d) $d(C,D)=\sqrt{(d_1-c_1)^2+(d_2-c_2)^2}=$

$=\sqrt{(4-2{,}12)^2+(4-2{,}39)^2}=\sqrt{1{,}88^2+1{,}61^2}=$

$=\sqrt{3{,}53+2{,}59}=\sqrt{6{,}12}=2{,}47$ cm

Aplicando una regla de tres y llamando x a la distancia real en centímetros:

$x=2{,}47\cdot 5\,000=12350$ cm $=123{,}5$ m.

e) Por el teorema de Pitágoras:

$\left|\overline{DE}\right|^2=3{,}5^2+3^2\Rightarrow\left|\overline{DE}\right|^2=12{,}25+9\Rightarrow$

$\Rightarrow\left|\overline{DE}\right|^2=21{,}25\Rightarrow\left|\overline{DE}\right|=\sqrt{21{,}25}=4{,}6$ cm

Con una regla de tres y llamando x a la distancia real en centímetros: $x=4{,}6\cdot 5000=23000$ cm $=230$ m.

5. a) $v_{0x}=20\cos 60°=10$ y $v_{0y}=20\operatorname{sen} 60°=17{,}32$

b) $y=-\frac{1}{2}gt^2+v_{0x}t+y_0=-\frac{1}{2}\cdot 9{,}8t^2+10t+1{,}8=$

$=-4{,}9t^2+10t+1{,}8$

c) El tiempo de vuelo de la pelota es el tiempo hasta que llega al suelo, esto es, hasta que $y = 0$. Así:

$-4{,}9t^2+10t+1{,}8=0\Rightarrow 4{,}9t^2-10t-1{,}8=0\Rightarrow$

$\Rightarrow t=\frac{10\pm\sqrt{10^2+35{,}28}}{9{,}8}\Rightarrow$

$\Rightarrow t=\frac{10\pm\sqrt{100+35{,}28}}{9{,}8}\Rightarrow$

$\Rightarrow t=\frac{10\pm\sqrt{135{,}28}}{9{,}8}\Rightarrow t=\frac{10\pm 11{,}63}{9{,}8}\Rightarrow$

$\Rightarrow t=\frac{21{,}63}{9{,}8}=2{,}2$ s

d) El balón alcanza la altura máxima en

$v_{0y}t-\frac{1}{2}gt^2+y_0=17{,}32\cdot 1{,}1-\frac{1}{2}\cdot 9{,}8\cdot 1{,}1^2+1{,}8=$

$=14{,}92$ m

e) La pelota recorrió $v_{0x}t=10\cdot 2{,}2=22$ m.

11. Estadística

ACTIVIDADES

1. Población: todos los alumnos del centro en cuestión.

Variable estadística: videojuegos preferidos.

En este caso la representatividad de la muestra es de 10 personas por aula, escogidas al azar.

2. Hallamos el número de alumnos de ESO de la ciudad: 1 300 + 1 250 + 1 100 + 1 350 = 5 000

Representamos por n_1, n_2, n_3 y n_4 el número de alumnos de 1.º ESO, 2.º ESO, 3.º ESO y 4.º ESO respectivamente, que deben formar la muestra.

$\frac{n_1}{1300}=\frac{300}{5\,000}\Rightarrow n_1=\frac{1300\cdot 300}{5\,000}=78$

$\frac{n_2}{1250}=\frac{300}{5\,000}\Rightarrow n_2=\frac{1250\cdot 300}{5\,000}=75$

$\frac{n_3}{1100}=\frac{300}{5\,000}\Rightarrow n_3=\frac{1100\cdot 300}{5\,000}=66$

$\frac{n_4}{1350}=\frac{300}{5\,000}\Rightarrow n_4=\frac{1350\cdot 300}{5\,000}=81$

3.

Número de hermanos	Frecuencia absoluta (n_i)	Frecuencia relativa (f_i)	Frecuencia absoluta acumulada (N_i)	Frecuencia relativa acumulada (F_i)
1	4	0,16	4	0,16
2	12	0,48	16	0,64
3	6	0,24	22	0,88
4	2	0,08	24	0,96
5	1	0,04	25	1
	$\sum n_i=25$	$\sum f_i=1$		

4.

Número de hermanos	n_i	f_i	N_i	F_i
0	2	0,08	2	0,08
1	8	0,32	10	0,40
2	10	0,4	20	0,80
3	4	0,16	24	0,9
4	1	0,04	25	1
	$\sum n_i=25$	$\sum f_i=1$		

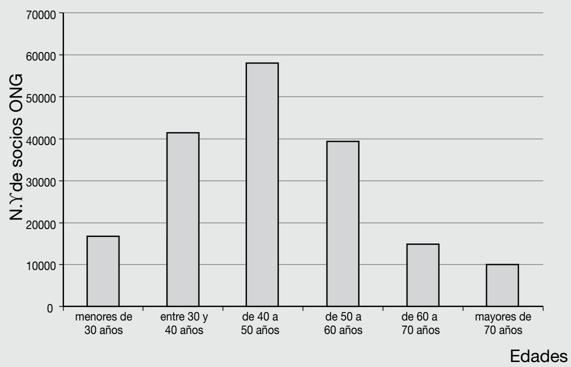
5.

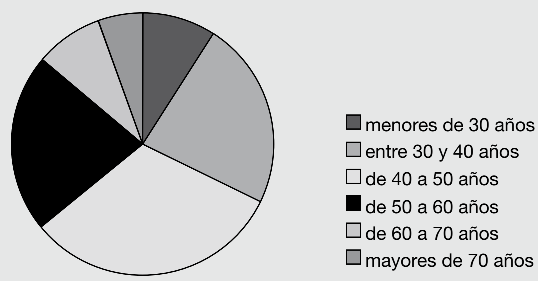
Temperatura mínima	Marca de clase	n_i	f_i	N_i	F_i
[0,2)	1	5	0,1876	5	0,1876
[2,4)	3	12	0,4286	17	0,6071
[4,6)	5	6	0,2143	23	0,8214
[6,8)	7	3	0,1071	26	0,9286
[8,10)	9	2	0,0714	28	1

6.

t (min)	Marca de clase	n_i	f_i	N_i	F_i
[0,5)	2,5	10	0,4	10	0,4
[5,10)	7,5	9	0,36	19	0,76
[10,15)	12,5	3	0,12	22	0,88
[15,20)	17,5	1	0,04	23	0,92
[20,25)	22,5	2	0,08	25	1
		$\sum n_i=25$	$\sum f_i=1$		

7.





8. La media del peso de los recién nacidos es 3,4 kg. El dato es representativo y podemos tomarlo como referencia ya que el coeficiente de variación es del 9,4 %.

9. El estudio se refiere a dos características de una población. Por tanto, se considera una variable estadística bidimensional (X, Y), donde X = peso de cada alumno e Y = altura de cada alumno.

10. Variable estadística: (X, Y), siendo X el número de horas de televisión, e Y el número de horas de estudio semanales.

La representatividad se ha obtenido por muestreo aleatorio.

11.

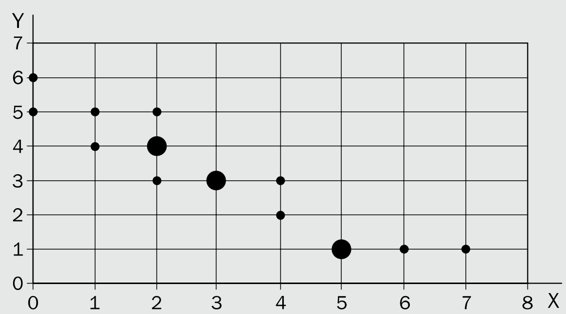
Y \ X	7	8	9	10	11	Total
2	1	1	1	1	1	5
3	0	1	0	0	0	1
4	1	3	1	0	1	6
Total	2	5	2	1	2	12

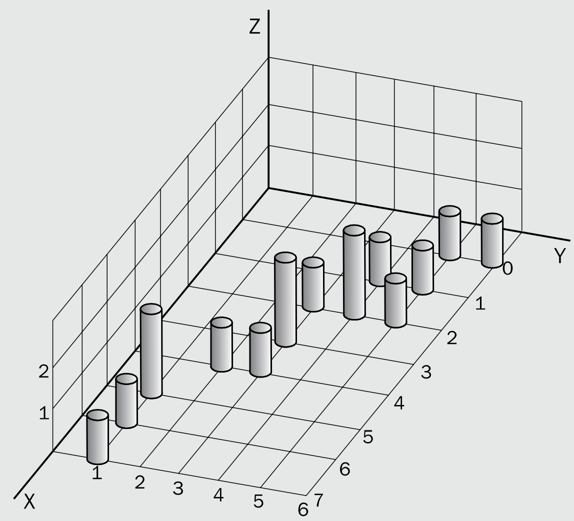
12.

Y \ X	7	8	9	10
[4,5; 5,0)	0	1	0	1
[5,0; 5,5)	6	2	0	8
[5,5; 6,0)	0	1	1	2
[6,0; 6,5)	0	0	1	1
Total	6	4	2	12

13.

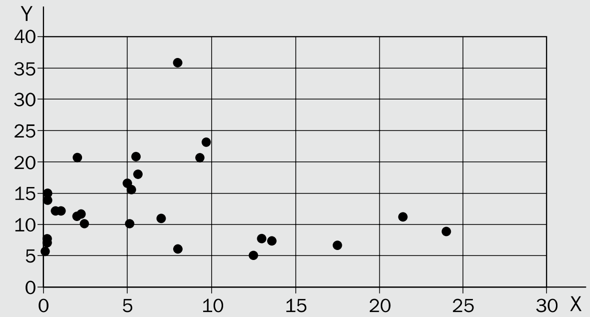
Y \ X	0	1	2	3	4	5	6	7	Total
1	0	0	0	0	0	2	1	1	4
2	0	0	0	0	1	0	0	0	1
3	0	0	1	2	1	0	0	0	4
4	0	1	2	0	0	0	0	0	3
5	1	1	1	0	0	0	0	0	3
6	1	0	0	0	0	0	0	0	1
Total	2	2	4	2	2	2	1	1	16

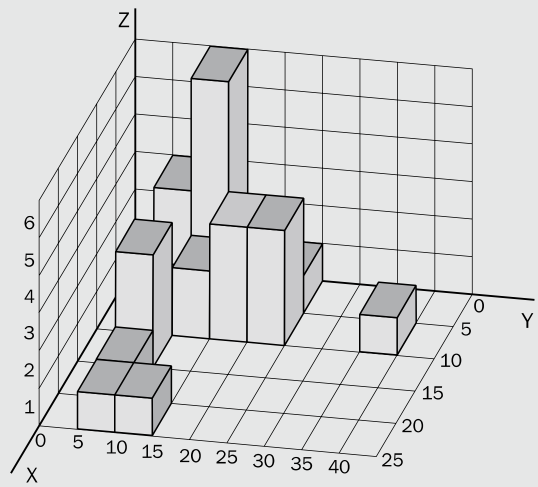




14.

Y \ X	[0, 5)	[5,10)	[10, 15)	[15, 20)	[20, 25)	[25, 30)
[5,10)	3	1	3	1	1	9
[10, 15)	6	2	0	0	1	9
[15, 20)	1	3	0	0	0	4
[20, 25)	1	3	0	0	0	4
[25, 30)	0	0	0	0	0	0
[35, 40)	0	0	0	0	0	0
[40, 45)	0	1	0	0	0	1
Total	11	10	3	1	2	27





15. a) Dependencia estadística, porque ambos valores están relacionados.

b) Independencia, porque el cociente intelectual y el peso nada tienen que ver.

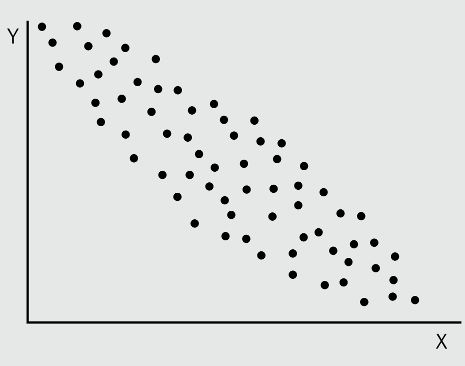
c) Dependencia funcional, porque el área del círculo es función del radio.

d) Independencia, porque el color del cabello y la profesión no guardan relación alguna.

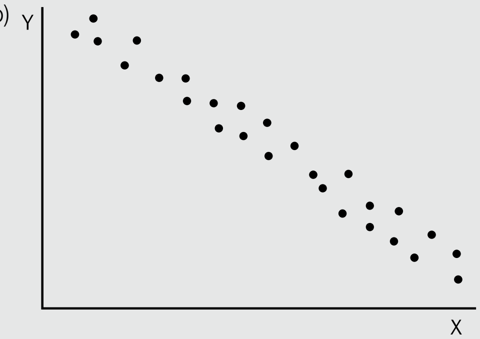
e) Dependencia estadística, porque ambos valores están relcionados.

f) Independencia, porque las horas de sueño no guardan relación con la época del año.

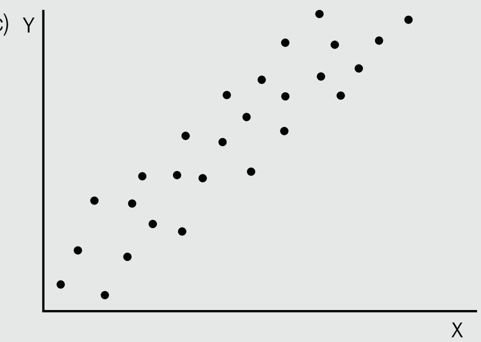
16. a)



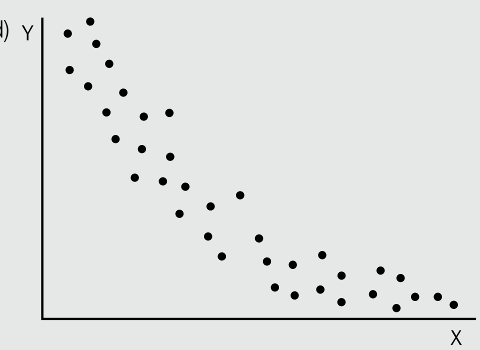
b)



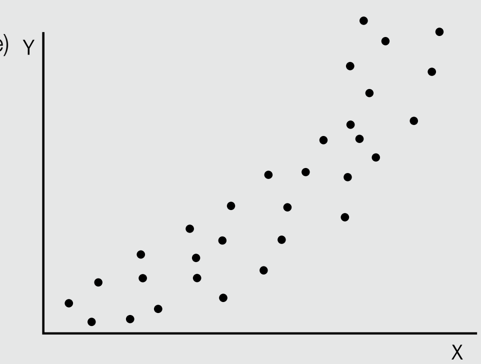
c)



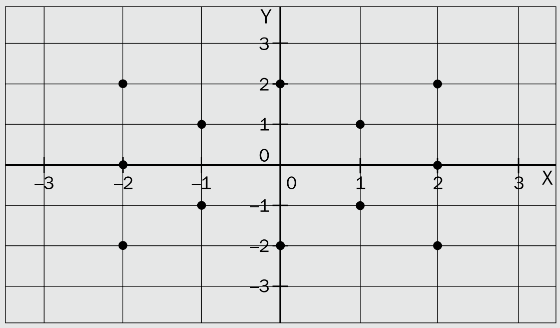
d)



e)



17. a)



No se observa ningún tipo de correlación.

b) $\bar{x} = \dfrac{\sum_i x_i n_i}{N} =$

$= \dfrac{-2 \cdot 3 + (-1) \cdot 2 + 0 \cdot 2 + 1 \cdot 2 + 2 \cdot 3}{12} = 0$

$\bar{y} = \dfrac{\sum_j y_j n_j}{N} =$

$= \dfrac{-2 \cdot 3 + (-1) \cdot 2 + 0 \cdot 2 + 1 \cdot 2 + 2 \cdot 3}{12} = 0$

$\sigma_x = \sqrt{\dfrac{\sum_i x_i^2 n_i}{N} - \bar{x}^2} =$

$= \sqrt{\dfrac{(-2)^2 \cdot 3 + (-1)^2 \cdot 2 + 0^2 \cdot 2 + 1^2 \cdot 2 + 2^2 \cdot 3}{12} - 0} =$

$= 1{,}528$

$\sigma_y = \sqrt{\dfrac{\sum_j x_j^2 n_j}{N} - \bar{y}^2} =$

$= \sqrt{\dfrac{(-2)^2 \cdot 3 + (-1)^2 \cdot 2 + 0^2 \cdot 2 + 1^2 \cdot 2 + 2^2 \cdot 3}{12} - 0} =$

$= 1{,}528$

$\sigma_{xy} = \dfrac{\sum_{i,j} x_i \cdot y_j \cdot n_{ij}}{N} - \bar{x} \cdot \bar{y} = \dfrac{0}{12} - 0 \cdot 0 = 0$

$r = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y} = \dfrac{0}{1{,}528 \cdot 1{,}528} = 0$

c) Al no existir ningún tipo de correlación, el coeficiente es 0.

18. a) El gráfico 2, porque no hay manera de ver una correlación (lineal o curvilínea) de los puntos representados.

b) El gráfico 1, porque se observa una alineación de los puntos en la dirección de una recta de pendiente negativa.

19.

Fase	Tareas
Concepción	• Definición del objetivo del estudio: qué se desea conocer y qué variables estadísticas se utilizarán. • Determinación de la población y valoración de la necesidad de tomar una muestra. • Previsión del costo económico.
Obtención de datos	• Obtención y selección de los datos para el análisis posterior. • Si los datos proceden de una encuesta: redacción de las preguntas, selección y formación del personal encuestador, diseño del plan de actuación y ejecución de la recogida de datos.
Análisis de los datos	• Tratamiento de los datos obtenidos. • Obtención de los parámetros estadísticos relevantes. • Confección de tablas y gráficos.
Presentación de los resultados	• Elaboración de un informe. • Análisis crítico de los resultados obtenidos

20. Los parámetros son valores que nos proporcionan información sobre los datos de la variable estadística que estamos estudiando.

Los parámetros de centralización son los valores que se consideran representativos de la serie de datos.

Los parámetros de dispersión son los valores que informan sobre la dispersión de los datos.

21. Se debe utilizar la muestra, ya que la población es muy grande para manejar correctamente los datos.

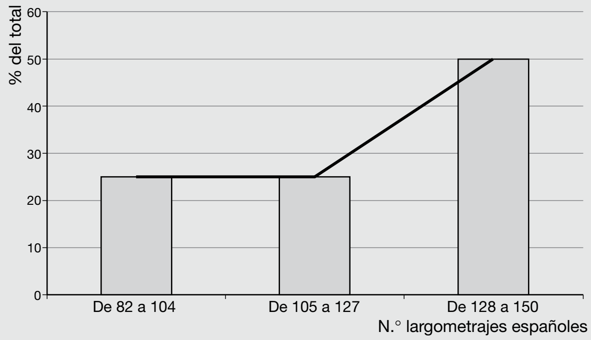
22. a) El grupo de población que lee más es el de mujeres de entre 14 y 24 años.

b) Las diferencias por sexos son mayores en el segmento de 14 a 24 años.

c) Hasta los 54 años, el porcentaje de mujeres que leen es superior al de los hombres. A partir de los 55 años se invierte este comportamiento. En ambos sexos, se observa que el porcentaje de lectores va disminuyendo a medida que aumenta la edad.

23.

N.º de filmes españoles	Marca de clase	n_i	f_i	N_i	F_i
[82, 105)	93,5	2	0,2500	2	0,25
[105, 128)	116,5	2	0,2500	4	0,50
[128, 151)	139,5	4	0,5000	8	1
		$\sum n_i = 8$	$\sum f_i = 1$		



24. Moda: Mo = 20

Media aritmética: $\bar{x} = \frac{\sum_i x_i \cdot n_i}{N} = \frac{580}{34} = 17,059$

Mediana: $\text{Me} = \frac{15+15}{2} = 15$

Recorrido: $r = 25 - 5 = 20$

Desviación media: $d_m = \frac{\Sigma|x_i - \bar{x}| \cdot n_i}{N} = \frac{150}{34} = 4,412$

Varianza: $\sigma^2 = \frac{\Sigma|x_i - \bar{x}|^2 \cdot n_i}{N} = \frac{905,866}{34} = 26,643$

Desviación típica: $\sigma = \sqrt{\sigma^2} = 5,162$

25. Moda: Mo = 7,5

Media aritmética: $\bar{x} = \frac{\sum_i x_i \cdot n_i}{N} = \frac{462}{50} = 9,240$

Mediana: $\text{Me} = L_i + h\frac{\frac{N}{2} - N_{i-1}}{n_i} = 6 + 3\frac{\frac{50}{2} - 5}{25} = 8,400$

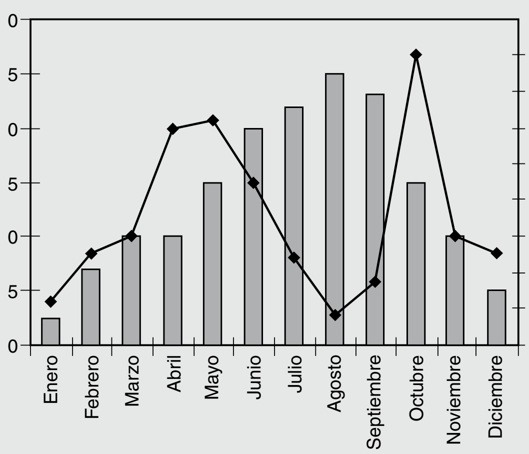
Recorrido: $r = 15 - 0 = 15$

Desviación media: $d_m = \frac{\Sigma|x_i - \bar{x}| \cdot n_i}{N} = \frac{140,4}{50} = 2,808$

Varianza: $\sigma^2 = \frac{\Sigma|x_i - \bar{x}|^2 \cdot n_i}{N} = \frac{505,64}{34} = 10,113$

Desviación típica: $\sigma = \sqrt{\sigma^2} = 3,180$

26. a)



b) TEMPERATURAS:

Calcularemos primero la media aritmética de las temperaturas a lo largo del año:

$$T = \frac{2,5+7+10+10+15+20+22+25+23+15+10+5}{12} = 13,5\text{°C}$$

es la temperatura media a lo largo del año.

Calculamos el cuadrado de la distancia de cada dato a la temperatura media:

Enero	$(2,5 - 13,5)^2 = 121$
Febrero	$(7 - 13,5)^2 = 42,25$
Marzo	$(10 - 13,5)^2 = 12,25$
Abril	$(10 - 13,5)^2 = 12,25$
Mayo	$(15 - 13,5)^2 = 2,25$
Junio	$(20 - 13,5)^2 = 42,25$
Julio	$(22 - 13,5)^2 = 72,25$
Agosto	$(25 - 13,5)^2 = 132,25$
Septiembre	$(23 - 13,5)^2 = 90,25$
Octubre	$(15 - 13,5)^2 = 2,25$
Noviembre	$(10 - 13,5)^2 = 12,25$
Diciembre	$(5 - 13,5)^2 = 72,25$

La varianza se obtiene sumando todos estos datos y dividiendo entre el número total de datos:

$$V = \frac{121 + 42,25 + 12,25 + 12,25 + 2,25 + 42,25 + 72,25 +}{12} =$$
$$= \frac{+132,25 + 90,25 + 2,25 + 12,25 + 72,25}{12} = 51,14$$

La desviación típica se obtiene calculando la raíz cuadrada de la varianza: $\sqrt{51,14} = 7,15$

c) A partir de los datos de la tabla:

Temperatura: CV = 52,2 %

Precipitaciones: CV = 61,1 %

En ambos casos el coeficiente de variación es muy grande (>15 %) por lo que las medias (13,71 ºC y 34,83 mm) no son representativas de la distribución de datos. En cuanto a comparar las distribuciones, la de precipitaciones es más variable que la de temperaturas por ser su coeficiente más alto.

27. $\frac{1,25 + 1,82 + 1,95 + 2,86 + x + 3,58}{6} = 2,43 \Rightarrow$

$\Rightarrow x = 3,12$ millones de €

28. Los 12 datos ordenados son: 4 5 5 7 8 8 8 10 11 12 13 13.

$$i = \left(\frac{1}{4}\right) \cdot 12 = 3$$

Así, $Q_1 = \frac{5+7}{2} = 6$.

29. Las frecuencias absolutas acumuladas son:

Intervalo de clase	[0, 10)	[10, 20)	[20, 30)	[30, 40)	[40, 50)
n_i	3	7	8	8	2
N_i	3	10	18	26	28

Tercer cuartil: $\left(\frac{3}{4}\right)\cdot 28 = 21$

Así, $Q_3 = 30 + \frac{21-18}{8}\cdot(40-30) = 30 + 3{,}75 = 33{,}75$.

30. Las frecuencias absolutas acumuladas son:

Intervalo de clase	[38, 44)	[44 ,50)	[50, 56)	[56, 62)	[62, 68)	[68, 74)
n_i	11	5	9	8	12	7
N_i	11	16	25	33	45	52

Percentil 30: $\left(\frac{30}{100}\right)\cdot 52 = 15{,}6$

Así, $P_{30} = 44 + \frac{15{,}6-11}{5}\cdot(50-44) = 44 + \frac{4{,}6}{5}\cdot 6 = 49{,}52$.

Percentil 65: $\left(\frac{65}{100}\right)\cdot 52 = 33{,}8$

Así, $P_{65} = 62 + \frac{33{,}8-33}{12}\cdot(68-62) = 62 + \frac{0{,}8}{12}\cdot 6 = 62{,}4$.

31. Ordenamos los datos:

11 11 12 13 14 14 14 14 15 15 15 16 16 17 17

Los cuartiles son $Q_1 = 13$, $Q_2 = 14$ y $Q_3 = 16$.

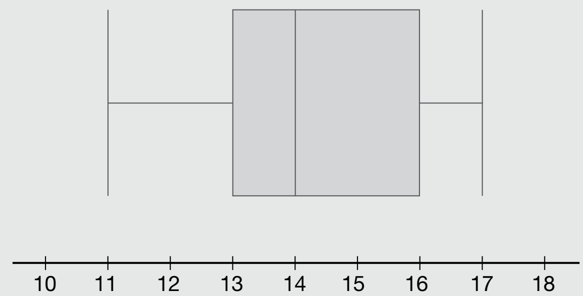
El rango intercuartílico es $IQR = Q_3 - Q_1 = 16 - 13 = 3$.

Los límites inferior y superior son:

$LI = Q_1 - 1{,}5 \cdot IQR = 13 - 1{,}5 \cdot 3 = 8{,}5$

$LS = Q_3 + 1{,}5 \cdot IQR = 16 + 1{,}5 \cdot 3 = 20{,}5$

Por tanto, el diagrama de caja y bigotes es:



32. Ordenamos los datos:

18 18 19 22 23 23 23 23 27 27 29 30 30 32 33 42

Los cuartiles son $Q_1 = \frac{22+23}{2} = \frac{45}{2} = 22{,}5$,

$Q_2 = \frac{23+27}{2} = \frac{50}{2} = 25$ y $Q_3 = \frac{30+30}{2} = 30$.

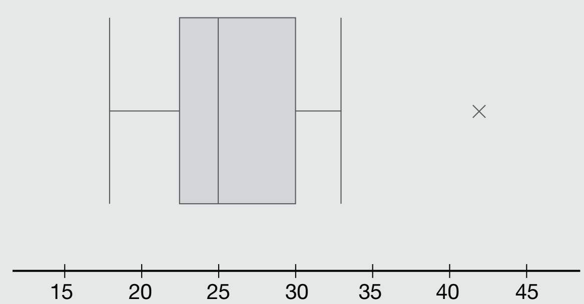
El rango intercuartílico es $IQR = Q_3 - Q_1 = 30 - 22{,}5 = 7{,}5$.

Los límites inferior y superior son:

$LI = Q_1 - 1{,}5 \cdot IQR = 22{,}5 - 1{,}5 \cdot 7{,}5 = 22{,}5 - 11{,}25 = 11{,}25$

$LS = Q_3 + 1{,}5 \cdot IQR = 30 + 1{,}5 \cdot 7{,}5 = 30 + 11{,}25 = 41{,}25$

Por tanto, el diagrama de caja y bigotes es:



33. Confeccionamos la tabla de frecuencias absolutas:

x_i	10	11	13	15	16	19	22	23
n_i	2	4	1	3	2	3	2	3

La media es:

$$\bar{x} = \frac{2\cdot 10 + 4\cdot 11 + 13 + 3\cdot 15 + 2\cdot 16 + 3\cdot 19 + 2\cdot 22 + 3\cdot 23}{20} \Rightarrow \bar{x} = \frac{20+44+13+45+32+57+44+69}{20} \Rightarrow \bar{x} = \frac{324}{20} = 16{,}2$$

La varianza es:

$$\sigma^2 = \frac{|10-16{,}2|^2\cdot 2 + |11-16{,}2|^2\cdot 4 + |13-16{,}2|^2 + |15-16{,}2|^2\cdot 3 + |16-16{,}2|^2\cdot 2 + |19-16{,}2|^2\cdot 3}{20} +$$

$$+ \frac{|22-16{,}2|^2\cdot 2 + |23-16{,}2|^2\cdot 3}{20} = \frac{6{,}2^2\cdot 2 + 5{,}2^2\cdot 4 + 3{,}2^2 + 1{,}2^2\cdot 3 + 0{,}2^2\cdot 2 + 2{,}8^2\cdot 3 + 5{,}8^2\cdot 2 + 6{,}8^2\cdot 3}{20} =$$

$$= \frac{76{,}88 + 108{,}16 + 10{,}24 + 4{,}32 + 0{,}08 + 23{,}52 + 67{,}28 + 138{,}72}{20} = \frac{429{,}2}{20} = 21{,}46$$

La desviación típica es: $\sigma^2 = 21{,}46 \Rightarrow \sigma = \sqrt{21{,}46} = 4{,}63$

Por tanto, el coeficiente de variación es $CV = \frac{\sigma}{\bar{x}} = \frac{4{,}63}{16{,}2} = 0{,}29$.

34. Las frecuencias absolutas acumuladas son:

Intervalo de clase	[3, 6)	[6 ,9)	[9, 12)	[12, 15)	[15, 18)	[18, 21)
n_i	5	2	6	7	3	1
N_i	5	7	13	20	23	24

Tenemos que $\left(\frac{1}{4}\right)\cdot 24 = 6$.

$Q_1 = 6 + \frac{6-5}{2}\cdot(9-6) = 6 + \frac{3}{2} = 7{,}5$

Tenemos que $\left(\frac{3}{4}\right)\cdot 24 = 18$.

$Q_3 = 12 + \frac{18-13}{7}\cdot(15-12) = 12 + \frac{15}{7} = 14{,}14$

Entonces, $IQR = Q_3 - Q_1 = 14{,}14 - 7{,}5 = 6{,}64$.

35. La gráfica tiene dos errores:

a) El valor de desempleados en 2007 no está alineado en la vertical con dicho año.

b) El valor de 2014 es mayor que el valor de 2009 y, sin embargo, en la gráfica el primero está más abajo que el segundo.

36.

x_i	n_i	$x_i \cdot n_i$	$x_i^2 \cdot n_i$
5	1	5	25
7	1	7	49
8	3	24	192
9	1	9	81
10	1	10	100

y_j	n_j	$y_j \cdot n_j$	$y_j^2 \cdot n_j$
11	1	11	121
12	3	36	432
13	2	26	338
16	1	16	256

$N = \sum_i n_i = 7 \qquad \sum_i x_i \cdot n_i = 55 \qquad \sum_i x_i^2 \cdot n_i = 447$

$N = \sum_j n_j = 7 \qquad \sum_j y_j \cdot n_j = 89 \qquad \sum_j y_j^2 \cdot n_j = 1147$

$(x_i \cdot y_j)$	n_{ij}	$x_i \cdot y_j \cdot n_{ij}$
(10, 16)	1	160
(8, 12)	3	288
(7, 11)	1	77
(5, 13)	1	65
(9, 13)	1	117

$N = \sum_{i,j} n_{ij} = 7 \qquad \sum_{i,j} x_i \cdot y_j \cdot n_{ij} = 707$

Hallamos las medias:

$$\overline{x} = \frac{\sum_i x_i \cdot n_i}{N} = \frac{55}{7} = 7{,}857$$

$$\overline{y} = \frac{\sum_j y_j \cdot n_j}{N} = \frac{89}{7} = 12{,}714$$

Las desviaciones típicas son:

$$\sigma_x = \sqrt{\frac{\sum_i x_i^2 \cdot n_i}{N} - \overline{x}^2} = \sqrt{\frac{447}{7} - 7{,}857^2} = \sqrt{2{,}125} = 1{,}458$$

$$\sigma_y = \sqrt{\frac{\sum_j y_j^2 \cdot n_j}{N} - \overline{y}^2} = \sqrt{\frac{1147}{7} - 12{,}714^2} = \sqrt{2{,}211} = 1{,}487$$

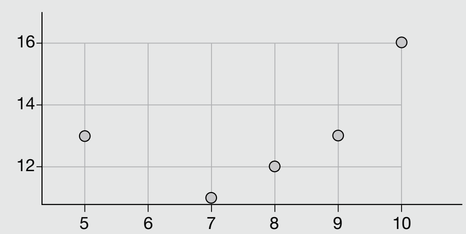
Calculamos la covarianza:

$$\sigma_{xy} = \frac{\sum_{i,j} x_i \cdot y_j \cdot n_{ij}}{N} - \overline{x}\cdot\overline{y} = \frac{707}{7} - 7{,}857 \cdot 12{,}714 = 1{,}106$$

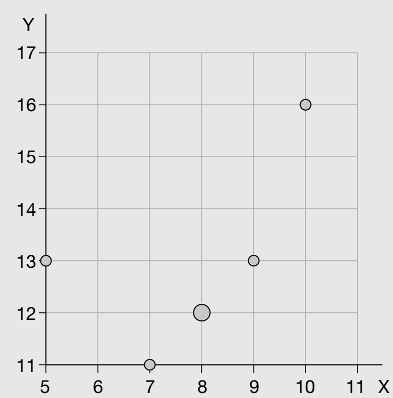
El coeficiente de Pearson es:

$$r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \frac{1{,}106}{1{,}458 \cdot 1{,}487} = 0{,}51$$

El diagrama de dispersión de la distribución es:



(Versión confeccionada con la hoja de cálculo de GeoGebra)



(Versión confeccionada con la hoja de cálculo de GeoGebra)

El valor del coeficiente de Pearson indica una correlación lineal positiva moderada. Así, por ejemplo, cuando hay un aumento en la temperatura mínima, en general, la temperatura máxima también se incrementa.

37. La dependencia funcional entre dos variables estadísticas se da cuando están relacionadas de forma que es posible determinar con exactitud los valores que toma una de ellas a partir de los que adopta la otra.

La dependencia estadística entre dos variables estadísticas se da cuando los valores que toma una de ellas están relacionados con los valores de la otra, pero no de manera exacta.

Dos variables estadísticas son independientes cuando no se puede establecer ninguna relación entre los valores que toma una de ellas y los que toma la otra.

38. a) X = Número de horas diarias dedicadas a ver la televisión.

Y = Número de horas de lectura diaria.

b) X = Autonomía de un vehículo en kilómetros recorridos.

Y = Consumo del vehículo en litros.

c) X = Volumen de un libro de bolsillo.

Y = Número de páginas de un libro.

d) X = Número de hijos de una familia.

Y = Consumo familiar de agua en litros.

39. La regresión lineal es el análisis que pretende determinar la recta que mejor se aproxima a un diagrama de dispersión. La recta resultante se denomina recta de regresión.

Las rectas de regresión permiten predecir el valor de una de las variables a partir de la otra.

40. a) Dependencia estadística

b) Dependencia estadística

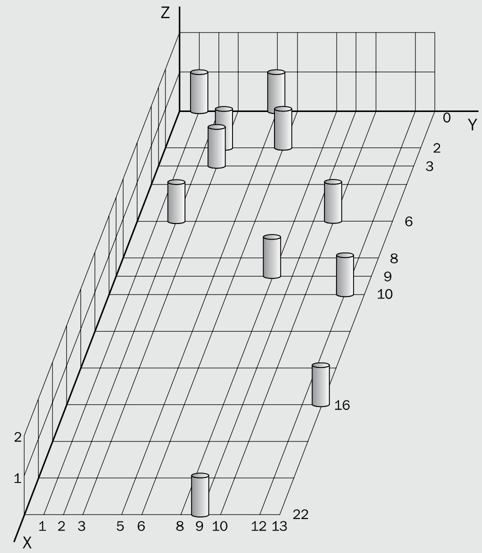
c) Dependencia estadística

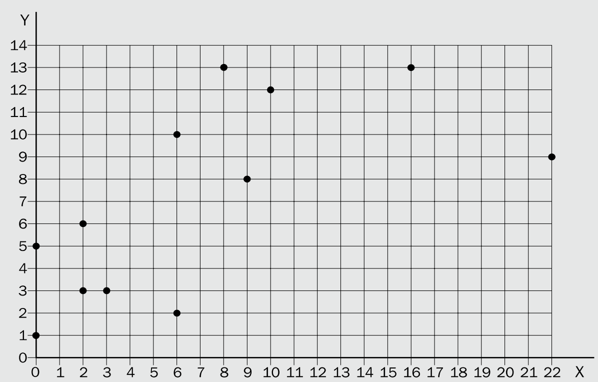
d) Dependencia estadística

e) Variables independientes

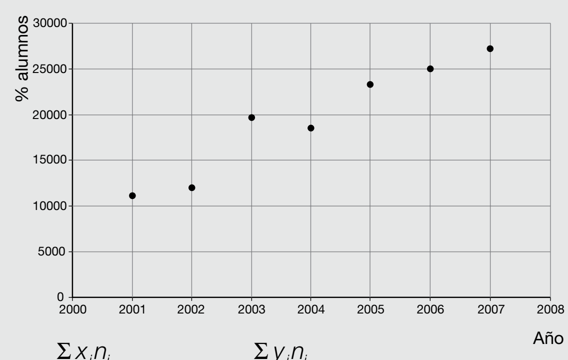
41.

Y \ X	0	2	3	6	8	9	10	16	22	Total
1	1	0	0	0	0	0	0	0	0	1
2	0	0	0	1	0	0	0	0	0	1
3	0	1	1	0	0	0	0	0	0	2
5	1	0	0	0	0	0	0	0	0	1
6	0	1	0	0	0	0	0	0	0	1
8	0	0	0	0	0	1	0	0	0	1
9	0	0	0	0	0	0	0	0	0	1
10	0	0	0	1	0	0	0	0	0	1
12	0	0	0	0	0	0	1	0	0	1
13	0	0	0	0	1	0	0	1	0	2
Total	2	2	1	2	1	1	1	1	1	12





42. a) En el diagrama de dispersión se observa una fuerte correlación lineal y positiva entre las dos variables:



b) $\overline{x} = \dfrac{\sum_i x_i n_i}{N} = 2004$; $\overline{y} = \dfrac{\sum_i y_i n_i}{N} = 19\,534{,}57$

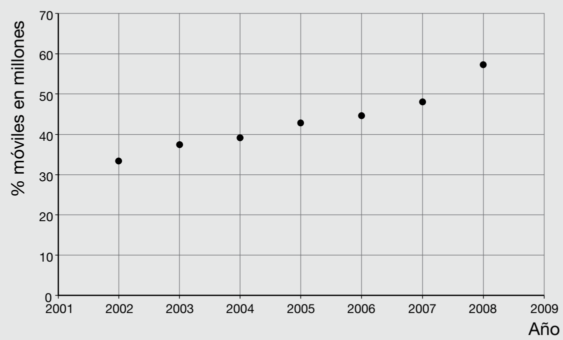
$$\sigma_x = \sqrt{\frac{\sum_i x_i^2 \cdot n_i}{N} - \overline{x}^2} = 2;$$

$$\sigma_y = \sqrt{\frac{\sum_i y_i^2 \cdot n_i}{N} - \overline{y}^2} = 5\,748{,}35$$

$$\sigma_{xy} = \frac{\sum_{i,j} x_i \cdot y_j \cdot n_{ij}}{N} - \overline{x} \cdot \overline{y} = 11116{,}43$$

$r = \dfrac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = 0{,}97$. Este valor es coherente con el resultado del apartado anterior.

43. a) En el diagrama de dispersión se observa una fuerte correlación lineal y positiva:



b) El coeficiente de Pearson que se obtiene con los ocho pares de valores de *X* e *Y* es:

$$\bar{x} = \frac{\sum_i x_i n_i}{N} = 2005; \quad \bar{y} = \frac{\sum_i y_i n_i}{N} = 43{,}31$$

$$\sigma_x = \sqrt{\frac{\sum_i x_i^2 \cdot n_i}{N} - \bar{x}^2} = 2;$$

$$\sigma_y = \sqrt{\frac{\sum_i y_i^2 \cdot n_i}{N} - \bar{y}^2} = 7{,}25$$

$$\sigma_{xy} = \frac{\sum_{i,j} x_i \cdot y_j \cdot n_{ij}}{N} - \bar{x} \cdot \bar{y} = 13{,}99$$

$r = \dfrac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = 0{,}97$. Este valor es coherente con el resultado del apartado anterior.

c) Calculamos la recta de regresión de *Y* sobre *X* sin los datos del año 2008. Se obtiene:

$$y = A + Bx$$

$$A = \bar{y} - \frac{\sigma_{xy}}{\sigma_x^2}\bar{x} = 40{,}98 - \frac{8{,}19}{2{,}92} \cdot 2\,004{,}5 =$$
$$= -5\,592$$

$$B = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{8{,}19}{23{,}24} = 2{,}81$$

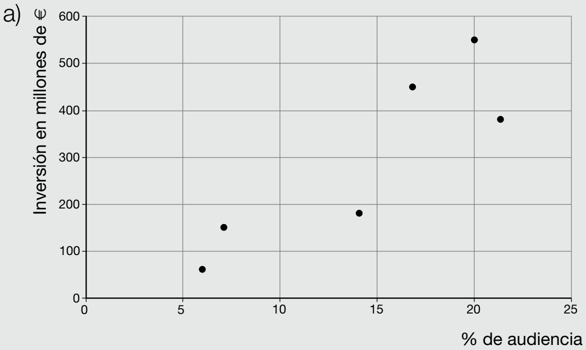
El valor predicho para el año 2008 es:

$y = -5592 + 2{,}81 \cdot 2008 = 50{,}48$

La predicción del año 2008 no se ajusta a una evolución lineal. Si embargo puede ser correcta si, por ejemplo, a partir del año 2007 se espera un aumento de clientes superior al predicho por un simple comportamiento lineal.

d) No es posible efectuar una predicción para el año 2020 porque, según los datos de la tabla, a partir del año 2007 hay un crecimiento distinto del lineal.

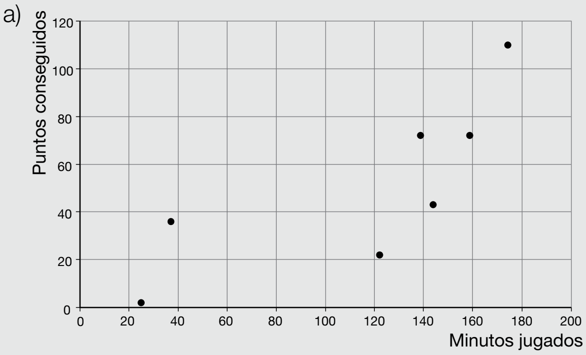
44. a)



b) Hay una correlación positiva aproximadamente lineal.

c) En general, a mayor audiencia, mayor inversión publicitaria, sobre todo en cadenas privadas.

45. a)



La correlación entre ambas variables es positiva pero no es lineal en todo el rango. Esto se debe a que cada componente del equipo tiene sus características propias. Los puntos que un jugador consigue no dependen sólo de los minutos jugados sino también de la posición en que juega (base, ala, pívot).

b) A la vista del gráfico, un componente que ha conseguido 60 puntos es de esperar que haya jugado un tiempo comprendido entre 120 y 160 minutos. Si realizamos una predicción basada en una recta de regresión lineal obtenemos:

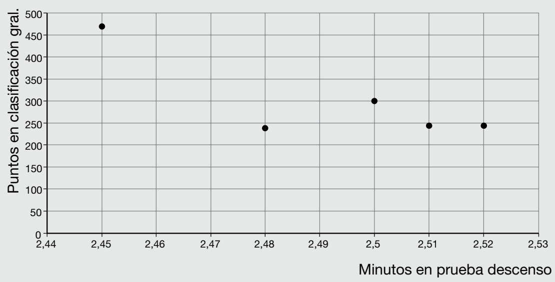
$$A = \bar{y} - \frac{\sigma_{xy}}{\sigma_x^2}\bar{x} = 51 - \frac{1\,436{,}29}{3\,009{,}06} \cdot 114{,}29 = -3{,}55$$

$$B = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{1\,436{,}29}{3\,009{,}06} = 0{,}477$$

$60 = -3{,}55 + 0{,}477 \cdot x \Rightarrow x = 133$ minutos.

No es una predicción demasiado fiable, puesto que la correlación lineal entre las variables no es fuerte. De hecho, si se calcula el coeficiente de Pearson se obtiene r = 0,78, que no está próximo a 1.

46. a)



$$\overline{x}=\frac{\sum_i x_i n_i}{N}=2{,}492;\quad \overline{y}=\frac{\sum_i y_i n_i}{N}=299{,}2$$

$$\sigma_x=\sqrt{\frac{\sum_i x_i^2\cdot n_i}{N}-\overline{x}^2}=0{,}025$$

$$\sigma_y=\sqrt{\frac{\sum_i y_i^2\cdot n_i}{N}-\overline{y}^2}=88{,}334$$

$$\sigma_{xy}=\frac{\sum_{i,j} x_i\cdot y_j\cdot n_{ij}}{N}-\overline{x}\cdot\overline{y}=-1{,}792$$

$$r=\frac{\sigma_{xy}}{\sigma_x\cdot\sigma_y}=-0{,}811$$

b) Calculamos la recta de regresión de Y sobre X:

$$y=A+B\,x$$

$$A=\overline{y}-\frac{\sigma_{xy}}{\sigma_x^2}\,\overline{x}=299{,}2-\frac{-1{,}792}{0{,}000\ 625}\cdot 2{,}492=7\ 444{,}3$$

$$B=\frac{\sigma_{xy}}{\sigma_x^2}=\frac{-1{,}792}{0{,}000\ 625}=-2\ 867{,}2$$

La puntuación que cabe esperar para un tiempo de 2,49 minutos es:

$y = 7444{,}3 - 2867{,}2 \cdot 2{,}49 = 305$

47. No. Debe haber otros factores que justifiquen este hecho, por ejemplo, que tanto las cigüeñas como los seres humanos elijan los territorios con clima más benévolo para vivir.

48. a) $r^2=0{,}64\cdot 0{,}75=0{,}480;\ r=\sqrt{0{,}480}=0{,}69$

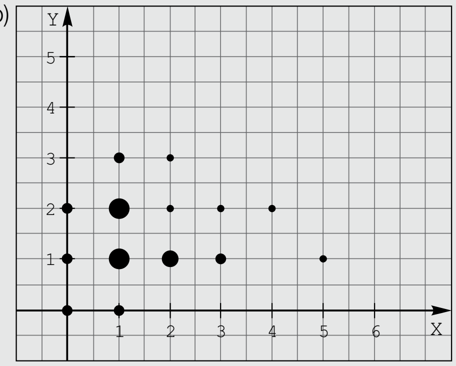
b) $r^2=(-0{,}35)\cdot(-0{,}22)=0{,}077;\ r=\sqrt{0{,}077}=-0{,}28$

En el apartado a) tomamos r positivo puesto que la correlación es positiva. En el segundo apartado, lo tomamos negativo porque la correlación es negativa.

49.

N.º de hijos / N.º de coches	0	1	2	3	4	5	**Total**
0	1	2	2			1	**6**
1	1	5	5	2			**13**
2		4	1	1			**6**
3		2	1				**3**
4			1				**1**
5		1					**1**
Total	**2**	**14**	**10**	**3**	**0**	**1**	**30**

b)



c) X = n.º hijos en la familia

Y = n.º coches en la familia

Media X = 1,43

Media Y = 1,6

Varianza X = 1,2

Varianza Y = 0,95

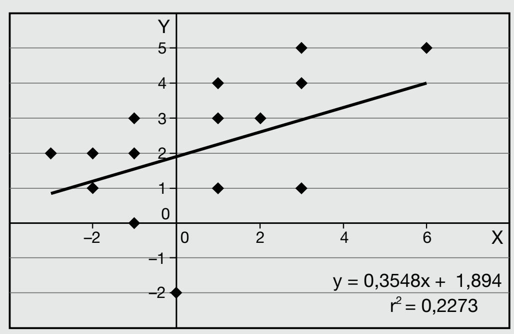
Covarianza = –0,39

Coeficiente de Pearson = –0,34

d) El coeficiente de correlación $r = -0{,}135$ es un valor muy próximo a cero, lo que indica que no podemos afirmar que haya relación entre las dos variables analizadas.

50. Covarianza = –0,11

Coeficiente de Pearson = –0,03



51. Respuesta abierta.

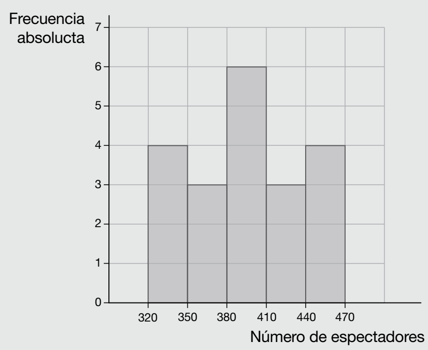
52. Respuesta abierta.

PON A PRUEBA TUS COMPETENCIAS

1. a)

Clases	[320,350)	[350,380)	[380,410)	[410,440)	[440,470)
N.º de espectadores	4	3	6	3	4

b) El gráfico más adecuado es un histograma por tratarse de datos agrupados.



c) La media es:

$$\bar{x} = \frac{452+437+441+463+460+412+405+411+402+399+398+408+390+373+368}{20} +$$

$$+\frac{351+340+338+347+326}{20} = \frac{7921}{20} = 396{,}05$$

d) Los cuartiles son valores que nos dividen la distribución de datos en cuatro partes iguales.

Los datos organizados son:

326 338 340 347 351 368 373 390 398 399 402 405 408 411 412 437 441 452 460 463

$Q_1 = \frac{351+368}{2} = 359{,}5$ $\quad$ $\text{Me} = \frac{399+402}{2} = 400{,}5$ $\quad$ $Q_3 = \frac{412+437}{2} = 424{,}5$

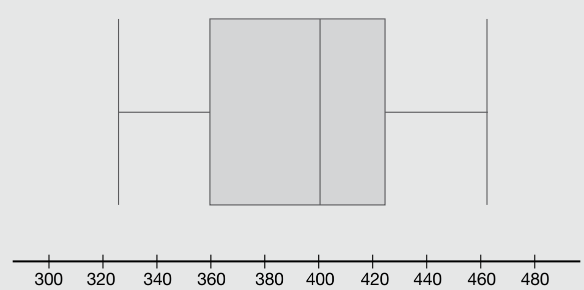
e) El rango intercuartílico es $IQR = Q_3 - Q_1 = 424{,}5 - 359{,}5 = 65$.

Los límites inferior y superior son:

$LI = Q_1 - 1{,}5 \cdot IQR = 360 - 1{,}5 \cdot 65 = 262{,}5$

$LS = Q_3 + 1{,}5 \cdot IQR = 430 + 1{,}5 \cdot 65 = 527{,}5$

Por tanto, el diagrama de caja y bigotes es:



2. a) $1{,}8 = \frac{0 \cdot 8 + 1 \cdot 4 + 2 \cdot 7 + 3 \cdot 2 + 5x + 6 \cdot 1}{25} \Rightarrow 1{,}8 = \frac{30+5x}{25} \Rightarrow 45 = 30 + 5x \Rightarrow 5x = 15 \Rightarrow x = 3$

Por tanto, tres personas habían realizado cinco compras en tiendas *online* en el último año.

b) Nos indica la variabilidad de las distribuciones de datos. Para calcularlo necesitamos la media y la desviación típica.

c) $\sigma^2 = \frac{|0-1{,}8|^2 \cdot 8 + |1-1{,}8|^2 \cdot 4 + |2-1{,}8|^2 \cdot 7 + |3-1{,}8|^2 \cdot 2 + |5-1{,}8|^2 \cdot 3 + |6-1{,}8|^2}{25}$

$= \frac{25{,}92 + 2{,}56 + 0{,}28 + 2{,}88 + 30{,}72 + 17{,}64}{25} = \frac{80}{25} = 3{,}2$

$\sigma^2 = 3{,}2 \Rightarrow \sigma = 1{,}79$ $\quad$ $CV = \frac{\sigma}{\bar{x}} = \frac{1{,}79}{1{,}8} = 0{,}994 = 99{,}4\ \%$

d) Como el CV es mayor que 15 %, la media aritmética es poco representativa de la distribución de datos; esto es, los datos están muy dispersos y no deberíamos hacer esa afirmación.

3. a) X: $\bar{x} = \frac{0+3000+6000+9000+12000+15000}{6} = \frac{45000}{6} = 7500$

Y: $\bar{y} = \frac{760+523+349+226+141+87}{6} = \frac{2086}{6} = 347{,}7$

b) X: $\sigma_x = \sqrt{\frac{\sum_i x_i^2 \cdot n_i}{N} - \bar{x}^2} = \sqrt{\frac{3000^2+6000^2+9000^2+12000^2+15000^2}{6} - 7500^2} = \sqrt{\frac{495000000}{6} - 56250000} =$
$= \sqrt{82500000 - 56250000} = \sqrt{26250000} = 5123{,}5$

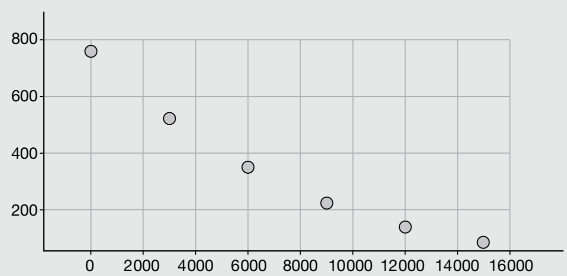
Y: $\sigma_y = \sqrt{\frac{\sum_j y_j^2 \cdot n_j}{N} - \bar{y}^2} = \sqrt{\frac{760^2+523^2+349^2+226^2+141^2+87^2}{6} - 347{,}7^2} = \sqrt{\frac{1051456}{6} - 120895{,}3} =$
$\sqrt{175242{,}7 - 120895{,}3} = \sqrt{54347{,}4} = 233{,}1$

c) $\sigma_{xy} = \frac{\sum_{i,j} x_i \cdot y_j \cdot n_{ij}}{N} - \bar{x} \cdot \bar{y} = \frac{3000 \cdot 523 + 6000 \cdot 349 + 9000 \cdot 226 + 12000 \cdot 141 + 15000 \cdot 87}{6} - 7500 \cdot 347{,}7$
$= \frac{8694000}{6} - 2607750 = 1449000 - 2607750 = -1158750$

d) $r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \frac{-1158750}{5123{,}5 \cdot 233{,}1} = -0{,}97$

Por tanto, existe una correlación lineal negativa fuerte.

e) El diagrama de dispersión es:



La recta de regresión es:

$y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2} \cdot (x - \bar{x}) \Rightarrow y - 347{,}7 = \frac{-1158750}{26250252{,}25} \cdot (x - 7500) \Rightarrow y - 347{,}7 = -0{,}044 \cdot (x - 7500)$
$\Rightarrow y - 347{,}7 = -0{,}044x + 330 \Rightarrow y = -0{,}044x + 677{,}7$

f) Como $y = -0{,}044x + 677{,}7$, para $x = 8\,848$: $y = -0{,}044 \cdot 8848 + 677{,}7 \Rightarrow y = 288{,}4$.

Por tanto, en la cima del monte Everest, la presión atmosférica es 288,4 mmHg.

4. a) X: $\bar{x} = \frac{9500+9488+9482+9477+9473}{5} = \frac{47420}{5} = 9484$

Y: $\bar{y} = \frac{150+190+215+240+265}{5} = \frac{1060}{5} = 212$

b) X: $\sigma_x = \sqrt{\frac{\sum_i x_i^2 \cdot n_i}{N} - \bar{x}^2} = \sqrt{\frac{9500^2+9488^2+9482^2+9477^2+9473^2}{5} - 9484^2} = \sqrt{\frac{449731726}{5} - 89946256} =$
$= \sqrt{89946345{,}2 - 89946256} = \sqrt{89{,}2} = 9{,}4$

Y: $\sigma_y = \sqrt{\frac{\sum_j y_j^2 \cdot n_j}{N} - \bar{y}^2} = \sqrt{\frac{150^2+190^2+215^2+240^2+265^2}{5} - 212^2} = \sqrt{\frac{232650}{5} - 44944} =$
$\sqrt{46530 - 44944} = \sqrt{1586} = 39{,}8$

c) $\sigma_{xy} = \dfrac{\sum_{i,j} x_i \cdot y_j \cdot n_{ij}}{N} - \bar{x} \cdot \bar{y} = \dfrac{9500 \cdot 150 + 9488 \cdot 190 + 9482 \cdot 215 + 9477 \cdot 240 + 9473 \cdot 265}{5} - 9484 \cdot 212$

$= \dfrac{10051175}{5} - 2010608 = 2010235 - 2010608 = -373$

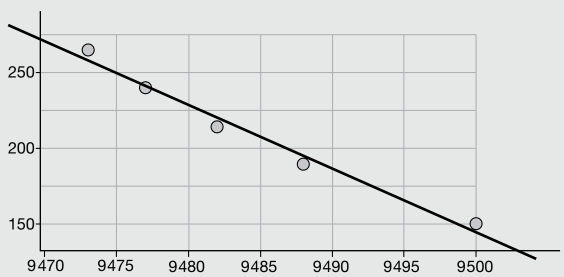
d) $r = \dfrac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \dfrac{-373}{9{,}4 \cdot 39{,}8} = -0{,}997$

Por tanto, existe una correlación lineal negativa fuerte.

e) La recta de regresión es:

$x - \bar{x} = \dfrac{\sigma_{xy}}{\sigma_y^2} \cdot \left(y - \bar{y}\right) \Rightarrow x - 9484 = \dfrac{-373}{1586} \cdot \left(y - 212\right) \Rightarrow x - 9484 = -0{,}235 \cdot \left(y - 212\right)$

$\Rightarrow x - 9484 = -0{,}235y + 49{,}82 \Rightarrow x = -0{,}235y + 9533{,}82$



f) Como $x = -0{,}235y + 9533{,}82$, para y = 180: $x = -0{,}235 \cdot 180 + 9533{,}82 = 9491{,}52$

Por tanto, el jugador habrá conseguido 9492 puntos en ese nivel.

5. a) Sea X = cantidad de vitamina C y Y = cantidad de magnesio.

X: $\bar{x} = \dfrac{228{,}3 + 60{,}9 + 47{,}8 + 36{,}4 + 30 + 53{,}2}{6} = \dfrac{456{,}6}{6} = 76{,}1$

Y: $\bar{y} = \dfrac{22 + 41 + 12 + 10 + 29 + 10}{6} = \dfrac{124}{6} = 20{,}7$

b) X: $\sigma_x = \sqrt{\dfrac{\sum_i x_i^2 \cdot n_i}{N} - \bar{x}^2} = \sqrt{\dfrac{228{,}3^2 + 60{,}9^2 + 47{,}8^2 + 36{,}4^2 + 30^2 + 53{,}2^2}{6} - 76{,}1^2} = \sqrt{\dfrac{63169{,}74}{6} - 5791{,}21} =$

$= \sqrt{10528{,}29 - 5791{,}21} = \sqrt{4737{,}08} = 68{,}8$

Y: $\sigma_y = \sqrt{\dfrac{\sum_j y_j^2 \cdot n_j}{N} - \bar{y}^2} = \sqrt{\dfrac{22^2 + 41^2 + 12^2 + 10^2 + 29^2 + 10^2}{6} - 20{,}7^2} = \sqrt{\dfrac{3350}{6} - 428{,}49} =$

$= \sqrt{558{,}33 - 428{,}49} = \sqrt{129{,}84} = 11{,}4$

c) $\sigma_{xy} = \dfrac{\sum_{i,j} x_i \cdot y_j \cdot n_{ij}}{N} - \bar{x} \cdot \bar{y} = \dfrac{228{,}3 \cdot 22 + 60{,}9 \cdot 41 + 47{,}8 \cdot 12 + 36{,}4 \cdot 10 + 30 \cdot 29 + 53{,}2 \cdot 10}{6} - 76{,}1 \cdot 20{,}7$

$= \dfrac{9859{,}1}{6} - 1575{,}27 = 1643{,}18 - 1575{,}27 = 67{,}9$

d) $r = \dfrac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \dfrac{67{,}9}{68{,}8 \cdot 11{,}4} = 0{,}087$

Por tanto, existe una correlación lineal positiva débil.

e) La nube de puntos es: