

14 Estadística

PIENSA Y CONTESTA

Las variables que se representan en los gráficos estadísticos son de lo más variado. Y no todas son numéricas. Encuentra gráficos estadísticos en los que se representen características no cuantitativas.

Respuesta libre. Por ejemplo, en la representación del deporte practicado, de la marca de coche preferida, del medio de transporte utilizado para ir al instituto...se representan características no cuantitativas.

ANALIZA Y REFLEXIONA

En la actualidad es difícil abrir una revista o un periódico y no encontrar un gráfico estadístico. Pero son un invento relativamente reciente en la historia de las matemáticas. Poco más de dos siglos.

¿En qué áreas de la vida cotidiana aparecen a menudo gráficos estadísticos? Piensa en una en la que no se utilicen.

Respuesta libre. Por ejemplo, se utilizan en economía, en estudio del crecimiento de la población...

¿Por qué crees que las gráficas estadísticas tienen tanto éxito en actividades tan dispares?

Respuesta libre.

Y TÚ, ¿QUÉ OPINAS?

William Playfair creó distintas representaciones gráficas y sin embargo “su nombre no se encuentra ni siquiera en las grandes enciclopedias”. ¿Crees que es justo que se le reconozca por este hecho? ¿Por qué motivo crees que es justo que se valore o reconozca a una persona?

Respuesta libre.

Actividades propuestas

- Unos grandes almacenes quieren hacer un estudio sobre el grado de satisfacción de sus clientes. Para ello, seleccionan al azar, entre ellos, a 100 que han gastado menos de 1000 € el último año, otros 100 entre los que han gastado entre 1000 € y 5000 € y otros 100 entre los que han gastado más de 5000 €. ¿Es representativa la muestra?**

La población en estudio, los clientes, se ha dividido en tres estratos o partes según el dinero gastado en los almacenes. De cada estrato se han elegido al azar 100 individuos. Por tanto, se ha realizado un muestreo aleatorio estratificado.

La muestra será representativa si la proporción de individuos en cada estrato de la misma coincide con la proporción del estrato en la población. Es decir, la muestra será representativa si una tercera parte de la clientela se gastó menos de 1000 €, otra tercera parte se gastó entre 1000 € y 5000 €, y otro tercio de la población más de 5000 €.

- Clasifica las siguientes variables estadísticas:**
 - Número de goles en una jornada de la liga.**
 - Cotización en bolsa de una empresa en una semana.**
 - Profesiones con menos índice de paro.**
 - Causas de mortalidad en una población.**
 - Variable cuantitativa discreta.
 - Variable cuantitativa continua.
 - Variable cualitativa.
 - Variable cualitativa.

3. Los resultados de una encuesta a 40 jóvenes sobre el número de horas que utilizan una consola el fin de semana son:

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

- a) ¿De qué tipo es la variable estadística?
 b) Haz una tabla de frecuencias indicando la frecuencia absoluta, relativa, y las frecuencias acumuladas de cada dato.
 c) ¿Qué porcentaje utiliza la consola menos de 3 horas? ¿Y más de 6?
- a) La variable es cuantitativa discreta.
 b)

N.º de horas: x_i	f_i	h_i	F_i	H_i	H_i (%)
0	3	0,075	3	0,075	7,5
1	4	0,1	7	0,175	17,5
2	2	0,05	9	0,225	22,5
3	9	0,225	18	0,45	45
4	7	0,175	25	0,625	62,5
5	4	0,1	29	0,725	72,5
6	4	0,1	33	0,825	82,5
7	5	0,125	38	0,95	95
8	2	0,05	40	1	100
N = 40		1			

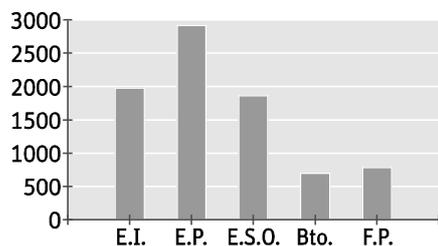
- c) El 22,5 % utiliza la consola menos de 3 horas, y el 100 % – 82,5 % = 17,5 %, más de 6 horas.

4. El número de alumnos, en miles, matriculados en enseñanzas no universitarias en España en el curso 2015/2016 fue:

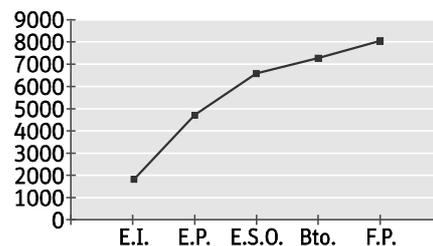
E. Infantil	1978
E. Primaria	2918
E.S.O.	1865
Bachillerato	698
F. Profesional	784

- a) Construye el diagrama de barras de la distribución escolar y el polígono de frecuencias acumuladas.
 b) Si los datos en 2012/2013 fueron respectivamente 1900, 2827, 1806, 692 y 663, haz un diagrama lineal con los datos de ambos cursos y comenta la evolución.

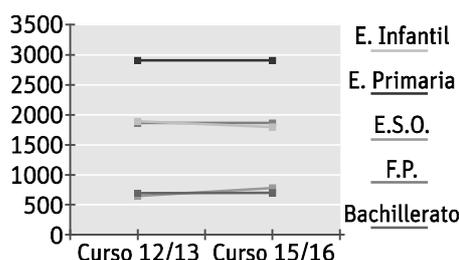
- a) Diagrama de barras



- Polígono de frecuencias acumuladas



- b) Diagrama lineal



El número de alumnos en Educación Primaria se ha mantenido constante y ha sido muy superior a los matriculados en otras enseñanzas.

Tanto el número de matriculados en E.S.O. como en Educación Infantil ha disminuido ligeramente, al contrario de lo que ha ocurrido con los matriculados en Formación Profesional y Bachillerato.

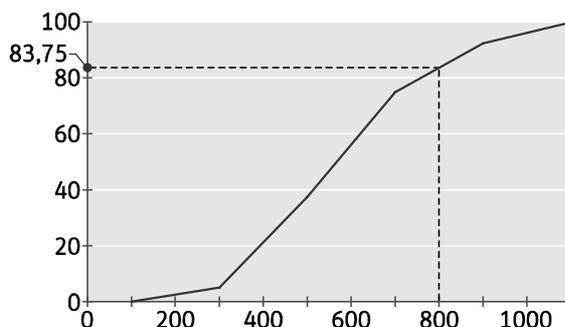
5. En una fábrica de bombillas se estudia la vida de un tipo de bombilla. Se ha tomado una muestra de 200 lámparas con los siguientes resultados.

Vida en horas	N.º de bombillas
[100, 300)	10
[300, 500)	65
[500, 700)	75
[700, 900)	35
[900, 1100)	15

Dibuja el polígono de frecuencias acumuladas en porcentaje. ¿Cuál es el porcentaje de bombillas que dura más de 800 horas?

Intervalos	Marca: x_i	f_i	h_i	F_i	H_i	H_i (%)
[100, 300)	200	10	0,05	10	0,05	5
[300, 500)	400	65	0,325	75	0,375	37,5
[500, 700)	600	75	0,375	150	0,75	75
[700, 900)	800	35	0,175	185	0,925	92,5
[900, 1100)	1000	15	0,075	200	1	100
		$N = 200$	1			

Como 800 horas es la mitad del 4 intervalo, se busca el tanto por ciento acumulado correspondiente a la mitad de ese intervalo, es decir, 83,75 %. Este resultado se puede ver de forma aproximada en la gráfica.



El $100\% - 83,75\% = 16,25\%$ de las bombillas duran más de 800 horas.

6. Actividad resuelta.

7. La edad de los 50 socios de un club deportivo juvenil viene dado por la tabla:

Edad	16	17	18	19	20	21
N.º	5	8	10	12	9	6

Halla la moda, la media y la mediana y los cuartiles.

Edad: x_i	f_i	F_i	$x_i \cdot f_i$
16	5	5	80
17	8	13	136
18	10	23	180
19	12	35	228
20	9	44	180
21	6	50	126
		$N = 50$	930

$$\bar{x} = \frac{930}{50} = 18,6 \quad M_0 = 19$$

El 25 % de 50 es 12,5. El primer cuartil es $Q_1 = 17$, ya que es el valor que deja por debajo al 25 % de los datos.

El 50 % de 50 es 25. El segundo cuartil es $Q_2 = M = 19$.

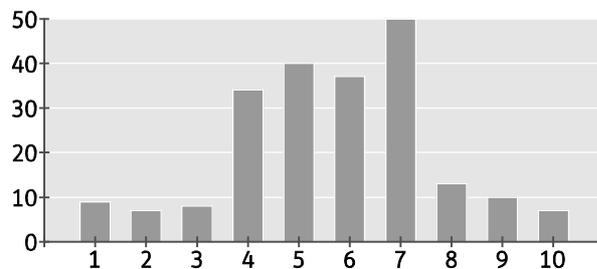
El 75 % de 50 es 37,5. El tercer cuartil es $Q_3 = 20$.

8. La evaluación de un test realizado a los 215 trabajadores de una empresa ha sido:

Nota	1	2	3	4	5	6	7	8	9	10
N.º Trab.	9	7	8	34	40	37	50	13	10	7

- Completa la tabla de frecuencias y representa la distribución con un diagrama de barras.
- Calcula la moda, la media, la mediana y los cuartiles.
- Deberán hacer un curso de formación los trabajadores por debajo del tercer decil. ¿Cuántos son?

Nota: x_i	f_i	h_i	F_i	H_i	H_i (%)	$x_i \cdot f_i$
1	9	0,042	9	0,042	4,186	9
2	7	0,033	16	0,074	7,442	14
3	8	0,037	24	0,112	11,163	24
4	34	0,158	58	0,270	26,977	136
5	40	0,186	98	0,456	45,581	200
6	37	0,172	135	0,628	62,791	222
7	50	0,233	185	0,860	86,047	350
8	13	0,060	198	0,921	92,093	104
9	10	0,047	208	0,967	96,744	90
10	7	0,033	215	1,000	100	70
	$N = 215$	1				1219



- $M_0 = 7$

$$\bar{x} = \frac{1219}{215} = 5,67$$

25 % de 215 = 53,78 $\Rightarrow Q_1 = 4$, 50 % de 215 = 107,5 $\Rightarrow Q_2 = M = 6$ y 75 % de 215 = 161,25 $\Rightarrow Q_3 = 8$
- El 30 % de 215 es 64,5. El tercer decil es $D_3 = 5$.

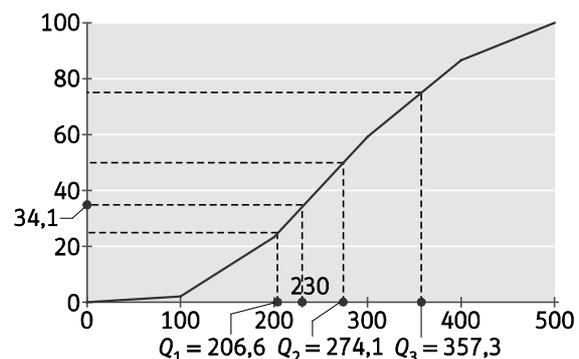
Deberán hacer un curso de formación todos los trabajadores que hayan obtenido menos de un 5, que son 58 en total.

9. En un saco de patatas, su masa se distribuye de la forma:

Peso	[0, 100)	[100, 200)	[200, 300)	[300, 400)	[400, 500)
Patatas	3	32	54	41	20

- Dibuja el gráfico de frecuencias acumuladas en porcentaje.
- Halla la mediana y los cuartiles.
- ¿Qué número aproximado de patatas pesarán menos de 230 g?

Extremos	F_i	H_i (%)
100	3	2
200	35	23,33
300	89	59,33
400	130	86,67
500	150	100



- $Q_1 = 206,6$; $M = Q_2 = 274,1$; $Q_3 = 357,32$.
- Aproximadamente, el 34,1 % de las patatas pesarán menos de 230 g.

10. El número de libros solicitados por los usuarios de una biblioteca ha sido:

Libros	1	2	3	4	5	6
Usuarios	8	12	9	6	3	2

- a) Calcula el recorrido y el recorrido intercuartílico.
 b) Halla la varianza y la desviación típica.

a)

Libros:	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
1	8	8	8
2	12	24	48
3	9	27	81
4	6	24	96
5	3	15	75
6	2	12	72
	$N = 40$	110	380

$$\text{Recorrido} = 6 - 1 = 5$$

$$\text{El } 25\% \text{ de } 40 = 10 \Rightarrow Q_1 = 2, \text{ el } 50\% \text{ de } 40 = 20 \Rightarrow Q_2 = M = 2 \text{ y el } 75\% \text{ de } 40 = 30 \Rightarrow Q_3 = 4$$

$$\text{Recorrido intercuartílico} = Q_3 - Q_1 = 4 - 2 = 2$$

b) $\bar{x} = \frac{110}{40} = 2,75 \Rightarrow s^2 = \frac{380}{40} - 2,75^2 = 1,94 \Rightarrow s = 1,39$

11. Las notas de 10 alumnos en una prueba han sido: 6, 5, 3, 6, 3, 7, 5, 8, 5, 4.

- a) Halla el recorrido y el recorrido intercuartílico.
 b) Calcula la desviación típica y el coeficiente de variación.
 c) El profesor decide puntuar sobre 20 y para ello multiplica las notas por 2. ¿Qué les ocurre a los tres parámetros estadísticos? ¿En qué caso hay mayor dispersión relativa?

Notas: 3, 3, 4, 5, 5, 5, 6, 6, 7, 8.

a) Recorrido = $8 - 3 = 5$

$$\text{El } 25\% \text{ de } 10 = 2,5 \Rightarrow Q_1 = 4, \text{ el } 50\% \text{ de } 10 = 5 \Rightarrow Q_2 = M = 5 \text{ y el } 75\% \text{ de } 10 = 7,5 \Rightarrow Q_3 = 6$$

$$\text{Recorrido intercuartílico} = Q_3 - Q_1 = 6 - 4 = 2$$

b) $\bar{x} = \frac{3 \cdot 2 + 4 + 5 \cdot 3 + 6 \cdot 2 + 7 + 8}{10} = 5,2 \Rightarrow s^2 = \frac{3^2 \cdot 2 + 4^2 + 5^2 \cdot 3 + 6^2 \cdot 2 + 7^2 + 8^2}{10} - 5,2^2 = 2,36 \Rightarrow s = 1,54$

$$CV = \frac{1,54}{5,2} = 0,3 \rightarrow 30\%$$

c) Notas: 6, 6, 8, 10, 10, 10, 12, 12, 14, 16.

$$\text{Recorrido} = 16 - 6 = 10 \Rightarrow \text{El recorrido queda multiplicado por } 2.$$

$$\text{El } 25\% \text{ de } 10 = 2,5 \Rightarrow Q_1 = 8 \Rightarrow Q_1 \text{ queda multiplicado por } 2.$$

$$\text{El } 50\% \text{ de } 10 = 5 \Rightarrow Q_2 = M = 10 \Rightarrow Q_2 \text{ queda multiplicado por } 2.$$

$$\text{El } 75\% \text{ de } 10 = 7,5 \Rightarrow Q_3 = 12 \Rightarrow Q_3 \text{ queda multiplicado por } 2.$$

$$\text{Recorrido intercuartílico} = Q_3 - Q_1 = 12 - 8 = 4 \Rightarrow \text{El rango intercuartílico queda multiplicado por } 2.$$

$$\bar{x} = \frac{6 \cdot 2 + 8 + 10 \cdot 3 + 12 \cdot 2 + 14 + 16}{10} = 10,4 \Rightarrow \text{La media queda multiplicada por } 2.$$

$$s^2 = \frac{6^2 \cdot 2 + 8^2 + 10^2 \cdot 3 + 12^2 \cdot 2 + 14^2 + 16^2}{10} - 10,4^2 = 9,44 \Rightarrow \text{La varianza queda multiplicada por } 4.$$

$$s = \sqrt{9,44} = 3,08 \Rightarrow \text{La desviación típica queda multiplicada por } 2.$$

$$CV = \frac{3,08}{10,4} = 0,3 \rightarrow 30\% \Rightarrow \text{El coeficiente de variación no varía, por tanto hay igual dispersión en los dos casos.}$$

12. Se han seleccionado dos pruebas A y B de 120 preguntas tipo test cada una. ¿En cuál es más representativa la media?

- En la prueba A la media de respuestas correctas ha sido 72, y la desviación típica, 15.
- En la prueba B la media ha sido 58, y la desviación típica, 13.

Se calculan los coeficientes de variación de las dos pruebas:

$$\text{Prueba A: } CV = \frac{15}{72} = 0,2083 \Rightarrow 20,83 \% \qquad \text{Prueba B: } CV = \frac{13}{58} = 0,2241 \Rightarrow 22,41 \%$$

Como $20,83 \% < 22,41 \%$, la media es más representativa en la prueba A.

13. Para medir la eficacia de un abono se han medido el crecimiento de 20 plantas en un mes.

Altura (cm)	[0, 10)	[10, 20)	[20, 30)	[30, 40)
N.º plantas	4	7	5	4

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

b) ¿Entre qué valores se encuentra el 68 % de los datos? ¿Y el 95 %?

a) Recorrido = $40 - 0 = 40$

Intervalos	Marca: x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[0, 10)	5	4	20	100
[10, 20)	15	7	105	1575
[20, 30)	25	5	125	3125
[30, 40)	35	4	140	4900
		$N = 20$	390	9700

$$\bar{x} = \frac{390}{20} = 19,5 \Rightarrow s^2 = \frac{9700}{20} - 19,5^2 = 104,75 \Rightarrow s = 10,23 \Rightarrow CV = \frac{10,23}{19,5} = 0,5246 \rightarrow 52,46 \%$$

b) El 68 % de los datos se encuentran entre $19,5 - 10,23 = 9,27$ y $19,5 + 10,23 = 29,73$.

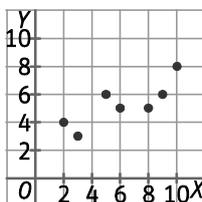
El 95 % de los datos se encuentran entre 0 y $19,5 + 2 \cdot 10,23 = 39,96$.

14. Actividad resuelta.

15. Dada la siguiente distribución bidimensional.

X	2	3	5	6	8	9	10
Y	4	3	6	5	5	6	8

Representa la nube de puntos e indica el tipo de correlación entre las variables.



Existe una correlación fuerte directa entre X e Y.

16. Las puntuaciones obtenidas por 40 personas en dos test que miden la comprensión lectora (X) y el cálculo numérico (Y) han sido:

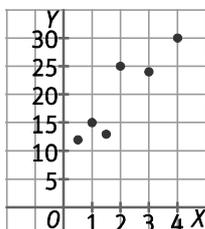
X \ Y	20	30	40	50
20	3	4	2	1
30	5	2	2	1
40	1	3	4	2
50	0	1	3	6

- a) ¿Cuántas personas han obtenido 30 puntos en comprensión lectora? ¿Y 50 puntos en cálculo numérico?
- b) ¿Hay correlación entre ambas habilidades?
- a) Han obtenido 30 puntos en comprensión lectora 10 personas.
Han obtenido 50 puntos en cálculo numérico 10 personas.
- b) Sí que hay correlación entre ambas habilidades porque, por lo general, las personas que han obtenido mejor nota en comprensión lectora también han obtenido mejor puntuación en cálculo numérico.

17. Se quiere estudiar la relación entre las ventas de un producto y el espacio destinado a su exposición. Se anotan las ventas, en unidades, y la longitud de los expositores, en metros, durante seis semanas:

Semana	1. ^a	2. ^a	3. ^a	4. ^a	5. ^a	6. ^a
Metros	0,5	1	1,5	2	3	4
Unidades	12	15	13	25	24	30

- a) Representa la nube de puntos e indica qué tipo de correlación existe.
- b) Calcula el centro de gravedad de la distribución.
- c) Calcula la media y la desviación típica de cada variable.
- d) Halla la covarianza y el coeficiente de correlación.
- a) Como los puntos de la nube se ajustan a una recta creciente, entonces existe correlación fuerte directa.



- b) Llamando X a los metros e Y a las unidades:

$$\bar{x} = \frac{0,5+1+1,5+2+3+4}{6} = \frac{12}{6} = 2 \quad \bar{y} = \frac{12+15+13+25+24+30}{6} = \frac{119}{6} = 19,83$$

El centro de gravedad de la distribución es (2; 19,83).

- c) Completamos la tabla para calcular los parámetros.

Semana	x_i	y_i	x_i^2	y_i^2	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
1. ^a	0,5	12	0,25	144	-1,5	-7,83	11,75
2. ^a	1	15	1	225	-1	-4,83	4,83
3. ^a	1,5	13	2,25	169	-0,5	-6,83	3,42
4. ^a	2	25	4	625	0	5,17	0,00
5. ^a	3	24	9	576	1	4,17	4,17
6. ^a	4	30	16	900	2	10,17	20,33
			32,5	2639			44,50

$$s_x^2 = \frac{32,5}{6} - 2^2 = 1,412 \Rightarrow s_x = 1,19 \text{ y } s_y^2 = \frac{2639}{6} - 19,83^2 = 46,6 \Rightarrow s_y = 6,83$$

- d) $s_{x,y} = \frac{44,50}{6} = 7,412 \Rightarrow r = \frac{7,412}{1,19 \cdot 6,83} = 0,912$

18. La tabla muestra los datos de importaciones y exportaciones de los años 2007 a 2014, en miles de millones de euros.

Año	08	09	10	11	12	13	14
X (I)	270	210	247	271	263	256	270
Y(E)	191	163	192	220	230	238	245

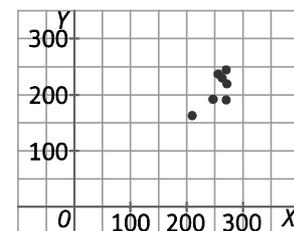
Dibuja la nube de puntos y calcula el coeficiente de correlación.

Año	x_i	y_i	x_i^2	y_i^2	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
08	270	191	72 900	36 481	268	-20,29	-5436,57
09	210	163	44 100	26 569	208	-48,29	-10 043,43
10	247	192	61 009	36 864	245	-19,29	-4725,00
11	271	220	73 441	48 400	269	8,71	2344,14
12	263	230	69 169	52 900	261	18,71	4884,43
13	256	238	65 536	56 644	254	26,71	6785,43
14	270	245	72 900	60 025	268	33,71	9035,43
	1787	1479	459 055	317 883			2844,43

$$\bar{x} = \frac{1787}{7} = 255,29 \Rightarrow s_x^2 = \frac{459\,055}{7} - 255,29^2 = 406,30 \Rightarrow s_x = 20,16$$

$$\bar{y} = \frac{1479}{7} = 211,29 \Rightarrow s_y^2 = \frac{317\,883}{7} - 211,29^2 = 768,39 \Rightarrow s_y = 27,72$$

$$s_{x,y} = \frac{2844,43}{7} = 406,35 \Rightarrow r = \frac{406,35}{20,16 \cdot 27,72} = 0,727$$



19. Se ha preguntado a 10 viandantes sobre el número de veces que han ido al teatro y al cine en el último año.

Teatro	0	1	2	3	3	4	5	5	6	7
Cine	6	12	15	10	5	8	6	9	4	2

¿Hay correlación entre las dos variables? Justifica tu respuesta.

Para comprobar si existe correlación entre el teatro, X, y el cine, Y, calculamos el coeficiente de correlación lineal.

x_i	y_i	x_i^2	y_i^2	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
0	6	0	36	-3,6	-1,7	6,12
1	12	1	144	-2,6	4,3	-11,18
2	15	4	225	-1,6	7,3	-11,68
3	10	9	100	-0,6	2,3	-1,38
3	5	9	25	-0,6	-2,7	1,62
4	8	16	64	0,4	0,3	0,12
5	6	25	36	1,4	-1,7	-2,38
5	9	25	81	1,4	1,3	1,82
6	4	36	16	2,4	-3,7	-8,88
7	2	49	4	3,4	-5,7	-19,38
36	77	174	731			-45,2

$$\bar{x} = \frac{36}{10} = 3,6 \Rightarrow s_x^2 = \frac{174}{10} - 3,6^2 = 4,44 \Rightarrow s_x = 2,11 \quad \bar{y} = \frac{77}{10} = 7,7 \Rightarrow s_y^2 = \frac{731}{10} - 7,7^2 = 13,81 \Rightarrow s_y = 3,72$$

$$s_{x,y} = \frac{-45,2}{10} = -4,52 \Rightarrow r = \frac{-4,52}{2,11 \cdot 3,72} = -0,576 \Rightarrow \text{Hay correlación inversa no muy fuerte.}$$

20. Los datos de importaciones y exportaciones de un país entre los años 2008 a 2014 en miles de millones de euros son:

I	270	210	247	271	263	256	270
E	191	163	192	220	230	238	245

Calcula la recta de regresión de Y sobre X.

I: x_i	E: y_i	x_i^2	y_i^2	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
270	191	72 900	36 481	14,71	-20,29	-298,49
210	163	44 100	26 569	-45,29	-48,29	2186,65
247	192	61 009	36 864	-8,29	-19,29	159,80
271	220	73 441	48 400	15,71	8,71	136,94
263	230	69 169	52 900	7,71	18,71	144,37
256	238	65 536	56 644	0,71	26,71	19,08
270	245	72 900	60 025	14,71	33,71	496,08
1787	1479	459 055	317 883			2844,43

$$\bar{x} = \frac{1787}{7} = 255,285 \Rightarrow s_x^2 = \frac{459\,055}{7} - 255,285^2 = 408,85 \Rightarrow s_x = 20,22$$

$$\bar{y} = \frac{1479}{7} = 211,29 \Rightarrow s_y^2 = \frac{317\,883}{7} - 211,29^2 = 768,39 \Rightarrow s_y = 27,72$$

$$s_{x,y} = \frac{2844,43}{7} = 406,35 \Rightarrow y - 211,29 = \frac{406,35}{408,85}(x - 255,285) \Rightarrow y = 0,994x - 42,43$$

La recta de regresión es $y = 0,994x - 42,43$.

21. Las edades de los actores que han protagonizado "Romeo y Julieta" en las últimas 8 representaciones de un teatro han sido:

Romeo	27	32	39	37	45	38	43	25
Julieta	23	30	32	40	39	38	32	34

Dibuja la nube de puntos y calcula la recta de regresión. ¿Cuál es la edad esperada de Julieta si Romeo tiene 36 años?

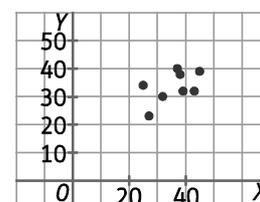
Llamamos X a la edad de Romeo e Y a la de Julieta.

x_i	y_i	x_i^2	y_i^2	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
27	23	729	529	-8,75	-10,50	91,88
32	30	1024	900	-3,75	-3,50	13,13
39	32	1521	1024	3,25	-1,50	-4,88
37	40	1369	1600	1,25	6,50	8,13
45	39	2025	1521	9,25	5,50	50,88
38	38	1444	1444	2,25	4,50	10,13
43	32	1849	1024	7,25	-1,50	-10,88
25	34	625	1156	-10,75	0,50	-5,38
286	268	10586	9198			153,01

$$\bar{x} = \frac{286}{8} = 35,75 \Rightarrow s_x^2 = \frac{10586}{8} - 35,75^2 = 45,19 \Rightarrow s_x = 6,72$$

$$\bar{y} = \frac{268}{8} = 33,5 \Rightarrow s_y^2 = \frac{9198}{8} - 33,5^2 = 27,5 \Rightarrow s_y = 5,24$$

$$s_{x,y} = \frac{153,01}{8} = 19,13 \Rightarrow y - 33,5 = \frac{19,13}{45,19}(x - 35,75) \Rightarrow y = 0,42x + 18,37$$



La recta de regresión es $y = 0,42x + 18,37$.

Si Romeo tiene 36 años, la edad esperada de Julieta será $y = 0,42 \cdot 36 + 18,37 = 33,5$ años.

22. Emprende

Se quiere realizar un estudio sobre el tipo de mascotas que tiene la población de una ciudad de 10 000 habitantes. Se proponen varias alternativas para seleccionar una muestra de 500 ciudadanos:

- Muestra aleatoria utilizando el censo municipal.
- Muestra estratificada por edad de forma aleatoria.
- Muestra aleatoria estratificada por barrios de la ciudad donde viven.

- a) ¿Qué muestra consideras más representativa?
- b) Se quiere hacer un estudio sobre el gasto que supone tener una mascota. Diseña el tipo de muestreo que consideres más adecuado y compáralo con el de tu compañero. ¿Qué muestra es más representativa?
- a) La muestra más representativa sería la muestra estratificada por edad.
- b) Respuesta libre.

23. Un supermercado realiza un estudio estadístico de sus clientes para conocer:

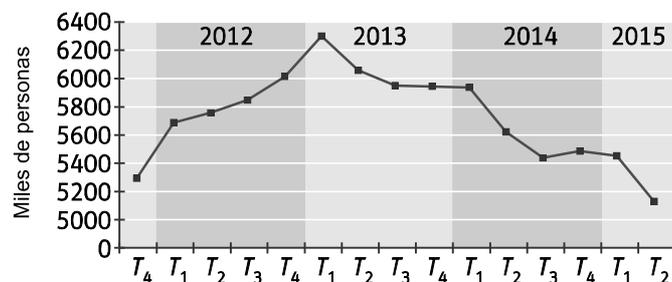
- a) El número de clientes cada hora.
- b) El importe de la compra de cada cliente.
- c) Medio de transporte utilizado para llegar.

Indica de qué tipo es cada una de estas variables.

- a) Variable cuantitativa discreta.
- b) Variable cuantitativa continua.
- c) Variable cualitativa.

24. La Encuesta de Población Activa (EPA), mide la tasa de parados y de población ocupada. La realiza el Instituto Nacional de Estadística (INE) sobre una muestra de unas 200 000 personas cada trimestre. La población total del estudio (población activa) asciende a 23 015 500 personas.

El gráfico muestra la evolución del número de parados desde 2011 a 2015.



- a) ¿Qué porcentaje de la población total forma la muestra? ¿Es representativa?
- b) ¿De qué tipo es el gráfico estadístico?
- c) Los puntos representan trimestres. ¿Cuándo se alcanza el máximo? ¿En cuánto aumentó el paro entre el 4.º trimestre de 2011 y el máximo del número de parados?
- d) Indica de forma aproximada la diferencia de parados entre el principio y el final de la serie.

- a) Cada trimestre, la muestra la forma el $\frac{200\,000}{23\,015\,500} = 0,0087 \Rightarrow 0,87\%$ de la población.

La muestra es representativa.

- b) El gráfico estadístico es un diagrama lineal.

- c) El máximo se alcanzó en el primer trimestre del 2013, con 6300 parados.

Entre el 4.º trimestre de 2011 y el máximo del número de parados, el paro aumentó en $6300 - 5300 = 1000$, es decir, en 1 millón de personas.

- d) Entre el principio y el final de la serie, el número de parados pasó de ser de 5300 mil personas a 5100.

El número de parados disminuyó en 200 mil personas.

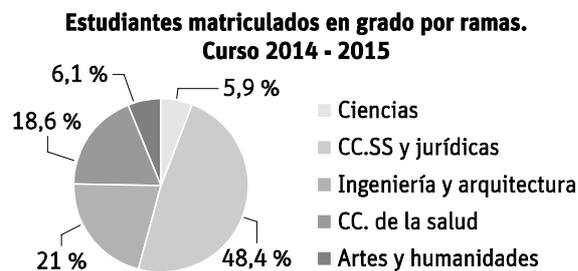
25. Durante el mes de junio las temperaturas máximas en una ciudad han sido (en grados centígrados):
28, 27, 28, 29, 25, 24, 23, 25, 27, 30, 29, 28, 29, 30, 26, 28, 31, 30, 30, 27, 29, 28, 30, 29, 30, 31, 32, 29, 30, 31

- Define la variable estadística. ¿De qué tipo es?
- Completa la tabla de frecuencias relativas y acumuladas.
- ¿Tiene sentido agrupar los datos en intervalos? Justifica tu respuesta.
 - La variable estadística "Temperatura máxima en una ciudad en junio" es cuantitativa discreta.

Temperaturas: x_i	f_i	h_i	F_i	H_i	H_i (%)
23	1	0,03	1	0,03	3,33
24	1	0,03	2	0,07	6,67
25	2	0,07	4	0,13	13,33
26	1	0,03	5	0,17	16,67
27	3	0,10	8	0,27	26,67
28	5	0,17	13	0,43	43,33
29	6	0,20	19	0,63	63,33
30	7	0,23	26	0,87	86,67
31	3	0,10	29	0,97	96,67
32	1	0,03	30	1	100
$N = 30$		1			

- No tiene sentido agrupar los datos en intervalos, porque se trata de una variable cuantitativa discreta con pocos datos diferentes.

26. Según datos del Ministerio de Educación, Cultura y Deporte (MECyD) los estudiantes matriculados en grados universitarios en 2014/2015 fueron 1 373 000, distribuidos así:



- Calcula la amplitud de los sectores correspondientes a los estudios de Ciencias y Ciencias de la salud.
- ¿Cuántos estudiantes cursan Ingeniería y Arquitectura?
 - Ciencias: $5,9 \% \text{ de } 360^\circ = 21,24^\circ$
Ciencias de la Salud: $18,6 \% \text{ de } 360^\circ = 66,96^\circ$
 - Cursan Ingeniería y Arquitectura $21 \% \text{ de } 1\,373\,000 = 288\,330$ estudiantes.

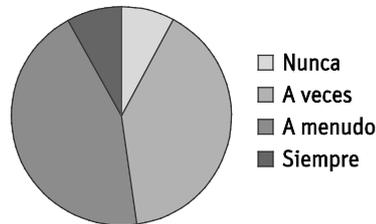
27. Se ha interrogado a 50 personas para saber si pasaban sus vacaciones en el extranjero. Los resultados han sido:

Extranjero	Nunca	A veces	A menudo	Siempre
Número	4	20	22	4

- a) ¿De qué tipo es la variable?
 b) Construye la tabla de frecuencias acumuladas.
 c) Representa la distribución. ¿Qué tipo de gráfico es el más aconsejable?
- a) La variable es cualitativa.
 b)

Extranjero	f_i	F_i	h_i	H_i (%)	Amplitud
Nunca	4	4	0,08	0,08	28,8
A veces	20	24	0,4	0,48	144
A menudo	22	46	0,44	0,92	158,4
Siempre	4	50	0,08	1	28,8
	$N = 50$		1		

- c) Representamos la variable en un diagrama de sectores.



28. Se recomienda ingerir las calorías que necesita una persona de la siguiente forma:

Hidratos de carbono	Grasas	Proteínas
50 %	35 %	15 %

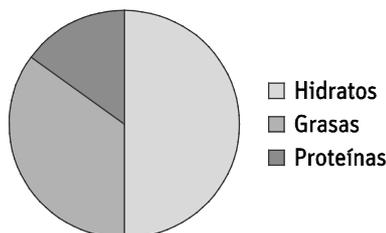
Representa los datos en un diagrama de sectores.

Amplitud de los sectores:

Hidratos de carbono: 50 % de $360^\circ = 180^\circ$

Grasas: 35 % de $360^\circ = 126^\circ$

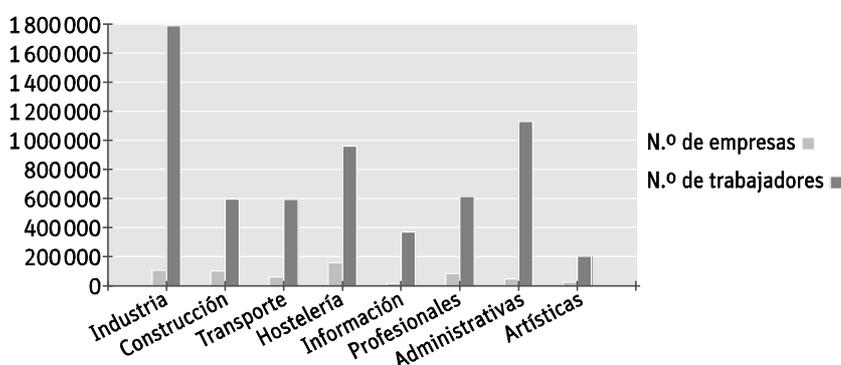
Proteínas: 15 % de $360^\circ = 54^\circ$



29. Según datos del Ministerio de empleo y seguridad social sobre actividad económica, el número de empresas y de trabajadores por sectores en 2014.

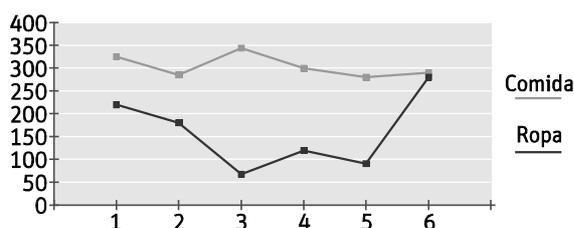
ACTIVIDAD ECONÓMICA	Nº de empresas	Nº de trabajadores
Industria	107 195	1 795 103
Construcción	105 103	597 357
Transporte y almacenamiento	63 407	594 952
Hostelería	159 488	962 144
Información, comunicaciones y actividades financieras, e inmobiliarias	18 753	369 762
Actividades profesionales, científicas y técnicas	86 862	617 004
Actividades administrativas y servicios auxiliares	47 894	1 133 914
Actividades artísticas, recreativas y de entretenimiento.	24 858	205 011

- a) ¿Cuál es el mejor tipo de gráfico para representar la participación de cada sector?
 b) Utiliza una hoja de cálculo para hacer los gráficos de cada columna. ¿Hay correspondencia entre ellos?
 a) El mejor tipo de gráfico para representar cada sector es un diagrama de barras.
 b)



30. Representa en el gráfico más adecuado el gasto mensual de una familia en comida y en ropa en los últimos 6 meses.

Comida	325	285	345	300	280	290
Ropa	220	180	68	120	90	280



31. Una empresa de marketing hace una encuesta sobre un producto entre 20 personas. Han de valorar el producto de 1 a 5:

1 = muy malo 2 = malo 3 = regular 4 = bueno 5 = muy bueno.

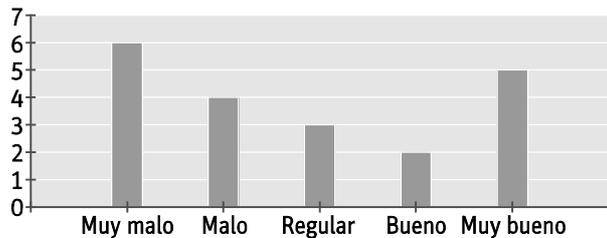
Los resultados obtenidos han sido: 1, 3, 3, 4, 1, 2, 2, 5, 1, 4, 5, 1, 5, 5, 1, 1, 2, 2, 3, 5

a) Halla la media, la moda y la mediana de las valoraciones.

b) Elige el gráfico adecuado para representar los resultados.

a) $\bar{x} = \frac{6 \cdot 1 + 4 \cdot 2 + 3 \cdot 3 + 2 \cdot 4 + 5 \cdot 5}{20} = 2,8$ $M_o = 1$ $M = \frac{2+3}{2} = 2,5$

b)



32. Las notas obtenidas por Juan en los exámenes escritos en matemáticas a lo largo de un curso han sido:

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

a) Calcula la moda, la media y la mediana.

b) Halla los cuartiles.

Notas: 3, 4, 5, 5, 6, 6, 6, 6, 7, 7, 7, 8

a) $\bar{x} = \frac{3+4+2 \cdot 5+4 \cdot 6+3 \cdot 7+8}{12} = 5,83$ $M_o = 6$ $M = 6$

b) $Q_1 = 5$ $Q_2 = 6$ $Q_3 = 7$

33. La cantidad mensual de lluvia recogida por un observatorio a lo largo de un año ha sido:

E	F	M	A	M	J	JL	A	S	O	N	D
182	170	137	117	149	80	40	122	133	156	198	182

a) ¿Se trata de un clima seco o húmedo?

b) Calcula la media y la mediana.

c) Halla los cuartiles.

Lluvias: 40, 80, 117, 122, 133, 137, 149, 156, 170, 182, 182, 198.

a) Por la cantidad de lluvia recogida mensualmente, se trata de un clima húmedo.

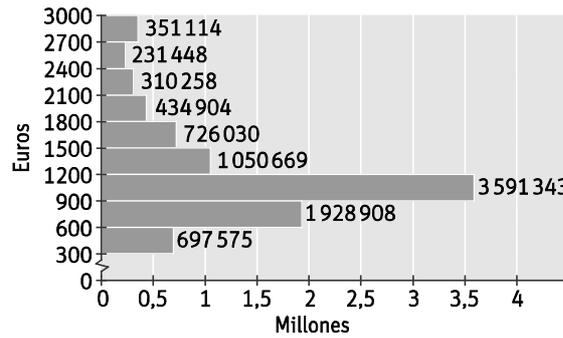
b) $\bar{x} = \frac{182+170+137+117+149+80+40+122+133+156+198+182}{12} = 138,83$ $M = \frac{137+149}{2} = 143$

c) $Q_1 = 122$

$Q_2 = M = 143$

$Q_3 = 182$

34. El número de pensionistas en España y la cuantía de las pensiones en 2015 viene dados en este gráfico.

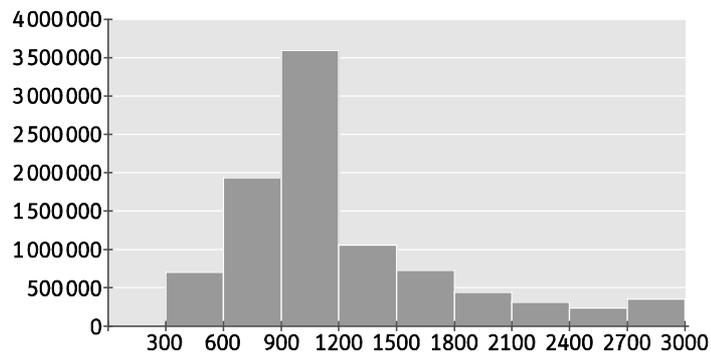


a) Haz la tabla de frecuencias relativas acumuladas (%) y dibuja el histograma.

b) Calcula la moda, la media y la mediana.

a)

Intervalos	Marca: x_i	f_i	$x_i \cdot f_i$	h_i	F_i	H_i	H_i (%)
[300, 600)	450	697 575	313 908 75	0,07	697 575	0,07	7,48
[600,900)	750	1 928 908	144 668 10	0,21	2 626 483	0,28	28,17
[900, 1200)	1050	3 591 343	377 091 01	0,39	6 217 826	0,67	66,70
[1200,1500)	1350	1 050 669	141 840 31	0,11	7 268 495	0,78	77,97
[1500,1800)	1650	726 030	119 794 95	0,08	7 994 525	0,86	85,76
[1800,2100)	1950	434 904	848 062 80	0,05	8 429 429	0,90	90,42
[2100,2400)	2250	310 258	698 080 50	0,03	8 739 687	0,94	93,75
[2400,2700)	2550	231 448	590 192 40	0,02	8 971 135	0,96	96,23
[2700,3000)	2850	351 114	100 067 49	0,04	9 322 249	1,00	100,00
		$N = 9\,322\,249$	$1,128 \cdot 10^{10}$				



b) $M_o = [900, 1200)$

$$\bar{x} = \frac{1,128 \cdot 10^{10}}{9\,322\,249} = 1210$$

$M = [900, 1200)$

35. Los resultados de una prueba de evaluación externa en un centro han sido:

Puntos	0	1	2	3	4	5	6	7	8	9	10
Alumnos	2	12	32	38	45	54	68	60	54	29	6

- Halla la moda, la media y la mediana de las notas.
- Dibuja el polígono de frecuencias acumuladas.
- Calcula los cuartiles.
- Dibuja el diagrama de cajas.

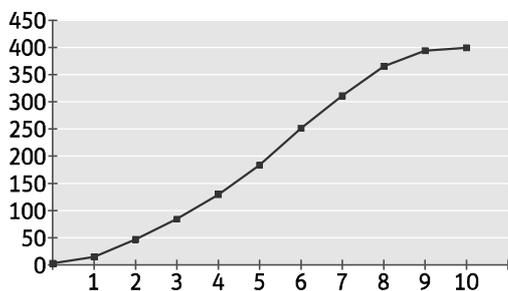
Puntos: x_i	f_i	h_i	F_i	H_i	H_i (%)	$x_i \cdot f_i$
0	2	0,005	2	0,005	0,5	0
1	12	0,03	14	0,035	3,5	12
2	32	0,08	46	0,115	11,5	64
3	38	0,095	84	0,21	21	114
4	45	0,1125	129	0,3225	32,25	180
5	54	0,135	183	0,4575	45,75	270
6	68	0,17	251	0,6275	62,75	408
7	60	0,15	311	0,7775	77,75	420
8	54	0,135	365	0,9125	91,25	432
9	29	0,0725	394	0,985	98,5	261
10	6	0,015	400	1	100	60
	$N = 400$	1				2221

a) $M_o = 6$

$$\bar{x} = \frac{2221}{400} = 5,55$$

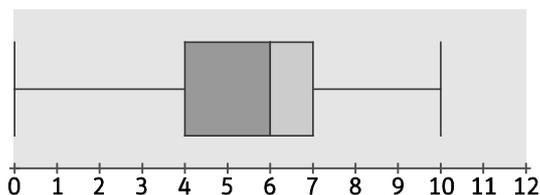
El 50 % de 400 = 200 $\Rightarrow M = 6$

b)



c) El 25 % de 400 = 100 $\Rightarrow Q_1 = 4$. El 50 % de 400 = 200 $\Rightarrow M = Q_2 = 6$. El 75 % de 400 = 300 $\Rightarrow Q_3 = 7$

d) Máximo = 10, mínimo = 0, $Q_1 = 4$, $M = Q_2 = 6$ y $Q_3 = 7$



36. Se ha preguntado a 30 personas sobre el número de horas diarias que ven la televisión, con el siguiente resultado:

Horas TV	0	1	2	3	4	5
N.º personas	4	6	8	7	4	1

- a) Calcula la media, la varianza y la desviación típica.
 b) Halla el coeficiente de variación.

Horas TV:	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
0	4	0	0
1	6	6	6
2	8	16	32
3	7	21	63
4	4	16	64
5	1	5	25
	$N = 30$	64	190

a) $\bar{x} = \frac{64}{30} = 2,13 \Rightarrow s^2 = \frac{190}{30} - 2,13^2 = 1,80 \Rightarrow s = 1,34$

b) $CV = \frac{1,34}{2,13} = 0,6291 \rightarrow 62,91\%$

37. Actividad resuelta.

38. En una empresa de madera una sierra automática corta tablones de 2 m de longitud. En un control se ha medido una muestra de 10 tablones y las longitudes han sido:

2,01 2,03 1,96 2,06 1,98 2,05 2,08 1,95 2,10 1,92

- a) Halla la media y la desviación típica.
 b) Se desechan los tablones cuyas longitudes no estén comprendidas entre $\bar{x} - s$ y $\bar{x} + s$. ¿Cuántos tablones de la muestra hay que desechar?

a) $\bar{x} = \frac{1,92 + 1,95 + 1,96 + 1,98 + 2,01 + 2,03 + 2,05 + 2,06 + 2,08 + 2,10}{10} = 2,014$

$s^2 = \frac{1,92^2 + 1,95^2 + 1,96^2 + 1,98^2 + 2,01^2 + 2,03^2 + 2,05^2 + 2,06^2 + 2,08^2 + 2,10^2}{10} - 2,014^2 = 0,003244 \Rightarrow s = 0,057$

- b) Se desechan los tablones que no estén comprendidos entre 1,957 cm y 2,071.

Es decir, se desechan los tablones que miden:

2,08 1,95 2,10 1,92.

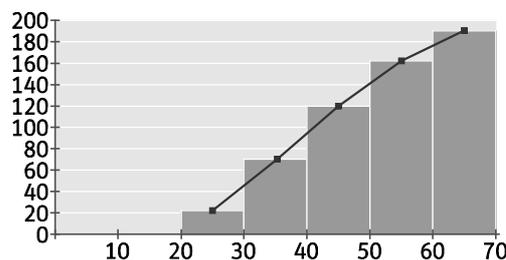
39. Las edades de los empleados de unos grandes almacenes vienen dadas en la tabla:

Edad	[20, 30)	[30, 40)	[40, 50)	[50, 60)	[60, 70)
N.º	22	48	50	42	28

- Haz el histograma de frecuencias y el polígono de frecuencias acumuladas.
- Halla la media y la mediana.
- Calcula el recorrido intercuartílico, la varianza y la desviación típica.

Intervalos	Marca: x_i	f_i	F_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[20, 30)	25	22	22	550	13 750
[30, 40)	35	48	70	1680	58 800
[40, 50)	45	50	120	2250	101 250
[50, 60)	55	42	162	2310	127 050
[60, 70)	65	28	190	1820	118 300
$N = 190$				8610	419 150

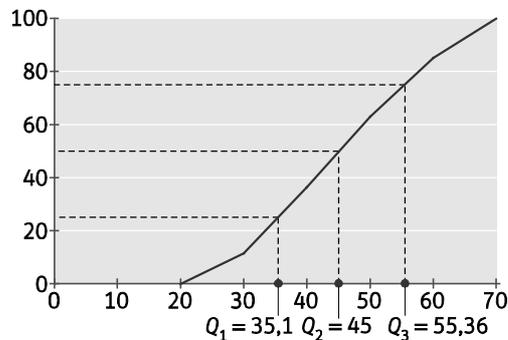
a) Histograma y polígono de frecuencias acumuladas



b) $\bar{x} = \frac{8610}{190} = 45,32$

El 50% de 190 = 95 $\Rightarrow M = 45$

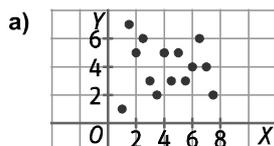
c) Representamos el polígono de frecuencias acumuladas en porcentaje, para hallar los cuartiles.



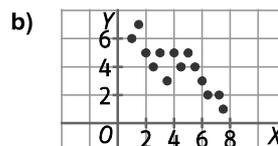
Cuartiles: $Q_1 = 35,31$, $Q_2 = M = 45$ y $Q_3 = 55,36 \Rightarrow$ Recorrido intercuartílico = $55,36 - 35,31 = 20,05$

$$s^2 = \frac{419 150}{190} - 45,32^2 = 152,15 \Rightarrow s = 12,33$$

40. Observa los siguientes diagramas de dispersión e indica el tipo de relación entre las variables.



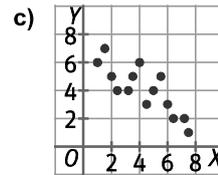
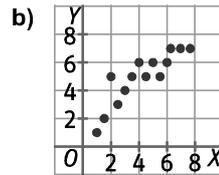
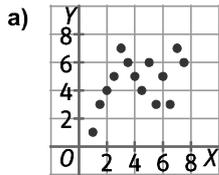
a) Correlación débil.



b) Correlación fuerte inversa.

41. Actividad resuelta.

42. Asocia los coeficientes de correlación r , con su nube de puntos.



A. $r = -0,8$

B. $r = 0,4$

C. $r = 0,9$

- a) La primera nube de puntos no se ajusta claramente a ninguna recta, por lo que su coeficiente de correlación lineal debe ser $r = 0,4$.
- b) La segunda nube de puntos se ajusta a una recta de pendiente positiva, por lo que su coeficiente de correlación debe ser $r = 0,9$.
- c) La tercera nube de puntos se ajusta a una recta de pendiente negativa, por lo que su coeficiente de correlación debe ser $r = -0,8$.

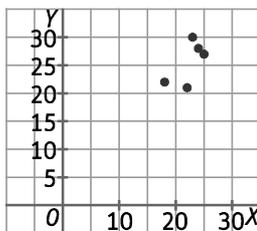
43. Para llegar al trabajo Isa puede ir en su coche o en autobús. Durante una semana va en su coche y la siguiente en transporte público para controlar el tiempo que tarda, en minutos.

	L	M	X	J	V
Coche	25	18	22	24	23
Autobús	27	22	21	28	30

- a) Dibuja la nube de puntos y el centro de gravedad.
- b) Halla la covarianza e interpreta el resultado.
- c) Calcula el coeficiente de correlación.

Día	Coche: x_i	Autobús: y_i	x_i^2	y_i^2	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
L	25	27	625	729	2,6	1,4	3,64
M	18	22	324	484	-4,4	-3,6	15,84
X	22	21	484	441	-0,4	-4,6	1,84
J	24	28	576	784	1,6	2,4	3,84
V	23	30	529	900	0,6	4,4	3,64
	112	128	2538	3338			27,8

a)



$$\bar{x} = \frac{112}{5} = 22,4 \quad \text{e} \quad \bar{y} = \frac{128}{5} = 25,6$$

El centro de gravedad de la distribución es (22,4; 25,6).

b) $s_{x,y} = \frac{27,8}{5} = 5,56 > 0 \Rightarrow$ Hay una correlación directa entre las dos variables.

c) $s_x^2 = \frac{2538}{5} - 22,4^2 = 5,84 \Rightarrow s_x = 2,42; s_y^2 = \frac{3338}{5} - 25,6^2 = 12,24 \Rightarrow s_y = 3,5$ y $r = \frac{5,56}{2,42 \cdot 3,5} = 0,66$

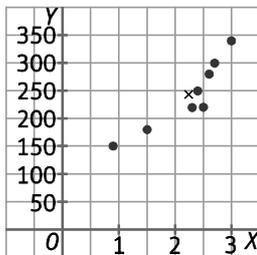
44. El director de un comercio quiere evaluar el impacto de los gastos en publicidad en el volumen de ventas de su negocio. Ha recogido los gastos en publicidad y ventas, en miles de euros, los últimos ocho meses:

G	2,6	2,5	2,4	1,5	0,9	3	2,7	2,3
V	280	220	250	180	150	340	300	220

- Dibuja la nube de puntos y marca el centro de gravedad.
- Halla la desviación típica de ambas variables.
- Calcula la covarianza y el coeficiente de correlación.
- Halla las ventas estimadas si el gasto en publicidad hubiese sido de 1,2 millones de euros.

G: x_i	V: y_i	x_i^2	y_i^2	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
2,6	280	6,76	78 400	0,36	37,5	13,5
2,5	220	6,25	48 400	0,26	-22,5	-5,85
2,4	250	5,76	62 500	0,16	7,5	1,2
1,5	180	2,25	32 400	-0,74	-62,5	46,25
0,9	150	0,81	22 500	-1,34	-92,5	123,95
3	340	9	115 600	0,76	97,5	74,1
2,7	300	7,29	90 000	0,46	57,5	26,45
2,3	220	5,29	48 400	0,06	-22,5	-1,35
17,9	1940	43,41	498 200			278,25

a)



$$\bar{x} = \frac{17,9}{8} = 2,2375 \text{ e } \bar{y} = \frac{1940}{8} = 242,5$$

El centro de gravedad de la distribución es (2,24; 242,5).

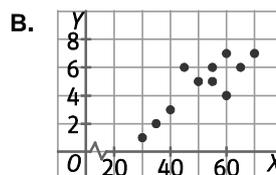
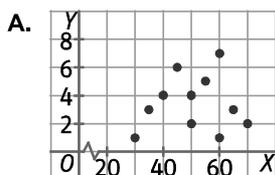
$$b) s_x^2 = \frac{43,41}{8} - 2,2375^2 = 0,4198 \Rightarrow s_x = 0,6479 \text{ y } s_y^2 = \frac{498\,200}{8} - 242,5^2 = 3468,75 \Rightarrow s_y = 58,8961$$

$$s_{x,y} = \frac{278,25}{8} = 34,78 \Rightarrow r = \frac{34,78}{0,6479 \cdot 58,8961} = 0,91$$

$$c) \text{ La recta de regresión de Y sobre X es } y - 242,5 = \frac{34,78}{0,4198} (x - 2,2375) \Rightarrow y = 82,85x + 57,12.$$

- d) Si se gastasen 1,2 miles de euros en publicidad, se estimarían unas ventas de $y = 82,85 \cdot 1,2 + 57,12 = 156,54$ miles de euros.

45. ¿Cuál de estas dos nubes de puntos representa la edad (X) y las notas obtenidas en un examen de acceso a la universidad de mayores de 25 años? ¿Por qué?



La gráfica A. tiene una correlación débil debido a que los puntos de la nube están bastante dispersos y no se ajustan a una recta. Sin embargo, en la gráfica B, hay una correlación directa fuerte porque los puntos de la nube se ajustan a una recta creciente.

Por tanto, la gráfica A representa la edad y las notas obtenidas en un examen de acceso a la universidad de mayores de 25 años, porque no existe correlación fuerte, ni directa ni inversa, entre estas dos variables.

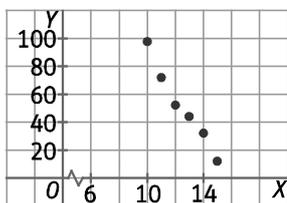
46. Una sociedad industrial vende máquinas a varias empresas. El precio mínimo es de 10 000 €. El número de máquinas vendidas, Y, está relacionada con el precio, X, en miles de euros según la tabla:

Precio	10	11	12	13	14	15
N.º máquinas	98	72	52	44	32	12

- Representa la nube de puntos.
- Calcula la recta de regresión.
- ¿Cuáles serían las ventas estimadas para un precio de 11 500 €?

Precio: x_i	N.º máquinas: y_i	x_i^2	y_i^2	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
10	98	100	9604	-2,5	46,33	-115,83
11	72	121	5184	-1,5	20,33	-30,50
12	52	144	2704	-0,5	0,33	-0,17
13	44	169	1936	0,5	-7,67	-3,83
14	32	196	1024	1,5	-19,67	-29,50
15	12	225	144	2,5	-39,67	-99,17
75	310	955	20 596			-279,00

a)



$$b) \quad \bar{x} = \frac{75}{6} = 12,5 \Rightarrow s_x^2 = \frac{955}{6} - 12,5^2 = 2,92 \quad \text{e} \quad \bar{y} = \frac{310}{6} = 51,67$$

$$s_{x,y} = \frac{-279}{6} = -46,5$$

$$\text{La recta de regresión de Y sobre X es } y - 51,67 = \frac{-46,5}{2,92} (x - 12,5) \Rightarrow y = -15,92x + 250,67$$

c) Para un precio de 11 500 €, se estiman unas ventas de $y = -15,92 \cdot 11,5 + 250,67 = 67,59$ máquinas.

47. La tabla muestra las emisiones de CO₂ por habitante (en toneladas métricas) y el uso de combustibles fósiles como fuente de energía 2013 (*Fuente Banco Mundial*).

	CO ₂ /Hab. (Tm)	C. fósiles (% sobre el total de energía)
Alemania	8,9	80,9
Argelia	3,3	99,9
Brasil	2,2	56,5
Canadá	14,1	73,4
China	6,7	88,2
Colombia	1,6	75,6
España	5,8	75,0
Ghana	0,4	43,6
EE.UU.	17,0	83,7

- ¿Existe alguna correlación entre las dos variables? Halla el coeficiente de correlación para justificar tu respuesta.
- Halla la recta de regresión del uso de combustibles fósiles sobre las emisiones de CO₂.
- ¿Serían fiables el valor aproximado de uso de combustibles fósiles para un país con unas emisiones de CO₂ de 10 Tm por habitante?

CO ₂ : x _i	C. fósiles: y _i	x _i ²	y _i ²	x _i - \bar{x}	y _i - \bar{y}	(x _i - \bar{x})(y _i - \bar{y})
8,9	80,9	79,21	6544,81	2,23	5,7	12,71
3,3	99,9	10,89	9980,01	-3,37	24,7	-83,24
2,2	56,5	4,84	3192,25	-4,47	-18,7	83,59
14,1	73,4	198,81	5387,56	7,43	-1,8	-13,37
6,7	88,2	44,89	7779,24	0,03	13	0,39
1,6	75,6	2,56	5715,36	-5,07	0,4	-2,028
5,8	75	33,64	5625	-0,87	-0,2	0,174
0,4	43,6	0,16	1900,96	-6,27	-31,6	198,13
17	83,7	289	7005,69	10,33	8,5	87,81
60	676,8	664	53 130,88			284,16

$$a) \bar{x} = \frac{60}{9} = 6,67 \Rightarrow s_x^2 = \frac{664}{9} - 6,67^2 = 29,29 \Rightarrow s_x = 5,41$$

$$\bar{y} = \frac{676,8}{9} = 75,2 \Rightarrow s_y^2 = \frac{53 130,88}{9} - 75,2^2 = 248,39 \Rightarrow s_y = 15,76$$

$$s_{x,y} = \frac{284,16}{9} = 31,573 \Rightarrow r = \frac{31,573}{5,41 \cdot 15,76} = 0,37$$

$$b) \text{ La recta de regresión de Y sobre X es } y - 75,2 = \frac{31,573}{29,29} (x - 6,67) \Rightarrow y = 1,078x + 68.$$

c) El coeficiente de correlación lineal es $r = 0,369$ e indica una dependencia débil. El modelo lineal únicamente tendría sentido a la hora de realizar estimaciones en puntos cercanos al centro de gravedad.

Por tanto, no sería fiable el valor obtenido al estimar el valor del uso de combustibles fósiles para un país con unas emisiones de CO₂ de 10 Tm por habitante.

48. Actividad resuelta.

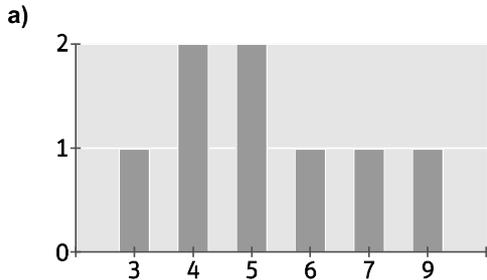
49. Calcula la recta de regresión de Y sobre X en una distribución bidimensional (X, Y) si $\bar{x} = 195$; $\bar{y} = 92,1$; $r = 0,94$; $s_x = 6,07$; $s_y = 6,56$.

$$\text{Como } r = \frac{S_{xy}}{s_x s_y} \Rightarrow S_{xy} = r s_x s_y \Rightarrow S_{xy} = 0,94 \cdot 6,07 \cdot 6,56 = 37,43$$

$$\text{Por tanto, la recta de regresión de Y sobre X es } y - 92,1 = \frac{37,43}{6,07^2} (x - 195) \Rightarrow y = 1,02x - 106,8.$$

50. Las notas de 8 alumnos en un examen de Matemáticas han sido: 4, 5, 6, 4, 7, 9, 3 y 5.

- Representa los datos en un gráfico adecuado.
- Calcula la media y la desviación típica.
- El profesor premia a sus alumnos y les sube 2 puntos a cada uno. ¿Qué les sucede a la media y la desviación típica?



b) $\bar{x} = \frac{3+4 \cdot 2+5 \cdot 2+6+7+9}{8} = 5,375$ y $s_x^2 = \frac{3^2+4^2 \cdot 2+5^2 \cdot 2+6^2+7^2+9^2}{8} - 5,375^2 = 3,23 \Rightarrow s_x = 1,79$

c) $\bar{x} = \frac{5+6 \cdot 2+7 \cdot 2+8+9+11}{8} = 7,375 \Rightarrow$ La media aumenta 2 puntos.

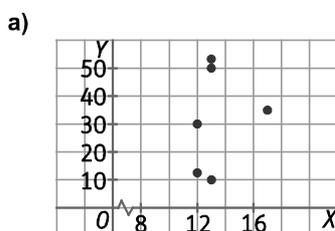
$s_x^2 = \frac{5^2+6^2 \cdot 2+7^2 \cdot 2+8^2+9^2+11^2}{8} - 7,375^2 = 3,23 \Rightarrow s_x = 1,79 \Rightarrow$ La desviación típica no varía.

51. Las previsiones de la Agencia Estatal de Meteorología AEMET para una semana en Zaragoza han sido:

- Dibuja la nube de puntos que relaciona la temperatura máxima con la probabilidad de precipitación.
- Calcula la covarianza y el coeficiente de correlación.
- ¿Existe alguna correlación entre la temperatura máxima y la probabilidad de precipitaciones? ¿De qué tipo es?

mar 15				mié 16		jue 17		vie 18	sáb 19
00-06 h	06-12 h	12-18 h	18-24 h	00-12 h	12-24 h	00-12 h	12-24 h		
6°C	9°C	10°C	6°C						
Probabilidad de precipitación									
75%	80%	5%	0%	5%	15%	5%	20%	50%	35%
Cota de nieve a nivel de provincia (m)									
900	900				1000	900	1300	1700	
Temperatura mínima y máxima (°C)									
6 / 13				5 / 13		2 / 12		2 / 13	6 / 17

Temperatura x_i	Precipitación y_i	x_i^2	y_i^2	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
13	53,3	169	2840,89	-0,33	21,50	-7,17
13	10	169	100	-0,33	-21,80	7,27
12	12,5	144	156,25	-1,33	-19,30	25,73
13	50	169	2500	-0,33	18,20	-6,07
17	35	289	1225	3,67	3,20	11,73
80	190,8	1084	7722,14			33,90



b) $\bar{x} = \frac{80}{6} = 13,333 \Rightarrow s_x^2 = \frac{1084}{6} - 13,333^2 = 2,9 \Rightarrow s_x = 1,7$

$\bar{y} = \frac{190,8}{6} = 31,8 \Rightarrow s_y^2 = \frac{7722,14}{6} - 31,8^2 = 275,78 \Rightarrow s_y = 16,6$

$s_{x,y} = \frac{33,9}{6} = 5,65 \Rightarrow r = \frac{5,65}{1,7 \cdot 16,6} = 0,2$

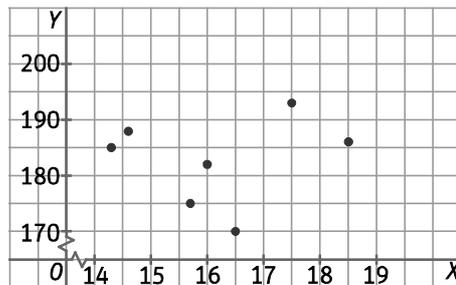
- c) Existe correlación directa entre ambas variables, pero muy débil porque r está próximo a cero.

52. Luis quiere comprarse una moto de 1000 c.c. Compara precios, potencia y peso de varios modelos.

Marca	Tuzuki	Conda	MMM	ZVK	Tamiha	Asawaki	MV Felicia
€	14 300	15 700	16 500	17 500	16 000	14 600	18 500
CV	185	175	170	193	182	188	186
Kg	208	199	202	209	216	208	212

- Representa la nube de puntos en la que X es el precio en miles de euros e Y la potencia. Halla el coeficiente de correlación. Explica el tipo de relación entre las dos variables.
- Representa la nube de puntos en la que X sea la potencia e Y el peso. ¿Qué puedes indicar de la correlación entre ambas?

a) Nube de puntos:



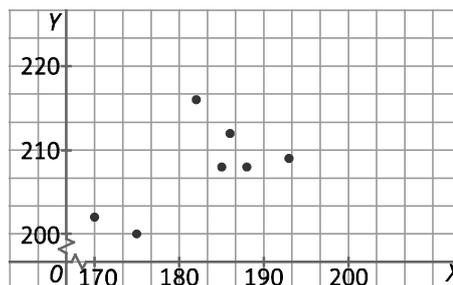
Precio: x_i	Potencia: y_i	x_i^2	y_i^2	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
14,3	185	204,49	34 225	-1,86	2,29	-4,24
15,7	175	246,49	30 625	-0,46	-7,71	3,53
16,5	170	272,25	28 900	0,34	-12,71	-4,36
17,5	193	306,25	37 249	1,34	10,29	13,81
16	182	256	33 124	-0,16	-0,71	0,11
14,6	188	213,16	35 344	-1,56	5,29	-8,23
18,5	186	342,25	34 596	2,34	3,29	7,70
113,1	1279	1840,89	234 063			8,31

$$\bar{x} = \frac{113,1}{7} = 16,157 \Rightarrow s_x^2 = \frac{1840,89}{7} - 16,157^2 = 1,936 \Rightarrow s_x = 1,39$$

$$\bar{y} = \frac{1279}{7} = 182,714 \Rightarrow s_y^2 = \frac{234 063}{7} - 182,714^2 = 53,166 \Rightarrow s_y = 7,29$$

$$s_{x,y} = \frac{8,31}{7} = 1,19 \Rightarrow r = \frac{1,19}{1,39 \cdot 7,29} = 0,117 \Rightarrow \text{Dependencia directa débil.}$$

b) La nube de puntos es la siguiente:

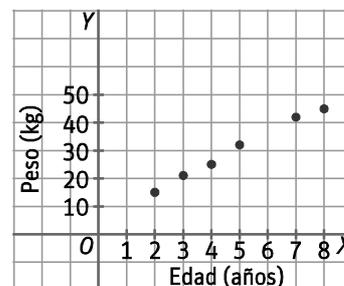


53. Un pediatra ha obtenido esta nube de puntos al estudiar la relación entre el peso y la edad de seis niños de su consulta.

Se conocen los datos de la distribución:

$$\bar{x} = 4,83, \bar{y} = 30, s_x^2 = 4,47, s_y^2 = 117,33, s_{xy} = 22,83$$

- Halla el coeficiente de correlación. ¿Qué indica?
- Calcula la ecuación de la recta de regresión de Y sobre X.
- ¿Cuál sería el peso aproximado de un niño de 6 años?
¿Es fiable esta estimación?



a) $r = \frac{22,83}{\sqrt{4,47} \cdot \sqrt{117,33}} = 0,997 \Rightarrow$ Correlación directa muy fuerte.

b) $y - 30 = \frac{22,83}{4,47} (x - 4,83) \Rightarrow y = 5,11x + 5,32$

c) El peso aproximado de un niño de 6 años sería $y = 5,11 \cdot 6 + 5,32 = 35,98$ kg. Es muy fiable esta estimación porque existe correlación directa muy fuerte entre el peso de un niño y la edad.

54. En una clase de 30 alumnos de 4.º ESO, las chicas calzan por término medio un 37, y los chicos, un 42. La media global de la clase es un 40. ¿Cuántas chicas hay en clase?

- A. 10 B. 12 C. 16 D. 20

Sea x el número de chicas y $30 - x$ el número de chicos.

$$\frac{37x + 42(30 - x)}{30} = 40 \Rightarrow 37x + 1260 - 42x = 1200 \Rightarrow 60 = 5x \Rightarrow x = 12$$

La respuesta correcta es la B.

55. La media de la estatura de 8 personas es 169 cm y la desviación típica 6,68. Se ha descubierto un error en los datos y una estatura de 180 cm se había anotado como 170 cm. ¿Cuál será la desviación típica correcta?

- A. 6,54 B. 6,70 C. 5,92 D. 7,62

Llamamos A a la distribución que contiene los datos $x_1, x_2, \dots, x_7, 170$ y B, a la distribución que contiene los datos $x_1, x_2, \dots, x_7, 180$.

$$\bar{x}_B = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + 180}{8} = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + 170}{8} + 1,25 = \bar{x}_A + 1,25 = 170,25$$

$$s_A^{-2} = \frac{x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2 + x_7^2 + 170^2}{8} - 169^2 = 6,68^2 \Rightarrow x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2 + x_7^2 = 199\,944,9792 \Rightarrow$$

$$s_B^{-2} = \frac{x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2 + x_7^2 + 180^2}{8} - 170,25^2 = \frac{199\,944,98 + 180^2}{8} - 170,25^2 = 58,0599 \Rightarrow \bar{s}_B = 7,62$$

La respuesta correcta es la D.

56. De una distribución bidimensional (X,Y) se sabe que $r = 0,9$, $s_{xy} = 2,268$ y que $s_y = 2,1$. La pendiente de la recta de regresión es:

- A. -2,52 B. 1,2 C. 1,575 D. 2,52

Como $r = \frac{s_{xy}}{s_x s_y} \Rightarrow s_x = \frac{s_{xy}}{r \cdot s_y} \Rightarrow s_x = \frac{2,268}{0,9 \cdot 2,1} = 1,2$ Por tanto, la pendiente de la recta de regresión es $\frac{2,268}{1,2^2} = 1,575$

La respuesta correcta es la C.

57. Las pulsaciones por minuto de un atleta pueden bajar con el entrenamiento.

Meses	0	3	6	9	12	15
Pulsaciones/min	65	64	62	58	57	55

¿Cuántas pulsaciones por minuto tendrá tras 4 años de entrenamiento ininterrumpido?

A. 50

B. 45

C. No se sabe

D. 31

Meses: x_i	Pulsaciones: y_i	x_i^2	y_i^2	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
0	65	0	4225	-7,5	4,83	-36,23
3	64	9	4096	-4,5	3,83	-17,24
6	62	36	3844	-1,5	1,83	-2,75
9	58	81	3364	1,5	-2,17	-3,26
12	57	144	3249	4,5	-3,17	-14,27
15	55	225	3025	7,5	-5,17	-38,78
45	361	495	21803			-112,50

$$\bar{x} = \frac{45}{6} = 7,5 \Rightarrow s_x^2 = \frac{495}{6} - 7,5^2 = 26,25$$

$$\bar{y} = \frac{361}{6} = 60,17$$

$$s_{x,y} = \frac{-112,50}{6} = -18,75$$

La recta de regresión de Y sobre X es $y - 60,17 = \frac{-18,75}{26,25} (x - 7,5) \Rightarrow y = -0,715x + 65,53$.

Tras 4 años = 48 meses de entrenamiento, tendrá $y = -0,715 \cdot 48 + 65,53 = 31,21$ pulsaciones.

La respuesta correcta es la D.

Encuentra el error

58. Una imagen vale más que mil palabras.

A veces los parámetros estadísticos pueden jugar malas pasadas. El estadístico Francis J. Anscombe se empeñó en demostrarlo con imágenes. Observa las dos tablas que inventó.

x	4	5	6	7	8	9	10	11	12	13	14
y	4,26	5,68	7,24	4,82	6,95	8,81	8,04	8,33	10,84	7,58	9,96

x	4	5	6	7	8	9	10	11	12	13	14
y	3,1	4,74	6	7,26	8,14	8,87	9,14	9,26	9,13	8,74	8,1

- La media de la variable y en las dos distribuciones es 7,5.
- En ambas, la variable x toma los mismos valores.
- En las dos el coeficiente de correlación es $r = 0,816$.
- La recta de regresión de las dos es: $y = 0,5x + 3$.

Y sin embargo, si decimos que el valor estimado de y para $x = 16$ es 11 cometo un error grave en una de las dos. ¿Por qué? Dibuja las dos nubes de puntos y explica la situación.

Tabla I

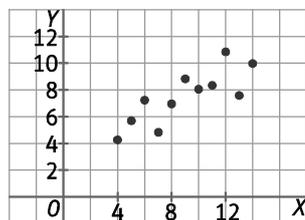
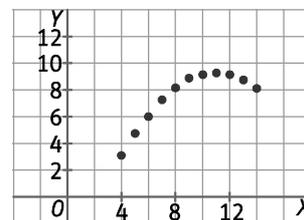


Tabla II



Los datos de las variables de la primera tabla presentan una regresión lineal. Sin embargo, los datos de las variables de la segunda tabla muestran una regresión polinómica.

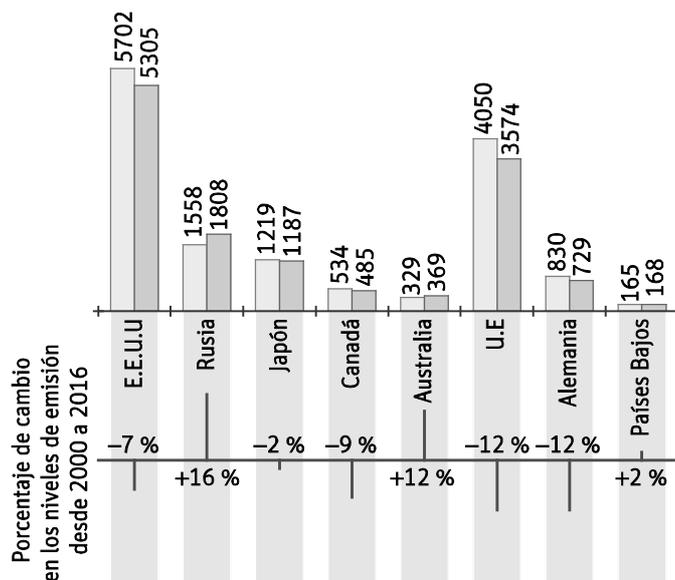
PONTE A PRUEBA

¿Influye el género en los índices de colesterol?

Actividad resuelta.

Los niveles de CO₂.

Desde hace varias décadas el aumento del gas CO₂ en la atmósfera está ayudando a producir un cambio climático. El diagrama muestra los niveles de emisión de CO₂ en 2000 (barras claras) de varios países, los niveles de emisión en 2011 (barras oscuras) y el porcentaje de cambio en los niveles (flechas).



1. El incremento de emisiones en la federación rusa ha sido del 16 %. Explica con los datos del gráfico cómo se ha obtenido ese 16 %.

El nivel de emisiones de CO₂ de Rusia en el año 2000 fue de 1558 y, en el año 2011, de 1808. Como $\frac{1808}{1558} = 1,16$, entonces se produjo un incremento del 16 % desde el año 2000 al año 2011.

2. Luisa se fija en los cambios en los niveles de Países Bajos (+2 %) y de la Unión Europea (-12 %) y le resulta extraño que mientras en el conjunto de la UE desciende, en Países Bajos aumenta. ¿Cuál es el porcentaje global de cambio del resto de los países para que esos datos sean ciertos?

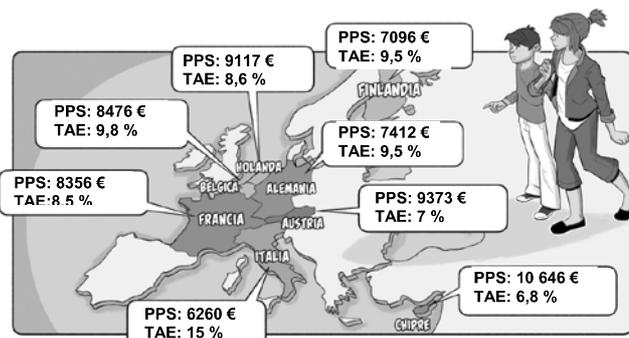
Sea x el índice global de cambio del resto de países de la UE: $1,02 \cdot x = 0,88 \Rightarrow x = 0,86$. El resto de países de la UE ha descendido sus emisiones en un 14 %.

3. Luisa y Antonio discuten sobre qué país o región tuvo el mayor aumento de emisiones. Cada uno llega a una conclusión distinta basándose en el diagrama. Da dos posibles respuestas "correctas" a esta pregunta y explica por qué ambas lo son.

El mayor aumento de emisiones lo tuvieron aquellos países cuyas emisiones en 2001 (barras oscuras) son mayores que sus emisiones en 2000 (barras claras). Rusia y Australia tuvieron el mayor aumento de emisiones.

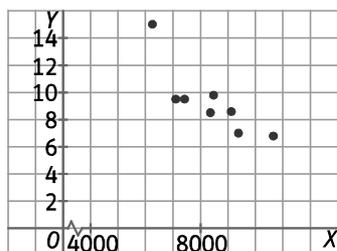
Abandono escolar y gasto por estudiante

España tenía en 2014 una de las tasas de abandono escolar prematuro más altas de la Unión Europea, con el 21,9 %. Es decir, casi 22 de cada 100 estudiantes abandonan los estudios sin obtener el título de secundaria. ¿Hay alguna relación entre la tasa de abandono y el gasto por estudiante de secundaria? En la ilustración puedes ver los datos del gasto por estudiante de secundaria expresado en términos de paridad de poder adquisitivo en euros (PPS) y la tasa de abandono escolar en tanto por ciento (TAE).



Países	PPS x_i	TAE y_i	x_i^2	y_i^2	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
FRANCIA	8356	8,5	69 822 736	72,25	14	-0,84	-11,73
BÉLGICA	8476	9,8	71 842 576	96,04	134	0,46	61,98
HOLANDA	9117	8,6	83 119 689	73,96	775	-0,74	-571,56
FINLANDIA	7096	9,5	50 353 216	90,25	-1246	0,16	-202,48
ALEMANIA	7412	9,5	54 937 744	90,25	-930	0,16	-151,13
AUSTRIA	9373	7	87 853 129	49	1031	-2,34	-2409,96
CHIPRE	10 646	6,8	113 337 316	46,24	2304	-2,54	-5846,40
ITALIA	6260	15	39 187 600	225	-2082	5,66	-11 789,33
	66 736	74,7	570 454 006	742,99			-20 920,60

1. Dibuja la nube de puntos ajustando las unidades de los ejes.



2. Halla el coeficiente de correlación. ¿Existe correlación entre las dos variables? ¿De qué tipo es?

$$\bar{x} = \frac{66\,736}{8} = 8342 \Rightarrow s_x^2 = \frac{570\,454\,006}{8} - 8342^2 = 1717\,786,75 \Rightarrow s_x = 1310,64$$

$$\bar{y} = \frac{74,7}{8} = 9,3375 \Rightarrow s_y^2 = \frac{742,99}{8} - 9,3375^2 = 5,685 \Rightarrow s_y = 2,38$$

$$s_{x,y} = \frac{-20\,920,60}{8} = -2615,075 \Rightarrow r = \frac{-2615,075}{1310,64 \cdot 2,38} = -0,838 \Rightarrow \text{Correlación fuerte inversa.}$$

3. Encuentra la ecuación de la recta de regresión de Y (abandono escolar) sobre X (gasto por estudiante).

$$\text{La recta de regresión de Y sobre X es } y - 9,3375 = \frac{-2615,075}{1717\,786,75} (x - 8342) \Rightarrow y = -0,0015x + 21,85.$$

4. El gasto por estudiante en España es de 7364 €. Según este dato, ¿qué tasa de abandono escolar sería la esperada? Compara el resultado con el valor real. Intenta explicar la diferencia.

Con un gasto de 7364 € se esperaría una tasa de abandono de $y = -0,0015 \cdot 7364 + 21,85 = 10,8\%$. España tiene una tasa de abandono del 21,9 %, muy superior a la 10,8 % esperado, porque se trata de un dato atípico.

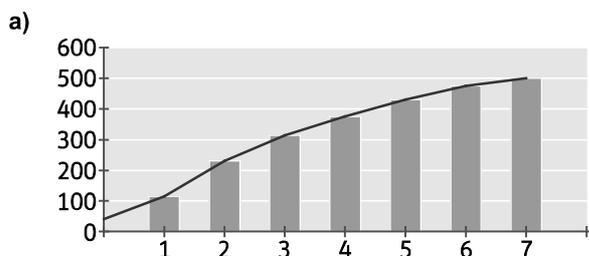
AUTOEVALUACIÓN

1. En una encuesta en una ciudad se ha preguntado el número de periódicos comprados a lo largo de una semana con los siguientes resultados expresados en porcentaje de encuestados:

N.º de periódicos	0	1	2	3	4	5	6	7
Personas (%)	8	15	23	17	12	11	9	5

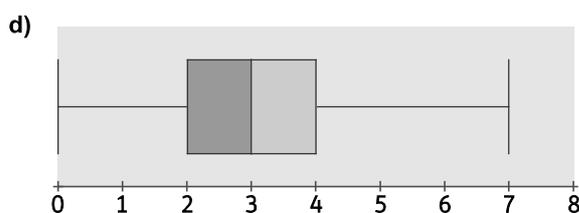
- Construye el diagrama de barras y el polígono de frecuencias.
- Calcula la media de periódicos comprados semanalmente.
- Halla la mediana y los cuartiles.
- Representa el diagrama de cajas.

Periódicos: x_i	%	f_i	$x_i \cdot f_i$	F_i
0	8	40	0	40
1	15	75	75	115
2	23	115	230	230
3	17	85	255	315
4	12	60	240	375
5	11	55	275	430
6	9	45	270	475
7	5	25	175	500
		500	1520	



b) $\bar{x} = \frac{1520}{500} = 3,04$

- c) 25 % de 500 es 125 $\Rightarrow Q_1 = 2$. El 50 % de 500 = 250 $\Rightarrow Q_2 = M = 3$. El 75 % de 500 = 375 $\Rightarrow Q_3 = 4$

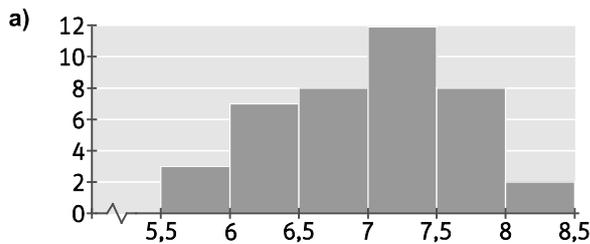


2. Una tienda de embutidos hace recuento del número de jamones y su peso. Los resultados son:

[5,5; 6)	[6; 6,5)	[6,5; 7)	[7; 7,5)	[7,5; 8)	[8; 8,5)
3	7	8	12	8	2

- Representa la distribución con un histograma.
- Calcula la media, la varianza y la desviación típica.
- Halla el coeficiente de variación

Intervalos	Marca: x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[5,5; 6)	5,75	3	17,25	99,19
[6; 6,5)	6,25	7	43,75	273,44
[6,5; 7)	6,75	8	54,00	364,50
[7; 7,5)	7,25	12	87,00	630,75
[7,5; 8)	7,75	8	62,00	480,50
[8; 8,5)	8,25	2	16,50	136,13
		$N = 40$	280,5	1984,5

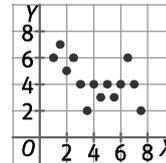
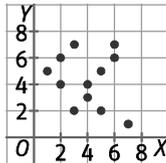
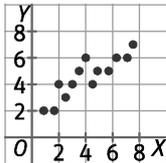


b) $\bar{x} = \frac{280,5}{40} = 7,0125$

$s^2 = \frac{1984,5}{40} - 7,0125^2 = 0,4373 \Rightarrow s = 0,6613$

c) $CV = \frac{0,6613}{7,0125} = 0,0943$

3. Asocia cada nube de puntos con su correspondiente coeficiente de correlación r :



A. $r = -0,2$

B. $r = 0,9$

C. $r = -0,5$

A. Segunda nube.

B. Primera nube.

C. Tercera nube.

4. Las notas obtenidas por 10 alumnos en Lengua (x_i) y en Música (y_i) han sido:

x_i	6	4	8	5	3	7	5	10	5	4
y_i	7	5	7	5	4	8	7	9	6	6

a) Representa la nube de puntos.

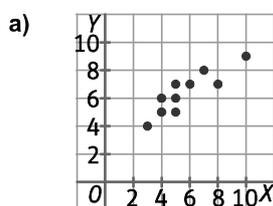
b) Calcula la media y las desviaciones típicas de las dos variables.

c) Halla la covarianza y el coeficiente de correlación.

d) Encuentra la recta de regresión de Y sobre X .

e) ¿Cuál será la nota esperada en Música por un alumno que haya tenido un 8,5 en Lengua?

Lengua: x_i	Música: y_i	x_i^2	y_i^2	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
6	7	36	49	0,3	0,6	0,18
4	5	16	25	-1,7	-1,4	2,38
8	7	64	49	2,3	0,6	1,38
5	5	25	25	-0,7	-1,4	0,98
3	4	9	16	-2,7	-2,4	6,48
7	8	49	64	1,3	1,6	2,08
5	7	25	49	-0,7	0,6	-0,42
10	9	100	81	4,3	2,6	11,18
5	6	25	36	-0,7	-0,4	0,28
4	6	16	36	-1,7	-0,4	0,68
57	64	365	430			25,2



b) $\bar{x} = \frac{57}{10} = 5,7 \Rightarrow s_x^2 = \frac{365}{10} - 5,7^2 = 4 \Rightarrow s_x = 2$ $\bar{y} = \frac{64}{10} = 6,4 \Rightarrow s_y^2 = \frac{430}{10} - 6,4^2 = 2,04 \Rightarrow s_y = 1,43$

c) $s_{x,y} = \frac{25,2}{10} = 2,52 \Rightarrow r = \frac{2,52}{2 \cdot 1,43} = 0,88$

d) La recta de regresión de Y sobre X es $y - 6,4 = \frac{2,52}{4} (x - 5,7) \Rightarrow y = 0,63x + 2,81$.

e) Un alumno que ha obtenido un 8,5 en Lengua se espera que obtenga $y = 0,63 \cdot 8,5 + 2,81 = 8,165$ en Música.