
- c) Representa los datos en un diagrama de caja.

a)

INTERVALOS	X_i	FRECUENCIAS
9,5-15,5	12,5	10
15,5-21,5	18,5	15
21,5-27,5	24,5	6
27,5-33,5	30,5	0
33,5-39,5	36,5	0
39,5-45,5	42,5	1

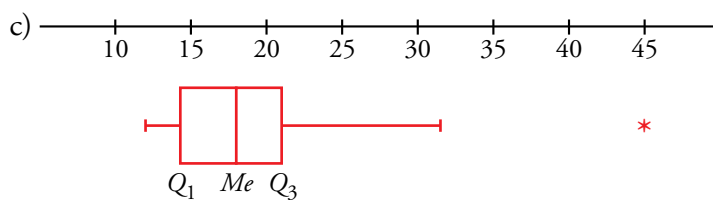
b)

EXTREMOS	F_i	% ACUM.
9,5	0	0
15,5	10	31,3
21,5	25	78,1
27,5	31	96,9
33,5	31	96,9
39,5	31	96,9
45,5	32	100



$Q_1 = 14,3$ $Me = 18$ $Q_3 = 21$ $p_{90} = 25,3$ $p_{95} = 27$

Planta de 26 cm $\rightarrow p_{93}$




Ejercicios y problemas

Página 209

Práctica

Tablas de frecuencias

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

Di cuál es la variable y de qué tipo es.

Haz una tabla de frecuencias y representa los datos en un diagrama adecuado.

- Variable: “Número de faltas de ortografía”

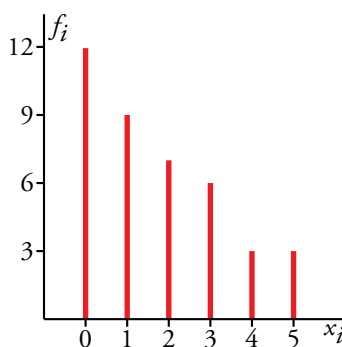
Es una variable cuantitativa discreta.

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

- Tabla de frecuencias:

x_i	f_i
0	12
1	9
2	7
3	6
4	3
5	3
	40

Diagrama de barras:



2.  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
3,3 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) ¿Cuál es la variable y de qué tipo es?

b) Construye una tabla con los datos agrupados en 6 intervalos desde 1,65 hasta 4,05.

Localizamos los valores extremos: 1,9 y 3,9. Recorrido = $3,9 - 1,8 = 2,1$


a) Variable: peso de los recién nacidos.

Tipo: cuantitativa continua.

b)

INTERVALOS	MARCA DE CLASE (x_i)	f_i
1,65-2,05	1,85	4
2,05-2,45	2,25	5
2,45-2,85	2,65	13
2,85-3,25	3,05	16
3,25-3,65	3,45	9
3,65-4,05	3,85	3
		50

Media, desviación típica y C.V.

3.  Halla la media, la desviación típica y el coeficiente de variación en estas distribuciones:

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

INTERVALO	f_i
1,65-2,05	4
2,05-2,45	5
2,45-2,85	13
2,85-3,25	17
3,25-3,65	8
3,65-4,05	3

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

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

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

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

$$\text{C.V.} = \frac{\sigma}{\bar{x}} = 0,9235 \rightarrow 92,35\%$$


INTERVALOS	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

$$\bar{x} = \frac{144,1}{50} = 2,9$$

$$\text{VAR.} = \frac{428,12}{50} - 2,8^2 = 0,1524$$

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

$$\text{C.V.} = \frac{0,39}{2,9} = 0,1345 \rightarrow 13,45\%$$

4.  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 más variación relativa.

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

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

Tiene mayor variación relativa la empresa B.

Parámetros de posición para datos aislados

5.  La altura, en centímetros, de un grupo de estudiantes de una misma clase es:

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

Halla 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 estudiantes tiene una estatura inferior a 175,5 cm.

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

El 25 % de los estudiantes 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 estudiantes tiene una estatura inferior a 181 cm.

6.  Halla la mediana, los cuartiles y el percentil 60 en cada una de las siguientes distribuciones correspondientes al número de respuestas correctas en un test realizado por dos grupos de estudiantes:

A: 25 - 22 - 27 - 30 - 23 - 22 - 31 - 18 B: 27 - 32 - 19 - 22 - 25 - 30 - 21

24 - 25 - 32 - 35 - 20 - 28 - 30 29 - 23 - 31 - 21 - 20 - 18 - 27

Colocamos en orden creciente los datos:

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) $\rightarrow Me = 25$


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

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

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

7.  Rellena la columna de los porcentajes acumulados en la siguiente tabla. Calcula, a partir de la tabla, la mediana, los cuartiles y los percentiles p_{70} y p_{90} .

x_i	f_i	F_i	% ACUM.
0	12	12	32,4
1	9	21	56,8
2	7	28	75,7
3	6	34	91,9
4	3	37	100


$Q_1 = 0$

$Me = 1$

$Q_3 = 2$

$p_{70} = 2$

$p_{90} = 3$

8.  En la fabricación de cierto tipo de bombillas se han detectado algunas defectuosas. Se analiza el contenido de 200 cajas de 100 bombillas cada una y se obtienen los siguientes resultados:

DEFECTUOSAS	1	2	3	4	5	6	7	8
N.º DE CAJAS	5	15	38	42	49	31	18	2

Calcula la mediana, los cuartiles y los percentiles p_{10} , p_{90} y p_{95} .

Hacemos la tabla de frecuencias acumuladas.

x_i	f_i	F_i	% ACUM.
1	5	5	2,5
2	15	20	10
3	38	58	29
4	42	100	50
5	49	149	74,5
6	31	180	90
7	18	198	99
8	2	200	100

Para $x_i = 4$, F_i iguala el 50 %, luego la mediana será el valor intermedio entre 4 y el siguiente, 5, esto es, $Me = 4,5$.

$Q_1 = p_{25} = 3$

$Q_3 = p_{75} = 6$

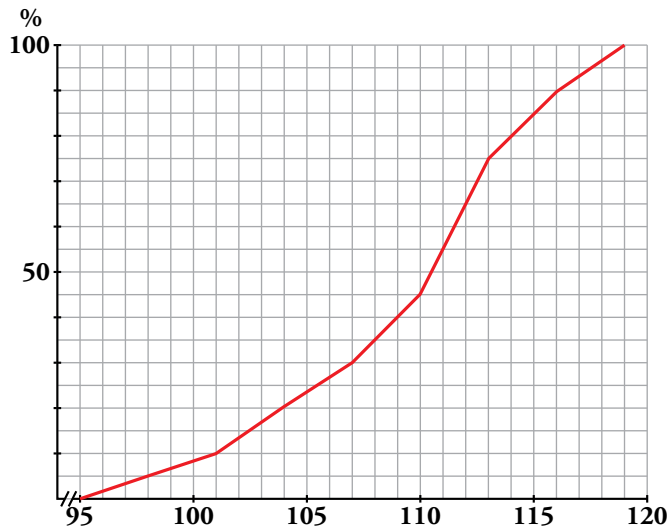
$p_{10} = 2,5$

$p_{90} = 6,5$

$p_{95} = 7$

Parámetros de posición para datos agrupados en intervalos

9. Este es el polígono de porcentajes acumulados de la distribución del CI (cociente intelectual) de un colectivo de 300 personas:



- a) Asigna, de forma aproximada, los valores de Q_1 , Me , Q_3 , p_5 , p_{10} , p_{40} , p_{80} , p_{90} y p_{95} .
- b) ¿Cuántas personas (aproximadamente) de este colectivo tienen un CI comprendido entre 104 y 116? ¿Cuántas tienen un CI superior a 115?
- c) ¿Qué percentil tiene una persona con CI de 112?

- a) $Q_1 = 105,5$ $p_5 = 98$ $p_{80} = 114$
 $Me = 110,5$ $p_{10} = 101$ $p_{90} = 116$
 $Q_3 = 113$ $p_{40} = 109$ $p_{95} = 117,5$

b) Un 20 % tiene un CI inferior a 104, es decir, 20 % de 300 = 60 personas.

Un 90 % tiene un CI inferior a 116, es decir, 90 % de 300 = 270 personas.

Por tanto, hay unas 210 personas con un CI entre 104 y 116.

Un 85 % tiene un CI inferior a 115, luego un 15 % lo tendrá superior, es decir, 15 % de 300 = 45 personas.

c) p_{65}

10. Las estaturas de un grupo de estudiantes de la misma clase vienen dadas en esta tabla:

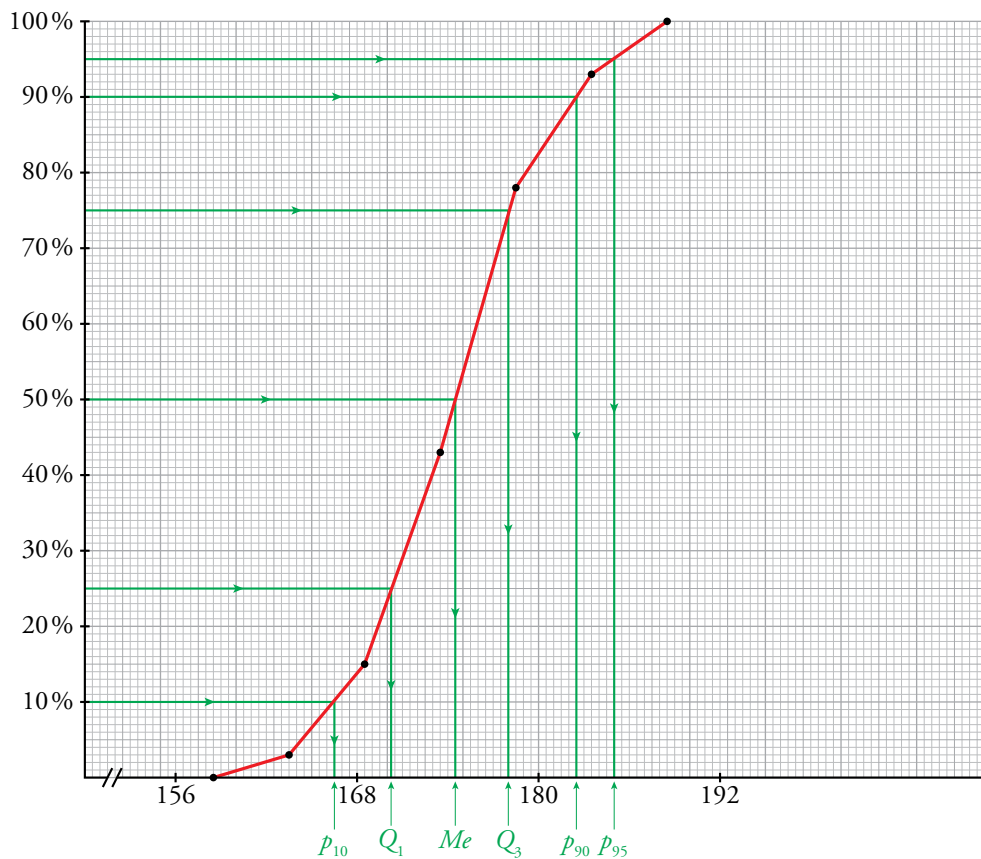
ALTURA	N.º DE ALUMNOS
158,5-163,5	1
163,5-168,5	5
168,5-173,5	11
173,5-178,5	14
178,5-183,5	6
183,5-188,5	3

- a) **Elabora la tabla de frecuencias y porcentajes acumulados. Indica en la tabla, en lugar de los intervalos, los extremos.**
- b) **Representa, a partir de la tabla, el polígono de porcentajes acumulados.**
- c) **Calcula, de forma aproximada (sobre el polígono), la mediana, los cuartiles y los percentiles p_{10} , p_{90} y p_{95} .**
- d) **¿Qué percentil corresponde a un estudiante de 180 cm de altura?**

a)

EXTREMOS	F_i	% ACUM.
158,5	0	0
163,5	1	2,5 \approx 3
168,5	6	15
173,5	17	42,5 \approx 43
178,5	31	77,5 \approx 78
183,5	37	92,5 \approx 93
188,5	40	100

b) y c)



$Q_1 \approx 170,5$

$Me \approx 174,5$

$Q_3 \approx 178$

$p_{10} \approx 166,5$

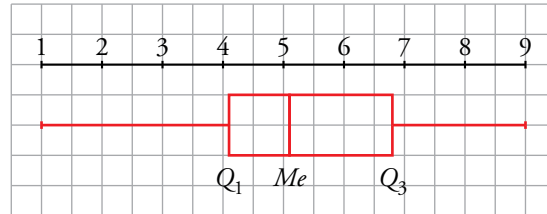
$p_{90} \approx 182,5$

$p_{95} \approx 185$

d) En la gráfica se puede observar que 180 cm $\rightarrow p_{82}$, aproximadamente.

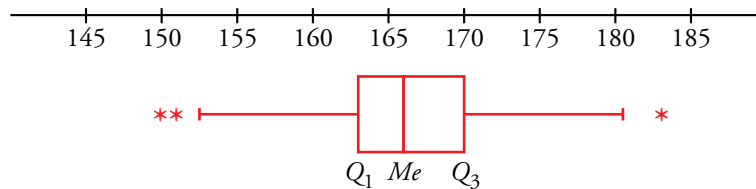
Diagramas de caja

11. Las puntuaciones obtenidas por 87 personas tienen los siguientes parámetros de posición: $Q_1 = 4,1$; $Me = 5,1$ y $Q_3 = 6,8$. Todas las puntuaciones están en el intervalo 1 a 9. Haz el diagrama de caja.



12. En una clase de 38 estudiantes de Primaria, las estaturas de 35 de ellos están comprendidas entre 153 cm y 179 cm. Los tres restantes miden 150 cm, 151 cm y 183 cm. Sabemos que $Q_1 = 163$; $Me = 166$ y $Q_3 = 170$.

Representa los datos en un diagrama de caja.



13. Haz el diagrama de caja correspondiente a las siguientes distribuciones.

a) La del ejercicio 5.

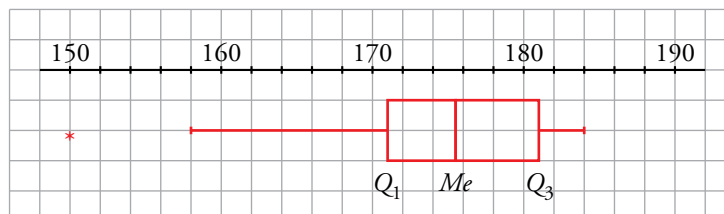
b) La A y la B del ejercicio 6.

c) La del ejercicio 7.

d) La del ejercicio 8.

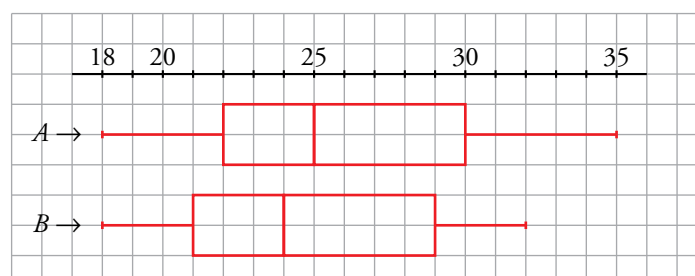
a) $Q_1 = 171$; $Me = 175,5$; $Q_3 = 181$

$$(Q_3 - Q_1) \cdot 1,5 = (181 - 171) \cdot 1,5 = 10 \cdot 1,5 = 15 \quad \begin{cases} 171 - 15 = 156 \\ 181 + 15 = 196 \end{cases}$$

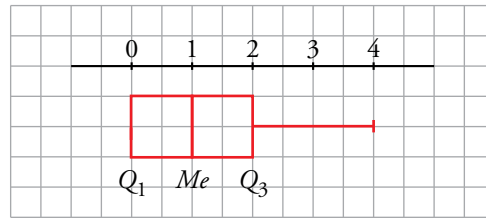


b) A: $Q_1 = 22$; $Me = 25$; $Q_3 = 30$

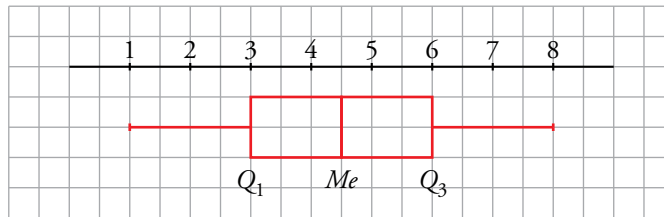
B: $Q_1 = 21$; $Me = 24$; $Q_3 = 29$



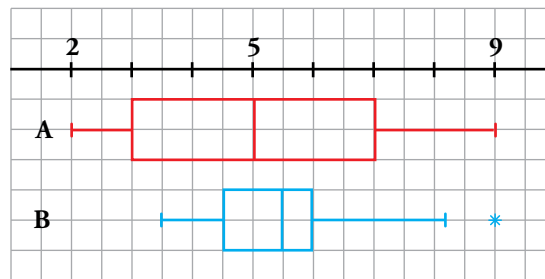
c) $Q_1 = 0$; $Me = 1$; $Q_3 = 2$



d) $Q_1 = 3$; $Me = 4,5$; $Q_3 = 6$



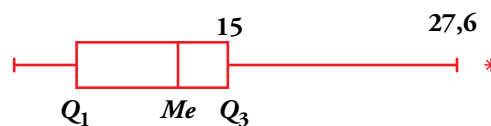
14. A los estudiantes de dos clases numerosas de un mismo centro les han puesto un test. Las notas vienen reflejadas en los siguientes diagramas de caja:



- a) ¿Cuál de las clases es más homogénea?
- b) ¿En cuál ha aprobado la mitad de la clase?
- c) En una de las clases, la tercera nota más alta ha sido un 6,5. ¿De qué clase se trata?
- d) ¿En qué clase las notas del 25 % de los estudiantes difieren en medio punto o menos?
- e) ¿Cuál es el rango de las notas de cada clase?

- a) Es más homogénea la clase B.
- b) En la clase A ha aprobado exactamente la mitad de la clase.
- c) En la clase B, puesto que en la clase A el 25 % tiene notas entre 5 y 7.
- d) En la clase B.
- e) En la clase A el rango es $9 - 2 = 7$.
En la clase B el rango es $9 - 3,5 = 5,5$.

15. Calcula el valor del primer cuartil correspondiente al siguiente diagrama de caja:




$27,6 - 15 = 12,6$

$12,6 : 1,5 = 8,4$, por tanto, $Q_3 - Q_1 = 8,4 \rightarrow 15 - Q_1 = 8,4$.

Luego, $Q_1 = 6,6$.

Muestreo


16.  Se quieren realizar estos estudios estadísticos:

- I. Tipo de transporte que utilizan los vecinos de un barrio para acudir a sus trabajos.
- II. Estudios que piensan seguir los estudiantes de un centro escolar al terminar la ESO.
- III. Edad de las personas que han visto una obra de teatro en una ciudad.
- IV. Número de horas diarias que ven la televisión los niños y las niñas de tu comunidad autónoma con edades comprendidas entre 5 y 10 años.
- V. Tiempo de conversación que aguantan las baterías de los móviles que fabrican en una empresa.
- VI. Preferencia de emisora de radio musical de los asistentes a un concierto.

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

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

- a) I → Los vecinos del barrio.
 - II → Alumnos y alumnas de la ESO de un centro.
 - III → Personas que han visto la obra.
 - IV → Niños y niñas de mi comunidad autónoma de entre 5 y 10 años.
 - V → Los móviles que fabrica la empresa.
 - VI → Los asistentes a un concierto.
- b) I → Dependiendo del número de vecinos del barrio: si son pocos, población; si son muchos, una muestra. Aunque teniendo en cuenta que es difícil cogerlos a todos y que todos contesten a la encuesta, quizás sería mejor una muestra.
 - II → Población. Con encuestas en clase en las que participan todos (obviamente, siempre falta alguno).
 - III → Muestra. Son muchas personas y sería inoportuno molestar a tanta gente, se formarían colas...
 - IV → Muestra. Son demasiadas personas.
 - V → Es necesario recurrir a una muestra para el estudio porque llevarlo a cabo requiere el desgaste de las baterías.
 - VI → Será necesario recurrir a una muestra.

17.  ¿Cómo se puede contar el número aproximado de palabras que tiene un cierto libro?

- Se seleccionan, abriendo al azar, unas cuantas páginas y se cuentan las palabras en cada una.
- Se calcula el número medio de palabras por página.
- Se da un intervalo en el que pueda estar comprendido el número total de palabras.

Hazlo con alguna novela que encuentres en casa. Cuanto más homogéneas sean sus páginas, más precisión tendrás en el resultado.


- En un libro de 200 páginas, seleccionamos al azar 5 páginas. Contamos el número de palabras de estas páginas: 537, 562, 548, 324, 600.

- Calculamos el número medio de palabras:

$$\frac{537 + 562 + 548 + 324 + 600}{5} = 514,2$$

En 200 páginas, habrá 102 840 palabras.

- El número de palabras del libro estará entre 100 000 y 105 000.

18.  Para hacer un sondeo electoral en un pueblo de 2 000 electores, aproximadamente, se va a elegir una muestra de 200 individuos. Di si te parece válido cada uno de los siguientes modos de seleccionarlos y explica por qué:

a) Se le pregunta al alcalde, que conoce a todo el pueblo, qué individuos le parecen más representativos.

b) Se eligen 200 personas al azar entre las que acuden a la verbena el día del patrón.

c) Se seleccionan al azar en la guía telefónica y se les encuesta por teléfono.

d) Se acude a las listas electorales y se seleccionan al azar 200 de ellos.


a) No es válido. Se trata de una elección subjetiva.

b) No es válido. Probablemente haya grupos de edades mucho más representados que otros.

c) Sí es válido.

d) Sí es válido.

Aplica lo aprendido

19.  El número de errores cometidos en un test por un grupo de personas viene reflejado en esta tabla:

N.º DE ERRORES	0	1	2	3	4	5	6
N.º DE PERSONAS	10	12	8	7	5	4	3

a) Halla la mediana, los cuartiles inferior y superior y los percentiles p_{20} , p_{40} y p_{90} . Explica su significado.

b) ¿Cuál es el número medio de errores por persona?

Completamos la siguiente tabla:

N.º DE ERRORES	0	1	2	3	4	5	6
N.º DE PERSONAS	10	12	8	7	5	4	3
F_i	10	22	30	37	42	46	49
% ACUMULADO	20,4	44,9	61,2	75,5	85,7	93,9	100


a) $p_{20} = 0$ $Q_1 = p_{25} = 1$

$p_{40} = 1$ $Me = p_{50} = 2$

$p_{90} = 5$ $Q_3 = p_{75} = 3$

$p_m = n$ significa que el $m\%$ de las personas comete un máximo de n errores.

b) $\bar{x} = \frac{[0 \cdot 10 + 1 \cdot 12 + 2 \cdot 8 + 3 \cdot 7 + 4 \cdot 5 + 5 \cdot 4 + 6 \cdot 3]}{49} = 2,18$ errores por persona.

20.  Al preguntar a los miembros de un grupo de lectura cuánto tiempo dedicaron a leer durante un fin de semana, se obtuvieron estos resultados:

a) Dibuja el histograma correspondiente (¡Atención! Los intervalos tienen distintos tamaños y las frecuencias deben ser proporcionales a las áreas de los rectángulos que forman el histograma).

b) Halla la media, la desviación típica y el coeficiente de variación.

c) Dibuja el polígono de porcentajes acumulados.

d) Halla Q_1 , Me , Q_3 y p_{90} .

e) A un miembro del grupo que en ese fin de semana ha leído 6 horas y cuarto, ¿qué percentil le corresponde?

f) Representa los datos en un diagrama de caja.

a) 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	N.º DE PERSONAS
[0; 0,5)	10
[0,5; 1,5)	10
[1,5; 2,5)	18
[2,5; 4)	12
[4, 8)	12
[8, 12]	8

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

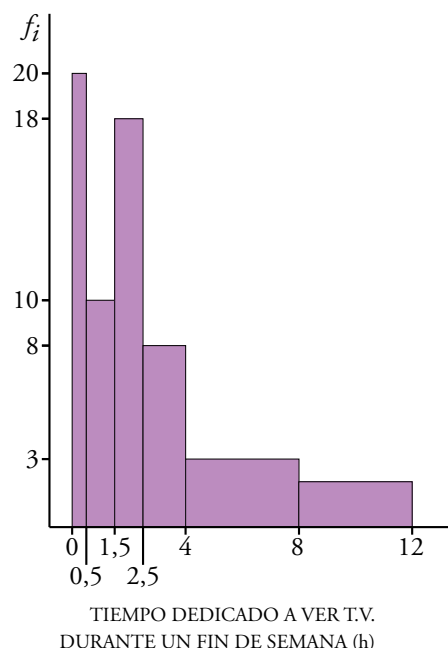
$[0,5; 1,5) \rightarrow a_2 = 1 \quad f_2 = 10 \rightarrow h_2 = 10$

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

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

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

$[8, 12] \rightarrow a_6 = 4 \quad f_6 = 8 \rightarrow h_6 = \frac{8}{4} = 2$



b)

TIEMPO	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
[8; 12]	10	8	80	800
		70	239,5	1 441,375

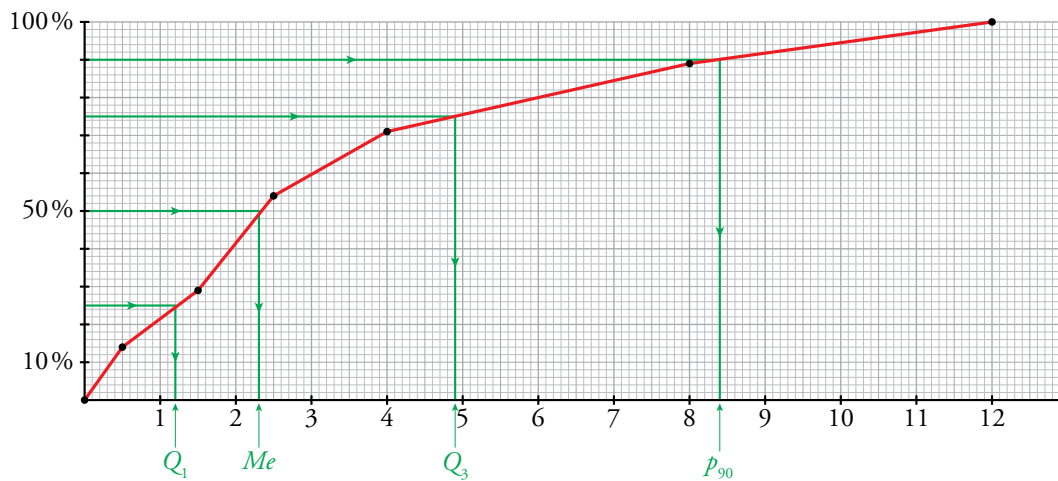
$$\bar{x} = \frac{239,5}{70} = 3,42$$

$$\sigma = \sqrt{\frac{1 441,375}{70} - 3,42^2} = \sqrt{8,89} = 2,98$$

$$C.V. = \frac{\sigma}{\bar{x}} = \frac{2,98}{3,42} = 0,87$$

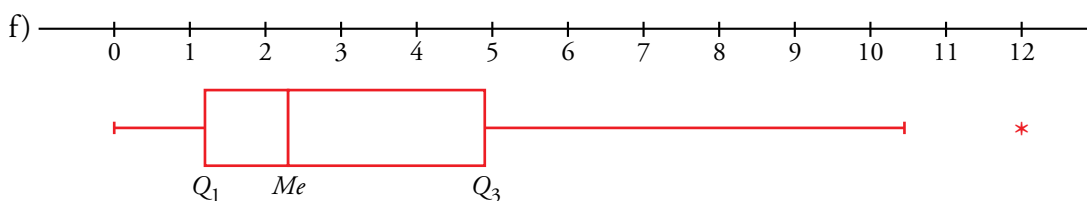
c)

EXTREMOS	F_i	% ACUM.
0	0	0
0,5	10	14,29 \approx 14
1,5	20	28,57 \approx 29
2,5	38	54,29 \approx 54
4	50	71,43 \approx 71
8	62	88,57 \approx 89
12	70	100



d) $Q_1 \approx 1,2$ $Me \approx 2,3$ $Q_3 \approx 4,9$ $p_{90} \approx 8,4$

e) El miembro que ha leído 6 horas y cuarto está en el percentil p_{81} .



21.  Deseamos hacer una tabla de datos agrupados a partir de 384 datos, cuyos valores extremos son 19 y 188.

a) Si queremos que sean 10 intervalos de amplitud 17, ¿cuáles serán esos intervalos?

b) Haz otra distribución en 12 intervalos de la amplitud que creas conveniente.

Recorrido $r = 188 - 19 = 169$

a) Buscamos un número mayor que r que sea múltiplo de 10 $\rightarrow r' = 170$.

Cada intervalo tendrá longitud 17.

Como $r' - r = 1$, comenzamos 0,5 antes del primer dato y finalizamos 0,5 después del último dato.

Los intervalos son:

[18,5; 35,5); [35,5; 52,5); [52,5; 69,5); [69,5; 86,5); [86,5; 103,5);

[103,5; 120,5); [120,5; 137,5); [137,5; 154,5); [154,5; 171,5); [171,5; 188,5)

b) Ahora buscamos un múltiplo de 12 mayor que 169 $\rightarrow r' = 180$.

Como $r' - r = 180 - 169 = 11$, comenzamos 5,5 antes del primer dato y finalizamos 5,5 después del último dato y cada intervalo tendrá amplitud $180 : 12 = 15$.

Los intervalos son:

[13,5; 28,5); [28,5; 43,5); [43,5; 58,5); [58,5; 73,5);

[73,5; 88,5); [88,5; 103,5); [103,5; 118,5); [118,5; 133,5);

[133,5; 148,5); [148,5; 163,5); [163,5; 178,5); [178,5; 193,5)

Página 212

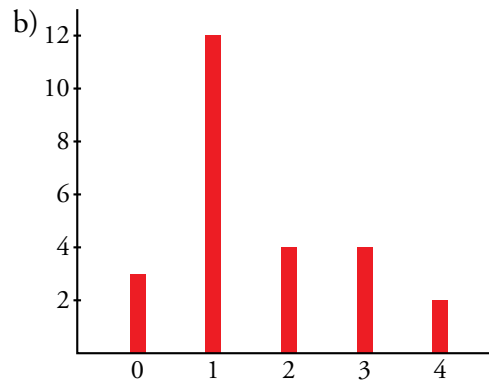
22. En una urbanización de 25 familias se ha observado la variable “número de coches que tiene la familia” y se han obtenido los siguientes datos:

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

- a) Construye la tabla de frecuencias.
- b) Haz el diagrama de barras.
- c) Calcula la media y la desviación típica.
- d) Halla la mediana, los cuartiles y los percentiles p_{40} y p_{90} .
- e) Dibuja el diagrama de caja.

a)

x_i	f_i
0	3
1	12
2	4
3	4
4	2



c)

x_i	f_i	$f_i x_i$	$f_i x_i^2$
0	3	0	0
1	12	12	12
2	4	8	16
3	4	12	36
4	2	8	32
	25	40	96

$$\bar{x} = \frac{40}{25} = 1,6$$

$$\sigma = \sqrt{\frac{96}{25} - 1,6^2} = 1,13$$

d)

x_i	f_i	F_i	% ACUM.
0	3	3	12
1	12	15	60
2	4	19	76
3	4	23	92
4	2	25	100

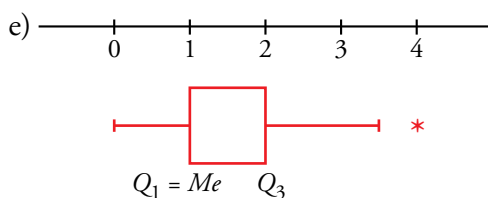
$$Q_1 = 1$$


$$Me = 1$$

$$Q_3 = 2$$

$$p_{40} = 1$$

$$p_{90} = 3$$



23.  El número de personas que acudieron cada día a las clases de natación de una piscina municipal fueron:

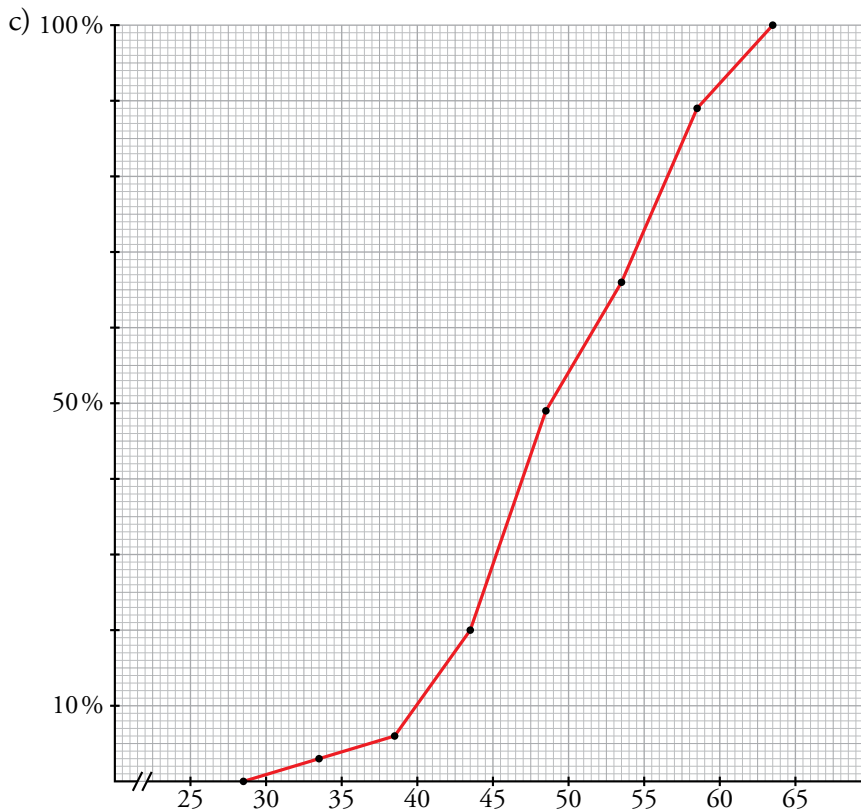
38	31	54	47	50	56	52	48	55	60
58	46	47	55	60	53	43	52	46	55
43	60	45	48	40	56	54	48	39	50
53	59	48	39	48					

- Elabora una tabla de frecuencias agrupando los datos en 7 intervalos de amplitud 5 empezando por 28,5.
- A partir de la tabla anterior, haz la tabla con los extremos y los porcentajes acumulados.
- Dibuja en un papel milimetrado el polígono de porcentajes acumulados.
- Calcula aproximadamente, a partir del polígono del apartado anterior, la mediana, los cuartiles y los percentiles p_{60} , p_{95} y p_{99} .
- ¿A qué percentil corresponde un día que acudieron 70 personas?

a) Agrupamos los datos en 7 intervalos de longitud 5. b)

INTERVALOS	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	1296
38,5-43,5	41	5	205	8405
43,5-48,5	46	10	460	21160
48,5-53,5	51	6	306	15606
53,5-58,5	56	8	448	25088
58,5-63,5	61	4	244	14884
		35	1730	87400

EXTREMOS	F_i	% ACUM.
28,5	0	0
33,5	1	2,9 ≈ 3
38,5	2	5,7 ≈ 6
43,5	7	20
48,5	17	48,6 ≈ 49
53,5	23	65,7 ≈ 66
58,5	31	88,6 ≈ 89
63,5	35	100



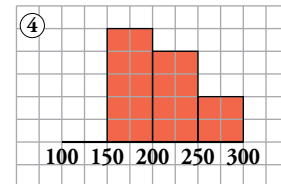
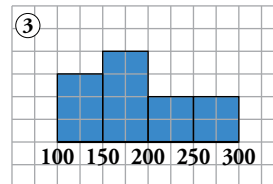
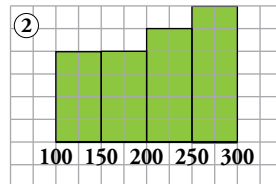
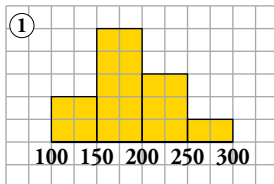
- Los valores aproximados son: $Q_1 = 44,5$ $Me = 49$ $Q_3 = 55,5$ $p_{60} = 52$ $p_{95} = 61$ $p_{99} = 63$
- El percentil que corresponde a un día en el que acudieron 70 personas es p_{100} puesto que ningún día acuden más de 60 personas.

Resuelve problemas

24. Se ha medido el nivel de colesterol en cuatro grupos de personas sometidas a diferentes dietas. Las medias y las desviaciones típicas son las de la tabla:

DIETA	A	B	C	D
\bar{x}	211,4	188,6	209,2	188,6
σ	37,5	52,6	56,3	43,1

Asocia a cada dieta la gráfica que le corresponde.



Observamos que las gráficas 1 y 3 muestran distribuciones con una media inferior a 200, mientras que las gráficas 2 y 4 muestran distribuciones con una media superior a 200.

Por tanto: A y C \rightarrow 2 y 4; B y D \rightarrow 1 y 3

Por otro lado, los datos en la gráfica 2 están más dispersos que en la gráfica 4, por tanto, la desviación típica es mayor. Así: C \rightarrow 2; A \rightarrow 4

De igual forma, los datos están más dispersos en la gráfica 3 que en la gráfica 1 y, por tanto: B \rightarrow 3; D \rightarrow 1.

25. En un grupo de estudiantes, cada uno cuenta el número de personas y el número de perros que viven en su portal. Juntan sus resultados y obtienen una muestra con la que se puede estimar el número de perros que hay en su ciudad.

Por ejemplo, supongamos que en su observación obtienen un total de 747 personas y 93 perros. Y saben que en su ciudad viven 75 000 personas.

- Con estos datos, ¿cuántos perros estiman que habrá en la ciudad?
- ¿Te parece que el procedimiento de elección de muestra es estadísticamente correcto? ¿Es aleatoria?
- Teniendo en cuenta la elección de muestra, ¿cómo de fiable es esta estimación?

a) $\frac{93}{747} \cdot 75\,000 \approx 9\,337$ perros

b) No es correcto el procedimiento de elección de la muestra porque se limita a una zona de la ciudad, donde viven los estudiantes que acuden a ese centro educativo. No es aleatoria.


c) Sería poco fiable la estimación puesto que la muestra no es representativa.

26. 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, por ejemplo, en un portal de 32 viviendas.

No se obtiene una muestra aleatoria, porque una familia que vive en una vivienda unifamiliar tiene más probabilidades de ser elegida que una familia que vive en un bloque de viviendas.

Problemas “+”

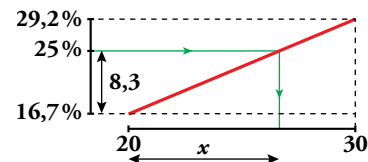
27.  Se ha pasado un test de 80 preguntas a 600 personas. La frecuencia y el porcentaje acumulado de respuestas correctas se refleja en la siguiente tabla:

EXTREMOS	FREC. ACUM.	% ACUM.
0	0	0
10	40	6,7
20	100	16,7
30	175	29,2
40	265	44,2
50	370	61,7
60	455	75,8
70	535	89,2
80	600	100

Hallamos Q_1 así:

$$\frac{x}{8,3} = \frac{30 - 20}{29,2 - 16,7} \rightarrow \frac{x}{8,3} = \frac{10}{12,5} \rightarrow$$

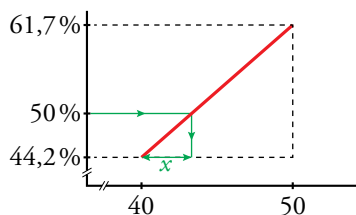
$$\rightarrow x = 6,64 \rightarrow Q_1 = 26,64$$



a) Calcula, de igual forma, Me , Q_3 , p_{20} y p_{85} .

b) ¿Cuál es el percentil de una persona que tiene 65 respuestas correctas?

a) Cálculo de Me :

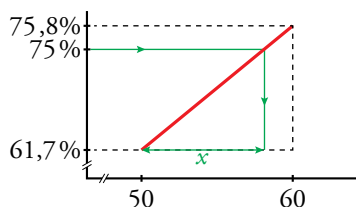


$$\frac{x}{50 - 44,2} = \frac{50 - 40}{61,7 - 44,2} \rightarrow$$

$$\rightarrow \frac{x}{5,8} = \frac{10}{17,5} \rightarrow$$

$$\rightarrow x = 3,31 \rightarrow Me = 43,31$$

Cálculo de Q_3 :

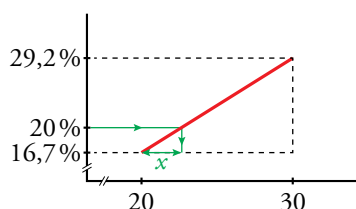


$$\frac{x}{75 - 61,7} = \frac{10}{75,8 - 61,7} \rightarrow$$

$$\rightarrow \frac{x}{13,3} = \frac{10}{14,1} \rightarrow$$

$$\rightarrow x = 9,43 \rightarrow Q_3 = 59,43$$

Cálculo de p_{20} :

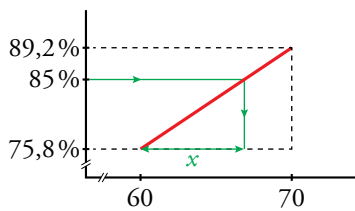


$$\frac{x}{20 - 16,7} = \frac{10}{29,2 - 16,7} \rightarrow$$

$$\rightarrow \frac{x}{3,3} = \frac{10}{12,5} \rightarrow$$

$$\rightarrow x = 2,64 \rightarrow p_{20} = 22,64$$

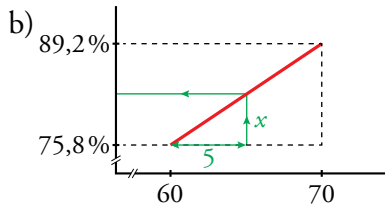
Cálculo de p_{85} :



$$\frac{x}{85 - 75,8} = \frac{10}{89,2 - 75,8} \rightarrow$$

$$\rightarrow \frac{x}{9,2} = \frac{10}{13,4} \rightarrow$$

$$\rightarrow x = 6,87 \rightarrow p_{85} = 66,87$$



$$\frac{5}{x} = \frac{10}{89,2 - 75,8} \rightarrow$$

$$\rightarrow \frac{5}{x} = \frac{10}{13,4} \rightarrow$$

$$\rightarrow x = 6,7 \rightarrow 75,8 + 6,7 = 82,5 \rightarrow p_{83}$$

28. En una fábrica se ha medido la longitud de 1 000 piezas de las mismas características y se ha obtenido:

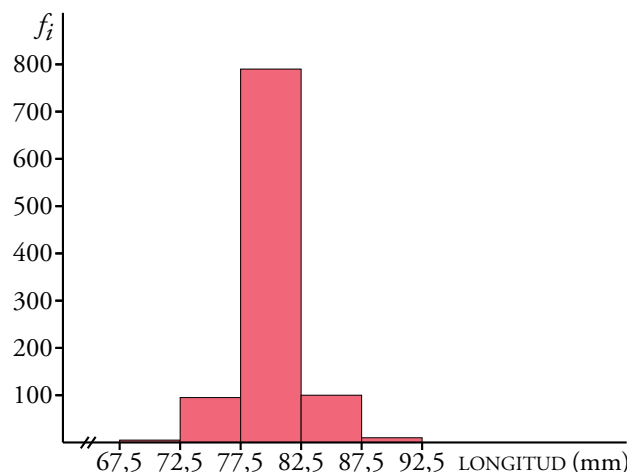
LONGITUD (en mm)	N.º DE PIEZAS
67,5-72,5	5
72,5-77,5	95
77,5-82,5	790
82,5-87,5	100
87,5-92,5	10

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 2.º intervalo habrá que rechazar las que midan entre 72,5 y 75. Calcula, como en el ejercicio anterior, qué tanto por ciento de la amplitud representa la diferencia 75 – 72,5 y calcula el porcentaje de la frecuencia correspondiente. Haz lo mismo para el 4.º intervalo.

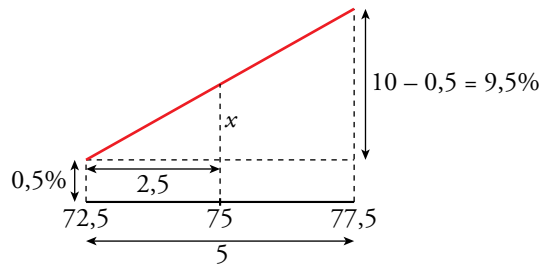
a) Por tener todos los intervalos la misma longitud, la altura de cada una de las barras coincidirá con la frecuencia de cada intervalo.



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

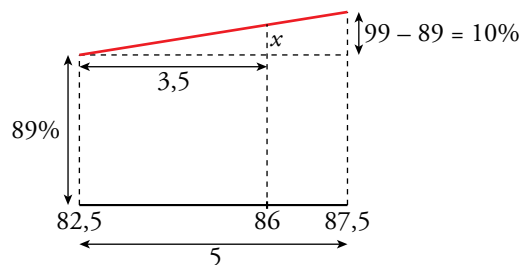
- Calculamos el porcentaje de piezas que hay por debajo de 75 mm:



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

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.

29. De una muestra de 75 pilas eléctricas, se dan estos datos sobre su duración:

- Halla \bar{x} y σ y calcula el porcentaje de pilas que hay en el intervalo $(\bar{x} - \sigma, \bar{x} + \sigma)$.
- Calcula Q_1 , Me , Q_3 , p_{30} , p_{60} y p_{95} .

TIEMPO (en horas)	N.º DE PILAS
25-30	3
30-35	5
35-40	21
40-45	28
45-55	12
55-70	6

a)

x_i	f_i	$f_i x_i$	$f_i x_i^2$
27,5	3	82,5	2 268,75
32,5	5	162,5	5 281,25
37,5	21	787,5	29 531,25
42,5	28	1 190	50 575
50	12	600	30 000
62,5	6	375	23 437,5
	75	3 197,5	141 093,75

$$\bar{x} = \frac{3197,5}{75} = 42,63$$

$$\sigma = \sqrt{\frac{141093,75}{75} - 42,63^2} = \sqrt{63,9331} \approx 8$$

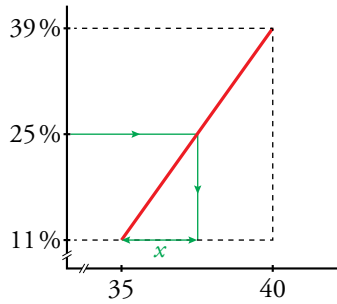
$$\bar{x} - \sigma = 34,63$$

$$\bar{x} + \sigma = 50,63$$

En el intervalo (34,63; 50,63) hay, aproximadamente, 61 pilas, lo que supone un 81,33%.

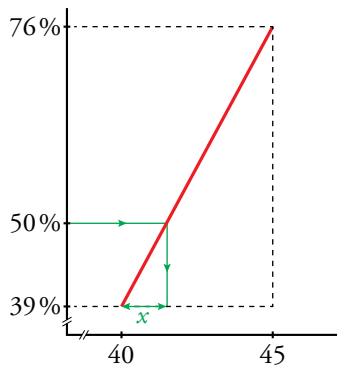
b)

EXTREMOS	F_i	% ACUM.
25	0	0
30	3	4
35	8	10,7 ≈ 11
40	29	38,7 ≈ 39
45	57	76
55	69	92
70	75	100



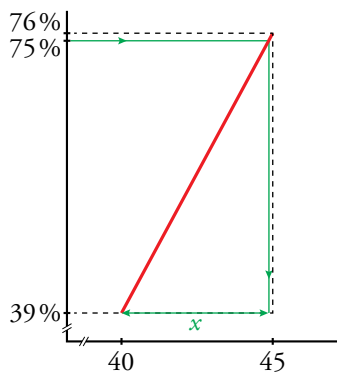
$$\frac{39 - 11}{25 - 11} = \frac{40 - 35}{x}$$

$$x = 2,5 \rightarrow Q_1 = 35 + 2,5 = 37,5$$



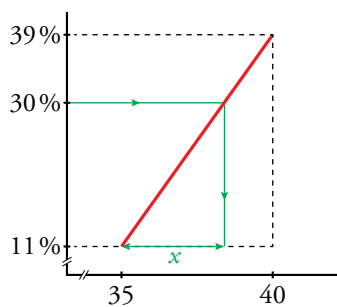
$$\frac{76 - 39}{50 - 39} = \frac{45 - 40}{x}$$

$$x = 1,86 \rightarrow Me = 40 + 1,86 = 41,86$$



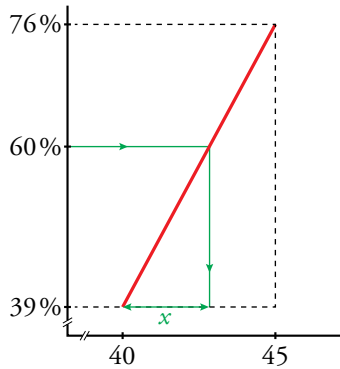
$$\frac{76 - 39}{75 - 39} = \frac{45 - 40}{x}$$

$$x = 4,86 \rightarrow Q_3 = 40 + 4,86 = 44,86$$



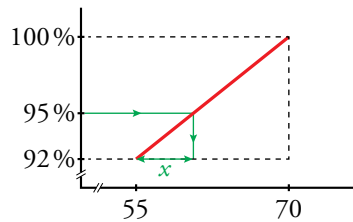
$$\frac{39 - 11}{30 - 11} = \frac{40 - 35}{x}$$

$$x = 3,39 \rightarrow p_{30} = 35 + 3,39 = 38,39$$



$$\frac{76 - 39}{60 - 39} = \frac{45 - 40}{x}$$

$$x = 2,84 \rightarrow p_{60} = 40 + 2,84 = 42,84$$



$$\frac{100 - 92}{95 - 92} = \frac{70 - 55}{x}$$

$$x = 5,63 \rightarrow p_{95} = 55 + 5,63 = 60,63$$

Reflexiona sobre la teoría

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

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

Aplicamos la definición de la media:

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

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

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

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

31. 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

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

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:

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

- DESVIACIÓN TÍPICA:

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

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

32. ▢ a) Justifica que la suma de las frecuencias relativas es siempre igual a 1.

b) Justifica también que la suma de las frecuencias porcentuales es 100.

a) Supongamos que tenemos n datos:

$$fr_1 + fr_2 + \dots + fr_k = \frac{f_1}{n} + \frac{f_2}{n} + \dots + \frac{f_k}{n} = \frac{f_1 + \dots + f_k}{n} = \frac{n}{n} = 1$$


b) Las frecuencias porcentuales son $\frac{f_i}{n} \cdot 100$:

$$\frac{f_1}{n} \cdot 100 + \dots + \frac{f_k}{n} \cdot 100 = \frac{(f_1 + \dots + f_k)}{n} \cdot 100 = 1 \cdot 100 = 100$$

33. ▢ La empresa A, con 500 trabajadores, tiene un ingreso anual medio por persona de 30 000 €, y la empresa B, con 750 trabajadores, tiene un ingreso anual medio de 25 000 €. Si las dos empresas se fusionan, ¿cuál será el ingreso anual medio tras la fusión?

$$\left. \begin{array}{l} 500 \cdot 30\,000 = 15\,000\,000 \\ 750 \cdot 25\,000 = 18\,750\,000 \end{array} \right\} \rightarrow 33\,750\,000 \text{ € y } 1\,250 \text{ personas}$$

$$\frac{33\,750\,000}{1\,250} = 27\,000 \text{ € será el ingreso anual medio tras la fusión.}$$

- 34.**  La nota media de los aprobados en un examen de matemáticas ha sido 6,8, y la de los suspensos, 3,5. Calcula la nota media de la clase sabiendo que hubo 35 aprobados y 15 suspensos.

$$6,8 = \frac{\text{Suma total notas aprobados}}{35} \rightarrow \text{Suma total notas probados} = 238$$

$$3,5 = \frac{\text{Suma total notas suspensos}}{15} \rightarrow \text{Suma total notas suspensos} = 52,5$$

$$\text{Nota media de la clase} = \frac{238 + 52,5}{35 + 15} = 5,81$$

- 35.**  La estatura media de los 38 estudiantes de una clase es de 168 cm. Las 17 chicas miden 162 cm de media. Calcula la media de los chicos.

$$162 = \frac{\text{Suma estatura chicas}}{17} \rightarrow \text{Suma estatura chicas} = 2754$$

$$168 = \frac{\text{Suma estatura chicos} + 2754}{38} \rightarrow \text{Suma estatura chicos} = 3630$$

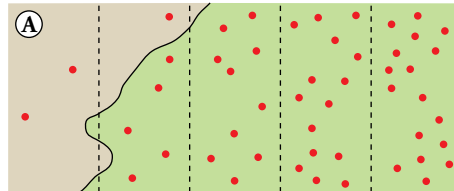
Por tanto, la media de los chicos será:

$$\frac{3630}{38 - 17} = \frac{3630}{21} = 172,86 \text{ cm}$$

Lee, resuelve y aprende por tu cuenta

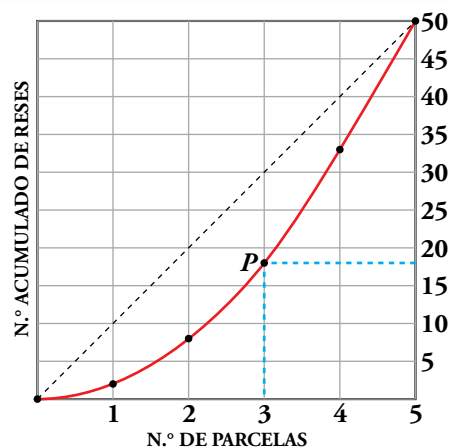
Terrenos, rebaños y gráficas

En la ilustración puedes ver una finca con su rebaño. Observa también que tiene una zona pantanosa en la que solo entran algunas reses despistadas.



En la gráfica que hay debajo se ha representado la distribución de los animales sobre el terreno.

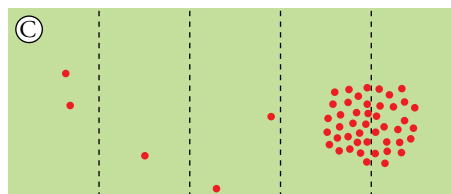
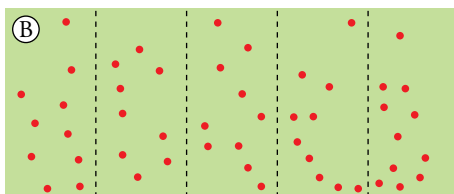
N.º DE PARCELAS	1	2	3	4	5
N.º ACUMULADO DE RESES	2	$2 + 6 = 8$	$8 + 10 = 18$	$18 + 15 = 33$	$33 + 17 = 50$



Así, por ejemplo, el punto $P(3, 18)$ indica que las 3 primeras parcelas (las más deshabitadas) albergan 18 reses.

- Dibuja, con el mismo criterio, las curvas relativas a las fincas B y C.

¿Cuál de las dos se acerca más a la diagonal del cuadrado?

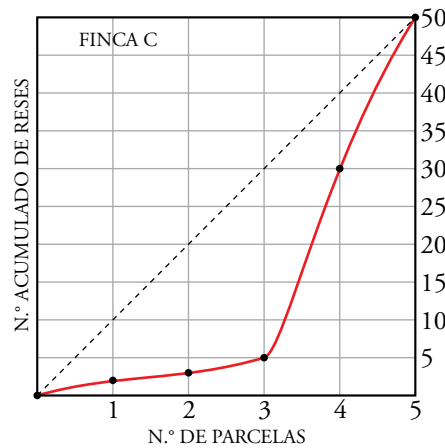
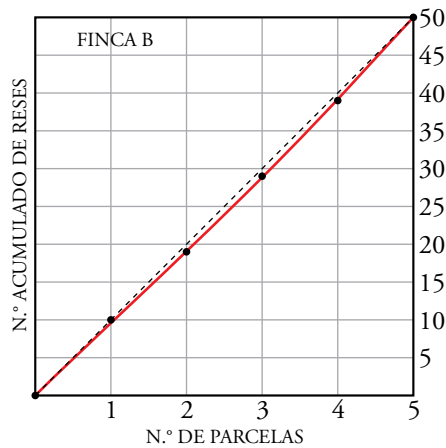


- Finca B

N.º DE PARCELAS	1	2	3	4	5
N.º ACUMULADO DE RESES	10	19	29	39	50

- Finca C

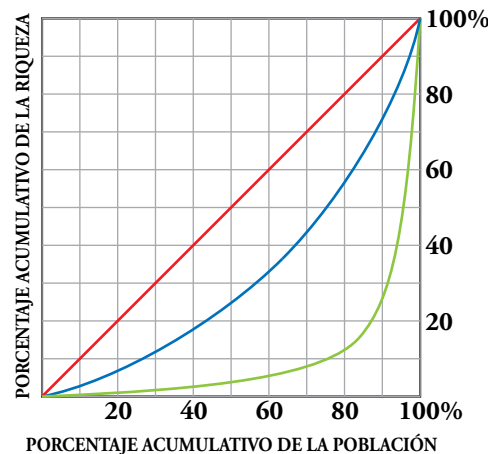
N.º DE PARCELAS	1	2	3	4	5
N.º ACUMULADO DE RESES	2	3	5	30	50



¿Cuántos abarcan cuánto?

- ¿Puedes identificar el país que corresponde a cada gráfica?

- A → País desarrollado → gráfica azul.
- B → País del Tercer Mundo → gráfica verde.
- C → País ideal → gráfica roja.



Sabías que...

- Si metieras en un bombo todas las letras de las dos líneas que estás leyendo y sacaras una al azar, ¿cuál de ellas tendría mayor probabilidad de ser elegida?

Contando las letras de las dos líneas, resulta: (a, 21); (e, 17); (s, 14); (l, 11); ...; a partir de aquí, el resto de letras, a simple vista, se observa que aparecen menos. Por tanto, la letra con mayor probabilidad en la extracción del bombo sería la “a”.

- ¿Qué letra es la más usada en castellano? Diseña un proyecto para averiguarlo.

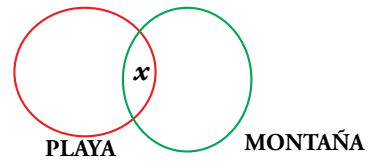
Para estimar la letra más usada en castellano, se sugiere apelar a la creatividad de los alumnos. Un camino podría ser:

- Abrir una novela actual por cualquier página, al azar.
- Asignar dos líneas a cada alumno de la clase, para que cada uno repita con ellas el mismo trabajo realizado en la actividad anterior.
- Reunir los datos recogidos por el grupo.
- Sacar conclusiones.

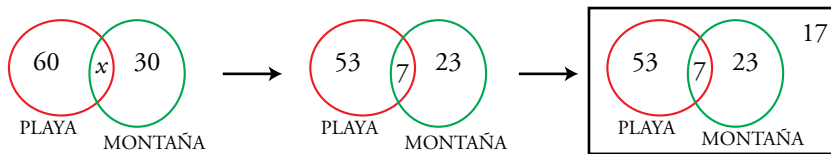
Entrénate resolviendo problemas

- Un encuestador ha preguntado a un grupo de personas sobre sus gustos a la hora de elegir el lugar para salir de vacaciones. Los resultados han sido:

- A 60 les gusta la playa.
- A 30 les gusta la montaña.
- A 7 les gustan ambos ambientes; es decir, la playa y la montaña.
- 17 manifiestan que no les gusta ni lo uno ni lo otro, por lo que no salen de vacaciones o van al extranjero.



¿Cuántas personas fueron encuestadas?



$$53 + 7 + 23 + 17 = 100$$

Se encuestó a 100 personas.

- En un zoológico existen dos tipos de visitas:

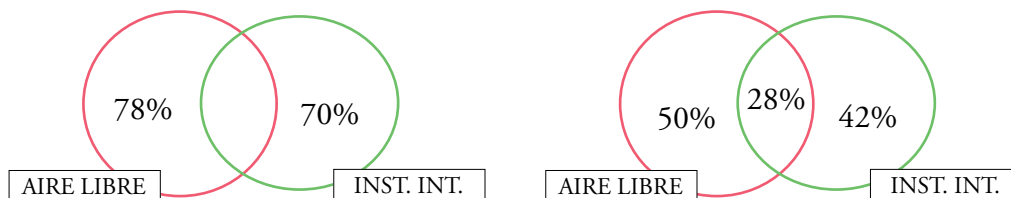
A. ANIMALES AL AIRE LIBRE B. INSTALACIONES INTERIORES

Se ha hecho un informe en el que se dice:

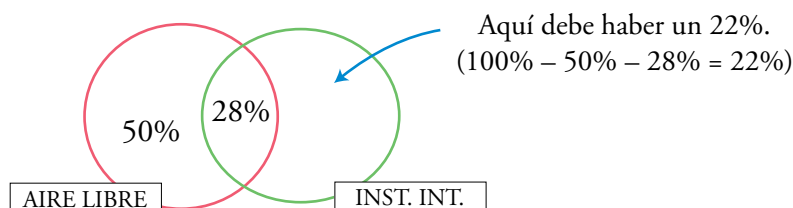
- El 78 % de los clientes visita a los animales que están al aire libre.
- El 70 % visita las instalaciones interiores.
- El 28 % efectúa ambas visitas.

El director, al leer el informe, pide a la persona que lo ha redactado que lo revise, pues indudablemente tiene que haber un error.

Se comprueba que el error está en el porcentaje de los asistentes a las instalaciones interiores. Corrige tú el error.



Es claro que hay un error, ya que $50\% + 28\% + 42\% = 120\%$.



Por tanto, las instalaciones interiores fueron visitadas por un $28\% + 22\% = 50\%$ (Hemos supuesto que todos los clientes visitaron al menos una de las dos instalaciones. En otro caso, no habrían pagado la entrada.)

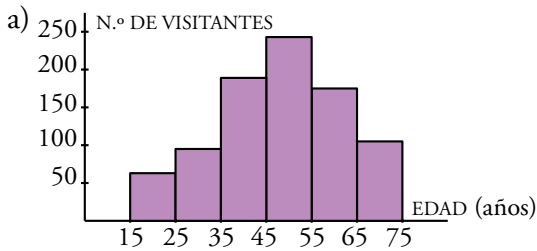
Autoevaluación

1. La edad de los visitantes de una exposición está recogida en la tabla de la derecha:

EDAD	N.º DE VIS.
15-25	63
25-35	95
35-45	189
45-55	243
55-65	175
65-75	105

a) Representa los datos en un gráfico adecuado.

b) Halla la media, la desviación típica y el C.V.



INTERVALO	x_i	f_i	$f_i x_i$	$f_i x_i^2$
15-25	20	63	1 260	25 200
25-35	30	95	2 850	85 500
35-45	40	189	7 560	302 400
45-55	50	243	12 150	607 500
55-65	60	175	10 500	630 000
65-75	70	105	7 350	514 500
		870	41 670	2 165 100

$$\bar{x} = \frac{41\,670}{870} = 47,90$$

$$\sigma = \sqrt{\frac{2\,165\,100}{870} - 47,90^2} = 13,94$$

$$\text{C.V.} = \frac{13,94}{47,90} = 0,29 \rightarrow 29\%$$

2. a) Calcula \bar{x} , σ , C.V, mediana, cuartiles y percentiles 20 y 80 de las notas de estos 20 estudiantes:

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

b) Representa los datos en un diagrama de barras.

x_i	f_i	$f_i x_i$	$f_i x_i^2$	F_i	% ACUM.
2	1	2	4	1	5
3	2	6	18	3	15
4	5	20	80	8	40
5	3	15	75	11	55
6	4	24	144	15	75
7	1	7	49	16	80
8	2	16	128	18	90
9	1	9	81	19	95
10	1	10	100	20	100
		109	679		

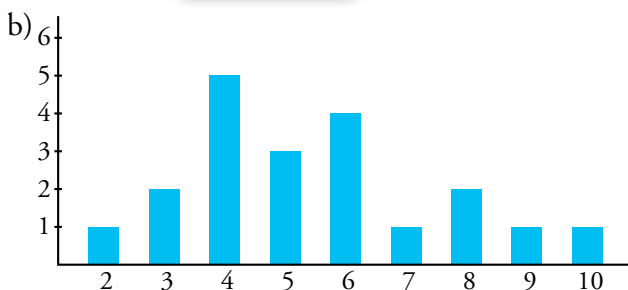
$$\bar{x} = \frac{109}{20} = 5,45$$

$$\sigma = \sqrt{\frac{679}{20} - 5,45^2} = 2,1$$

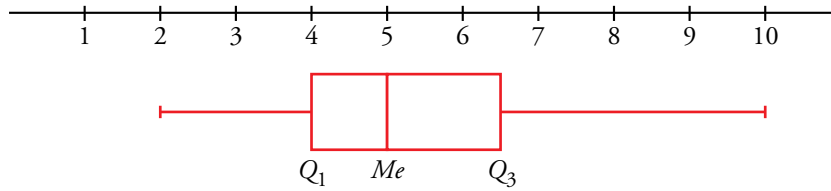
$$\text{C.V.} = \frac{\sigma}{\bar{x}} = \frac{2,1}{5,45} = 0,39$$

$$Me = 5 \quad Q_3 = 6,5 \quad p_{80} = 7,5$$

$$Q_1 = 4 \quad p_{20} = 4$$



3. A partir de los cuartiles y de la mediana de los datos del ejercicio anterior, representa un diagrama de caja y bigotes.



4. Calcula la mediana y los cuartiles de la siguiente distribución. Representalos en un diagrama de caja.

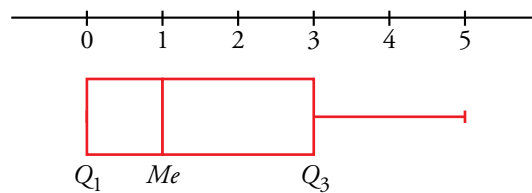
x_i	0	1	2	3	4	5
f_i	12	9	7	6	3	3

x_i	f_i	F_i	% ACUM.
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 F_i supera el 25% para $x_i = 0$.

$Q_3 = 3$, porque F_i supera el 75% para $x_i = 3$.

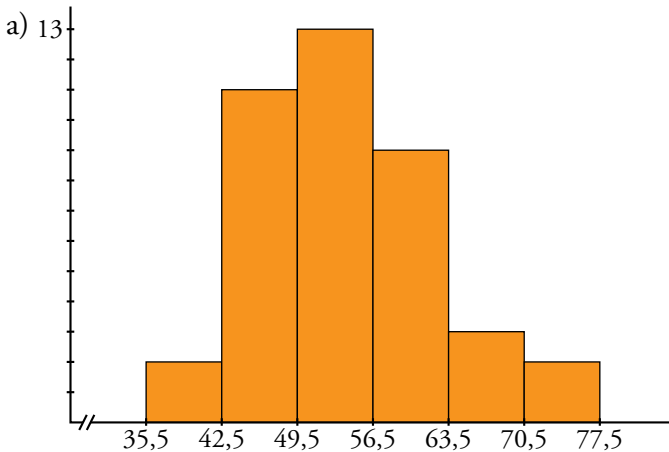


5. Esta tabla muestra los pesos de 40 estudiantes:

PESOS	FRECUENCIA
35,5-42,5	2
42,5-49,5	11
49,5-56,5	13
56,5-63,5	9
63,5-70,5	3
70,5-77,5	2

a) Representalos en un histograma.

b) Calcula la mediana y los cuartiles y estima los percentiles que corresponden a 40 kg, 60 kg y 70 kg.



b)

EXTREMOS	F_i	% ACUM.
35,5	0	0
42,5	2	5
49,5	13	32,5 ≈ 33
56,5	26	65
63,5	35	87,5 ≈ 88
70,5	38	95
77,5	40	100

Cálculo de la mediana, cuartiles y percentiles correspondientes a 40 kg, 60 kg y 70 kg.

$$\frac{65 - 33}{50 - 33} = \frac{56,5 - 49,5}{x} \rightarrow x = 3,72 \rightarrow Me = 49,5 + 3,72 = 53,22$$

$$\frac{33 - 5}{25 - 5} = \frac{49,5 - 42,5}{x} \rightarrow x = 5 \rightarrow Q_1 = 42,5 + 5 = 47,5$$

$$\frac{88 - 65}{75 - 65} = \frac{63,5 - 56,5}{x} \rightarrow x = 3,04 \rightarrow Q_3 = 56,5 + 3,04 = 59,54$$

Percentil que corresponde a 40 kg:

$$\frac{5 - 0}{x} = \frac{42,5 - 35,5}{40 - 35,5} \rightarrow x = 3,21 \rightarrow 0 + 3,21 = 3,21 \rightarrow p_3$$

Percentil que corresponde a 60 kg:

$$\frac{88 - 65}{x} = \frac{63,5 - 56,5}{60 - 56,5} \rightarrow x = 11,5 \rightarrow 65 + 11,5 = 76,5 \rightarrow p_{77}$$

Percentil que corresponde a 70 kg:

$$\frac{95 - 88}{x} = \frac{70,5 - 63,5}{70 - 63,5} \rightarrow x = 6,5 \rightarrow 88 + 6,5 = 94,5 \rightarrow p_{95}$$

6. Indica, en cada caso, si hay que recurrir a la población o tomar una muestra. Razona por qué.

- Estudio del número de suspensos de los estudiantes de 4.º ESO de un centro docente.
- Estudio del tiempo de caducidad de los cartones de leche de una fabrica de envasar.
- Estudio de las rentas anuales, gastos mensuales, hábitos... de la población de un cierto país.
- Encuesta electoral a pie de urna en unas elecciones.

En aquellos casos en los que haya que recurrir a muestras, describe una forma razonable de seleccionarlas.

- Población. Podemos comprobar el resultado de todos los individuos objeto de nuestro estudio.
- Muestra. No se puede controlar la caducidad de todos los cartones. Para seleccionar una muestra se eligen de forma aleatoria los cartones que se envasan y se hace un seguimiento de su tiempo de caducidad.
- Muestra. No se puede controlar toda la población de un país. Para seleccionar una muestra se toman del censo de forma aleatoria el número de habitantes a estudiar.
- Muestra. No se puede preguntar absolutamente a todas las personas que van a votar. Se pueden escoger aleatoriamente de entre los votantes de un conjunto representativo de colegios electorales.