Unidad 9

Actividades

1. a) Sobre toda la población, ya que hay un número pequeño de individuos: los alumnos de la clase.

b) Sobre una muestra, ya que el número de individuos es muy grande.

c) Sobre toda la población.

d) Sobre una muestra.

2. Primero se comprobará qué porcentaje sobre el total representan los alumnos de cada curso de ESO. A continuación, se agrupará la muestra de 300 alumnos manteniendo los porcentajes de cada grupo.

	Alumnos	% sobre el total	Alumnos de la muestra
1.º de ESO	1300	26 %	26 % * 300 = 78
2.º de ESO	1250	25 %	25 % * 300 = 75
3.º de ESO	1100	22 %	22 % * 300 = 66
4.º de ESO	1350	27 %	27 % * 300 = 81

$\sum = 5\,000$ $\sum = 100\ \%$ $\sum = 300$

3. La serie de datos está constituida por datos no agrupados, por lo que la tabla de frecuencias es de la siguiente forma:

N.º de hermanos	n_i	f_i	N_i	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

$\sum n_i = 25 \quad \sum f_i = 1$

4. La serie está constituida por datos agrupados, por lo que se deberán crear los intervalos necesarios.

El valor máximo es 8,4 y el valor mínimo, 1. Por tanto, el recorrido es $8{,}4 - 1 = 7{,}4$.

Número de datos: N = 26. Número de intervalos: $\sqrt{26} \simeq 5{,}10$. Se tomarán 6 intervalos.

Se considerarán los siguientes intervalos:

[0, 1,5); [1,5, 3); [3, 4,5); [4,5, 6); [6, 7,5); [7,5, 9)

Las correspondientes marcas de clase son:

0,75; 2,25; 3,75; 5,25; 6,75; 8,25

La tabla de frecuencias correspondiente a la temperatura mínima es:

Temperaturas	Marca de clase	n_i	f_i	N_i	F_i
[0 , 1,5)	0,75	4	0,1538	4	0,1538
[1,5 , 3)	2,25	9	0,3462	13	0,5
[3 , 4,5)	3,75	5	0,1923	18	0,6923
[4,5 , 6)	5,25	3	0,1154	21	0,8077
[6 , 7,5)	6,75	2	0,0769	23	0,8846
[7,5 , 9)	8,25	3	0,1154	26	1,0

$\sum n_i = 26 \sum f_i = 1$

5. Diagrama de barras:

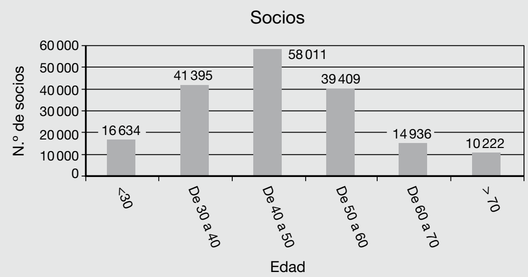
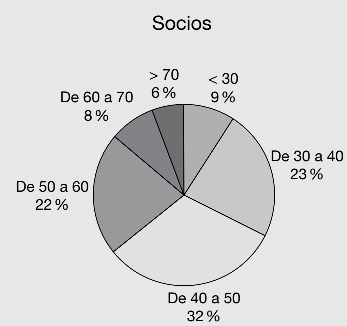


Diagrama de sectores:



6. Media:

$$\bar{x} = \frac{\sum x_i}{N} = \frac{87+63+65+86+92+72+69+90}{8} =$$

$$= \frac{624}{8} = 78$$

Mediana: Para su cálculo deben ordenarse los datos de mayor a menor.

92; 90; 87; 86; 72; 69; 65; 63

N = 8 es par, por lo que la mediana será la media aritmética de los dos valores centrales:

$$Me = \frac{72+86}{2} = 79$$

Recorrido: $Re = 92 - 63 = 29$

Varianza: $\sigma^2 = \dfrac{\sum |x_i - \bar{x}|^2 \cdot n_i}{N} =$

$$\frac{|87-78|^2 + |63-78|^2 + |65-78|^2 + |86-78|^2 + |92-78|^2 + |72-78|^2 + |69-78|^2 + |90-78|^2}{8} =$$

$$\frac{81+225+169+64+196+36+81+144}{8}=$$
$$=\frac{996}{8}=124{,}5$$

Desviación típica: $\sigma=\sqrt{\sigma^2}=\sqrt{124{,}5}=11{,}16$

Coeficiente de variación: $CV=\dfrac{\sigma}{\bar{x}}=\dfrac{11{,}16}{78}=0{,}14$

7. Se trata de una variable estadística bidimensional (X, Y), en la que X = número de horas de televisión e Y = número de horas de estudio.

La muestra podría no ser representativa de la población, ya que se toman 15 alumnos de 4.º de ESO de cada centro, sin tener en cuenta el número de alumnos que cursan ese grado en cada centro.

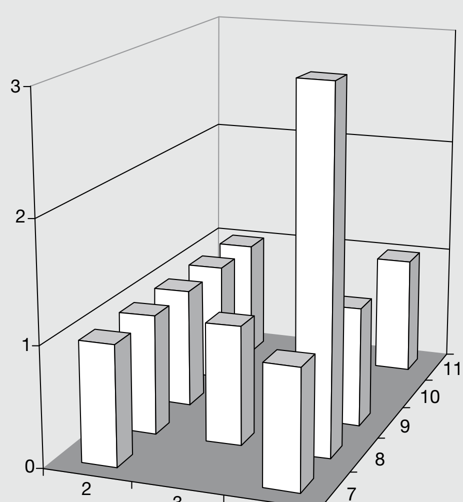
8. X = número de plantas.

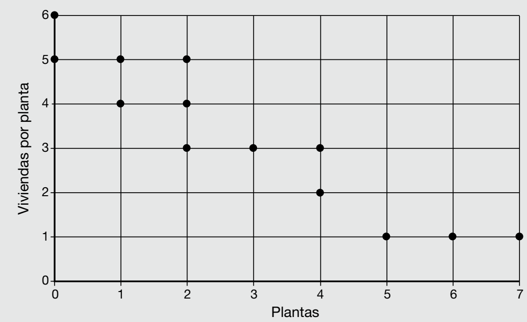
Y = número de viviendas por planta.

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	2
Total	2	5	2	1	2	12

9.

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





10. a) Dependencia funcional lineal.

b) Independientes.

c) Dependencia funcional parabólica.

11. a) Dependencia estadística.

b) Independencia.

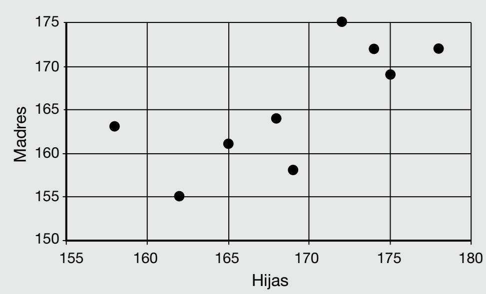
c) Dependencia funcional.

d) Independencia.

e) Dependencia estadística.

f) Independencia.

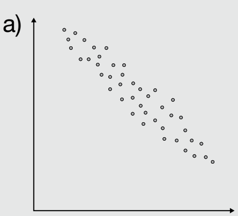
12. a)



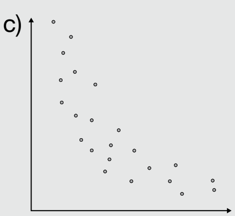
b) Parece una correlación lineal positiva débil.

13. Respuesta abierta. Sirvan como ejemplo los siguientes:

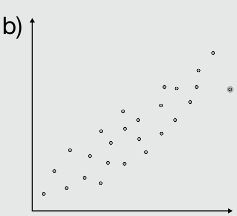
a)



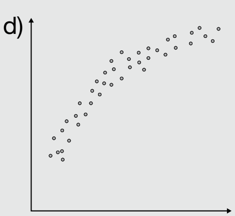
c)



b)



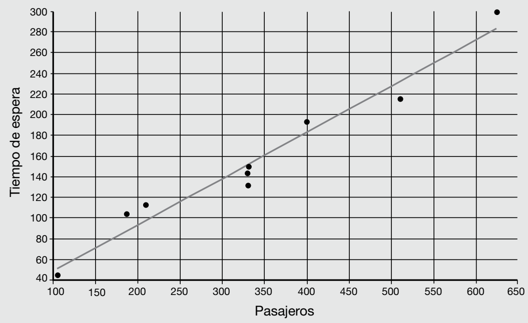
d)



14. Respuesta abierta.

15. Poner en práctica en una hoja de cálculo lo expuesto paso a paso en el ejemplo.

16.



17.

X_i	n_i	N_i	$x_i \cdot n_i$	$\lvert x_i - \bar{x} \rvert$	$\lvert x_i - \bar{x} \rvert^2 \cdot n_i$
4	5	5	20	5,81	169,01
6	3	8	18	3,81	43,64
8	8	16	64	1,81	26,32
10	12	28	120	0,19	0,42
12	7	35	84	2,19	33,45
14	6	41	84	4,19	105,14
16	2	43	32	6,19	76,53

$\sum = 43$ $\sum = 422$ $\sum = 454{,}51$

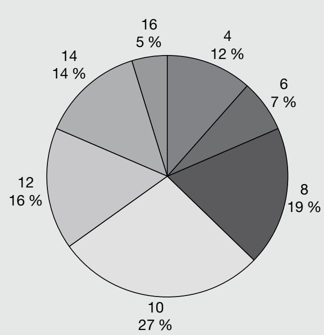
Moda: 10

Media aritmética: $\bar{x} = \dfrac{\sum_i x_i \cdot n_i}{N} = \dfrac{422}{43} = 9{,}81$

Mediana: 10

Varianza: $\sigma^2 = \dfrac{\sum_i \lvert x_i - \bar{x} \rvert^2 \cdot n_i}{N} = \dfrac{454{,}51}{43} = 10{,}57$

Desviación típica: $\sigma = \sqrt{\sigma^2} = \sqrt{10{,}57} = 3{,}25$



Estadística unidimensional

18. a) Alumnos de la ESO.

b) Alumnos de la ESO con una nota media de matemáticas mayor o igual que 7.

c) Se selecciona una muestra representativa para el estudio que se desea.

d) El tamaño de la muestra debe ser adecuado al estudio, teniendo en cuenta que si la muestra es demasiado pequeña, puede llegarse a conclusiones erróneas. Si la muestra es muy grande, el estudio puede encarecerse o alargarse en exceso.

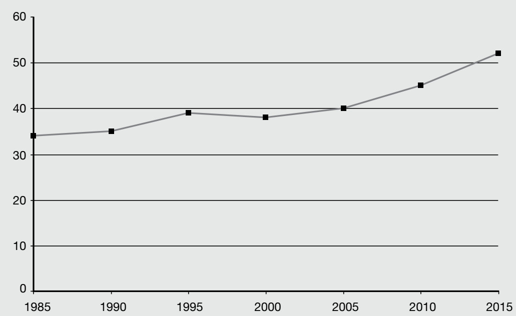
Existen fórmulas estadísticas para orientarnos en el cálculo del tamaño de la muestra.

19. Variables cualitativas: grupo sanguíneo, estado civil, actividad de ocio preferido y preferencia en la alimentación.

Variables cuantitativas discretas: edad y número de hermanos.

Variable cuantitativa continua: peso.

20.

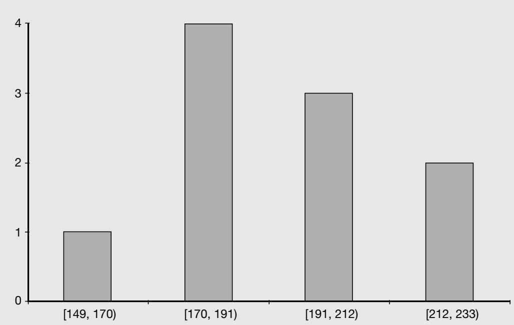


21. El valor máximo es 231 y el valor mínimo, 150. Por tanto, el recorrido es 231 –150 = 81.

Número de datos: $N = 10$. Número de intervalos: $\sqrt{10} \simeq 3{,}16$. Se tomarán 4 intervalos, cuya amplitud será: $\dfrac{81}{4} = 20{,}25 \rightarrow 21$.

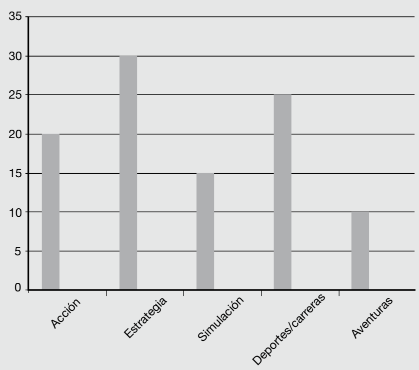
Largo-metrajes	Marca de clase	n_i	f_i	N_i	F_i
[149, 170)	159,5	1	0,1	1	0,1
[170, 191)	180,5	4	0,4	5	0,5
[191, 212)	201,5	3	0,3	8	0,8
[212, 233)	222,5	2	0,2	10	1,0

$\sum = 10$ $\sum = 1$



22.

Tipología	Frecuencia absoluta	% del total
Acción	8	20
Estrategia	12	30
Simulación	6	15
Deportes/carreras	10	25
Aventuras	4	10
	$\sum = 40$	$\sum = 100$



23. 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.

24. a) Histogramas. Son gráficos, sobre ejes, que muestran distribuciones de datos agrupados en clases. Se componen de rectángulos con bases que coinciden con la amplitud de cada clase, y las alturas son de forma que el área de los rectángulos es proporcional a las frecuencias.

Diagrama de barras. Son gráficos, sobre ejes, formados por rectángulos que tienen bases iguales, y las alturas son de forma que el área de los rectángulos es proporcional a las frecuencias absolutas.

b) Para mostrar evolución en el tiempo de una variable.

c) Un gráfico comparativo.

25. Son los valores que se usan en la estadística unidimensional. Los parámetros de centralización informan sobre la tendencia a agruparse hacia un valor central de los datos. Los más frecuentemente utilizados son la moda, la mediana y la media aritmética.

Los parámetros de dispersión son valores que informan sobre la dispersión de los datos, es decir, en qué medida están o no agrupados entorno a los valores centrales. Los más frecuentemente utilizados son el recorrido, la desviación media, la varianza, la desviación típica y el coeficiente de variación.

26. Moda: $Mo = 20$

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

Mediana: $Me = \dfrac{15+20}{2} = 17{,}5$

Recorrido: $r = 25 - 5 = 20$

Desviación media:

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

Varianza:

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

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

Coeficiente de variación: $\dfrac{\sigma}{\bar{x}} = \dfrac{5{,}162}{17{,}059} = 0{,}303$

27. Moda: $Mo = 7{,}5$

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

Mediana:

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

Recorrido: $r = 15 - 0 = 15$

Desviación media:

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

Varianza: $\sigma^2 = \dfrac{\sum_i |x_i - \bar{x}|^2 \cdot n_i}{N} = \dfrac{505{,}62}{50} = 10{,}112$

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

Coeficiente de variación: $\dfrac{\sigma}{\bar{x}} = \dfrac{3{,}180}{9{,}240} = 0{,}344$

28. 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.

Estadística bidimensional

29. 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.

30. 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. Existe una función que relaciona los valores de una variable con los valores de 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.

31. a) Dependencia estadística.

b) Dependencia estadística.

c) Dependencia estadística.

d) Dependencia estadística.

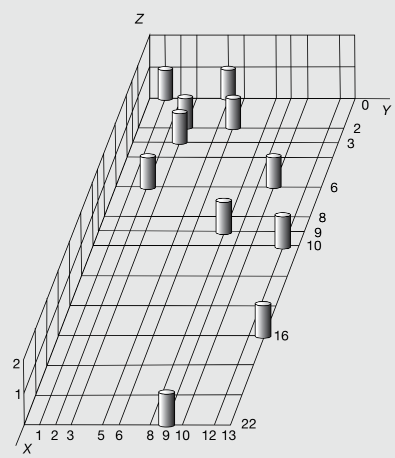
e) Variables independientes.

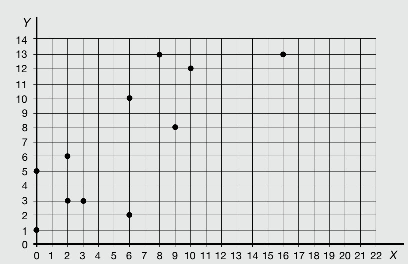
32. La regresión lineal es el análisis que pretende determinar la recta que mejor se aproxima a un diagrama de dispersión. La recta resultante se denomina recta de regresión.

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

33.

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	1	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





34. Para construir la tabla de doble entrada para datos agrupados es necesario agrupar los datos en intervalos.

Intervalos de clase para la variable X:

Recorrido: r = valor máximo – valor mínimo = 39,6 – 7,2 = 32,4.

Número de intervalos de clase (generalmente entre 5 y 20): $N = 15 \rightarrow \sqrt{15} = 3{,}87 \rightarrow 5$ intervalos.

Amplitud de cada intervalo: $\dfrac{32{,}4}{5} = 6{,}48 \rightarrow 7$

Intervalos: [5, 12); [12, 19); [19, 26); [26, 33); [33, 40)

Marca de clase: 8,5; 15,5; 22,5; 29,5; 36,5

Intervalos de clase para la variable *Y:*

Recorrido: r = valor máximo – valor mínimo = 98 – 11 = 87.

Número de intervalos de clase (generalmente entre 5 y 20): $N = 15 \quad \rightarrow \quad \sqrt{15} = 3{,}87 \quad \rightarrow$ 5 intervalos.

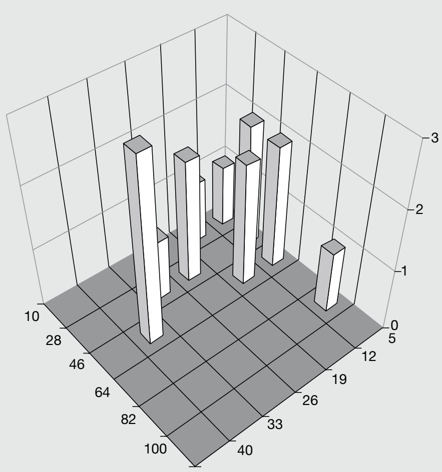
Amplitud de cada intervalo: $\frac{87}{5} = 17{,}4 \quad \rightarrow \quad 18$

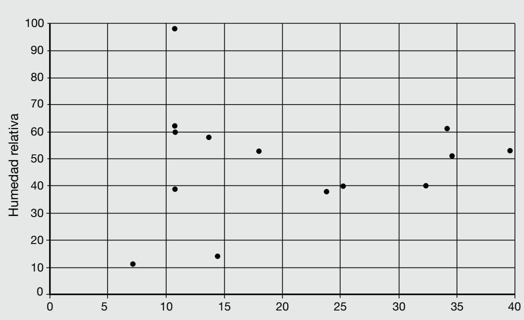
Intervalos: [10, 28); [28, 46); [46, 64); [64, 82); [82, 100)

Marca de clase: 19; 37; 55; 73; 91

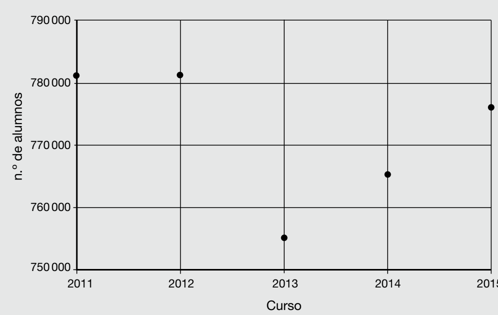
Tabla de doble entrada:

Y \ *X*	[5, 12) 8,5	[12, 19) 15,5	[19, 26) 22,5	[26, 33) 29,5	[33, 40) 36,5	Total
[10, 28) 19	1	1	0	0	0	2
[10, 28) 19	2	0	2	1	0	5
[10, 28) 19	2	2	0	0	3	7
[10, 28) 19	0	0	0	0	0	0
[10, 28) 19	1	0	0	0	0	1
Total	6	3	2	1	3	15





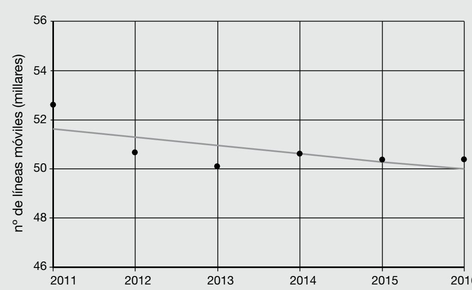
35. a)



b) En los dos primeros cursos existe una relación lineal constante. Los tres últimos también se relacionan linealmente.

c) Correlación lineal positiva fuerte.

36. a)

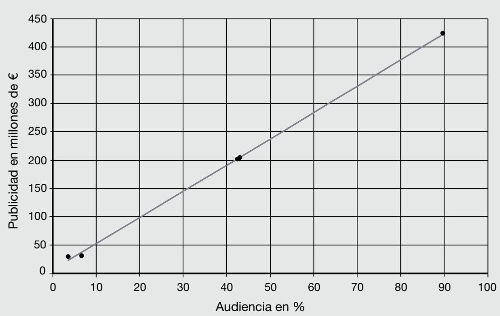


b) Correlación lineal negativa.

c) Si se calcula la recta de regresión de Y sobre X sin los datos del año 2016, se obtiene: $y = 957\,318 - 450{,}3\,x$. Para x = 2016, el valor predicho por la recta es 49 513, que tiene una diferencia del 1,70 % del valor predicho de 50 368, por lo que podemos afirmar que es una buena predicción.

d) Puede predecirse el valor para 2020 a partir de la recta de regresión, pero el periodo de tiempo entre el último dato real, el de 2015, y el dato que se predice es demasiado largo para un mercado como el de la telefonía móvil, sujeto a imprevistos y variaciones que pueden hacer que la realidad sea muy distinta. Sin embargo, como primera aproximación, y siendo conscientes de lo anteriormente dicho, esa predicción puede ser útil.

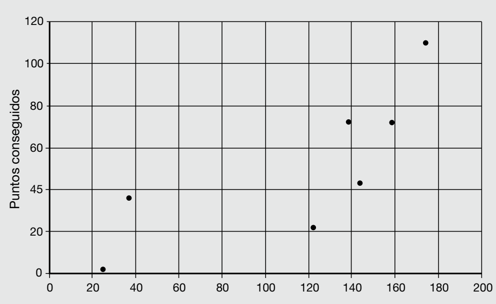
37. a)



b) Relación lineal positiva fuerte.

c) El objetivo de la publicidad es mostrar su producto al mayor número de público posible; por lo tanto, las marcas se interesarán por publicitar sus productos en las emisoras con mayor audiencia.

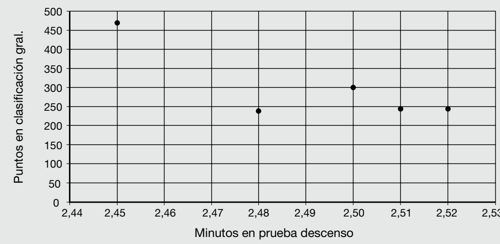
38. a)



b) 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 solo de los minutos jugados, sino también de la posición en la que juega (base, ala, pívot).

c) 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.

39. a)



Sí existe correlación.

b) La puntuación que cabe esperar para un tiempo de 2,49 minutos es:
$y = 7\,444{,}3 - 2\,867{,}2 \cdot 2{,}49 = 305$

40. a)

Salario	n_i	f_i	N_i	F_i
[0, 15]	2 145	0,2145	2 145	0,2145
[15, 20]	1 520	0,1520	3 665	0,3665
[20, 25]	840	0,0840	4 505	0,4505
[25, 30]	955	0,0955	5 460	0,5460
[30, 35]	1 110	0,1110	6 570	0,6570
[35, 40]	2 342	0,2342	8 912	0,8912
[40, 50]	610	0,0610	9 522	0,9522
[50, 100]	328	0,0328	9 850	0,9850
[100, 300]	150	0,0150	10 000	1,0000

$\sum = 10\,000 \quad \sum = 1$

b) Media aritmética:

$$\bar{x} = \frac{\sum_i x_i \cdot n_i}{N} = \frac{293\,800}{10\,000} = 29{,}38$$

El salario medio es de 29 380 €.

c) El salario más frecuente está en el intervalo 35 000-40 000.

El salario tal que la mitad de los restantes sea inferior se corresponde con la mediana:

$$Me = L_i + h \cdot \frac{\frac{N}{2} - N_{i-1}}{n_i} = 25 + 5 \cdot \frac{\frac{10\,000}{2} - 4\,505}{955} =$$
$$= 27\,592$$

d) $Q_1 = \frac{1}{4} \cdot N = \frac{1}{4} \cdot 10\,000 = 2\,500$, el siguiente valor que lo supera es el 2 501, que se encuentra en el intervalo [15, 20].

$Q_3 = \frac{3}{4} \cdot N = \frac{3}{4} \cdot 10\,000 = 7\,500$, el siguiente valor que lo supera es el 7 501, que se encuentra en el intervalo [35, 40].

41.

X	Y	X · Y	X^2	Y^2
1	3	3	1	9
0	–2	0	0	4
1	4	4	1	16
–1	3	–3	1	9
–3	2	–6	9	4
–2	1	–2	4	1
3	4	12	9	16
–2	2	–4	4	4
6	5	30	36	25
–1	2	–2	1	4
3	1	3	9	1
–1	0	0	1	0
1	1	1	1	1
2	3	6	4	9
$\sum = 7$	$\sum = 29$	$\sum = 42$	$\sum = 81$	$\sum = 103$

Medias aritméticas:

$$\bar{x} = \frac{7}{14} = 0{,}5 \qquad \bar{y} = \frac{29}{14} = 2{,}071$$

Covarianza:

$$\sigma_{xy} = \frac{\sum_i x_i \cdot y_i}{N} - \bar{x} \cdot \bar{y} = \frac{42}{14} - 0{,}5 \cdot 2{,}071 = 1{,}964$$

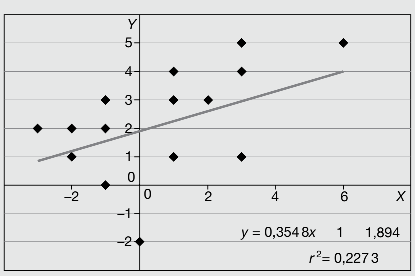
Desviaciones típicas:

$$\sigma_x = \sqrt{\frac{81}{14} - 0{,}5^2} = 2{,}353$$

$$\sigma_y = \sqrt{\frac{103}{14} - 2{,}071^2} = 1{,}752$$

Coeficiente de correlación lineal:

$$r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \frac{1{,}964}{2{,}353 \cdot 1{,}752} = 0{,}476$$



42. Serie 1:

Media X = 108,9

Media Y = 2,75

Covarianza = –10,775

Desviación típica X = 19,463

Desviación típica Y = 1,199

Coeficiente de correlación lineal: R = –0,462

Serie 2:

Media X = 58,458

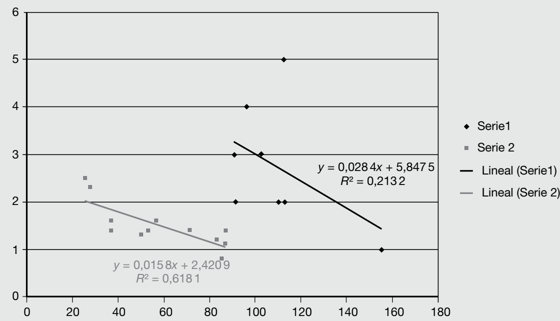
Media Y = 1,5

Covarianza = –8,109

Desviación típica X = 22,689

Desviación típica Y = 0,455

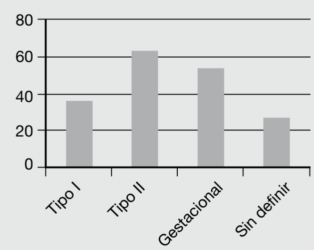
Coeficiente de correlación lineal: R = –0,786

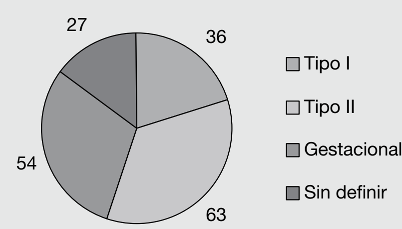


Estadística con recursos digitales

43. Respuesta abierta.

44.



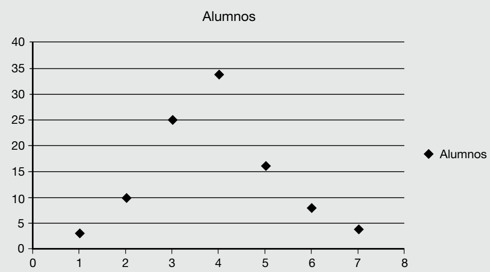


45.

Notas	n_i
1	2
2	1
3	2
4	3
5	6
6	5
7	3
8	2
9	1
Total	25

Moda:	5
Media:	5,12
Mediana:	5
Máximo:	9
Mínimo:	1
Recorrido:	8
Desv. media:	1,5744
Varianza:	4,1056
Desv. típica:	2,02622802
Coef. var:	0,39574766

46.



Problemas

47. Valor máximo = 68

Valor mínimo = 49

Recorrido = 68 – 49 = 19

N.º de intervalos = 4

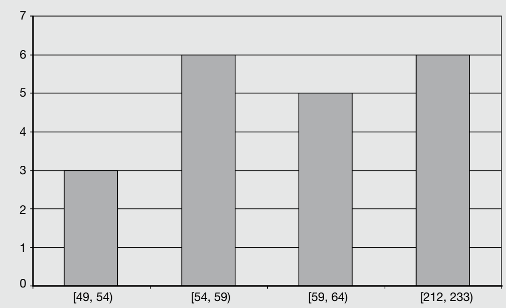
Amplitud del intervalo = $\frac{19}{4} = 4{,}75 \rightarrow 5$

Peso	Marca de clase	n_i	f_i	N_i	F_i
[49, 54)	51,5	3	0,15	3	0,15
[54, 59)	56,5	6	0,30	9	0,45
[59, 64)	61,5	5	0,25	14	0,70
[64, 69)	66,5	6	0,30	20	1,00

$\sum = 20 \quad \sum = 1$

Moda: 59 Mediana: 59

Media: 59,45 Varianza: 29,847 5



48. Equipos de 3.º:

Valor máximo = 10

Valor mínimo = 6

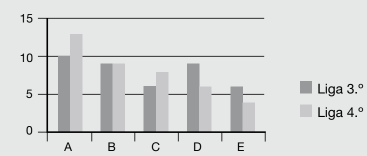
Recorrido = 10 – 6 = 4

Equipos de 4.º:

Valor máximo = 13

Valor mínimo = 4

Recorrido = 13 – 4 = 9



La liga más competitiva es la de los equipos de 3.º, ya que su recorrido es menor por estar las puntuaciones más igualadas.

Pon a prueba tus competencias

1. a)

	Valencia	Sevilla	Cuenca
Mes más cálido y más frío	Julio 29 ˚C Enero 14 ˚C	Julio 35 ˚C Enero 16 ˚C	Agosto 27 ˚C Enero 5 ˚C
Temperatura media anual	21,3 ˚C	22,83 ˚C	16,16 ˚C
Oscilación térmica máxima anual	15 grados	19 grados	22 grados
Mes de mayor precipitación y de menor	Octubre Julio	Marzo Agosto	Mayo Julio
Precipitación media anual	405,4 mm	571,8 mm	557,3 mm
Épocas de sequía y épocas húmedas	No hay época húmeda y la época seca es de marzo a septiembre	No hay época húmeda y la época seca es de marzo a septiembre	No hay época húmeda y la época seca es de julio a agosto

b) Respuesta abierta.

2. En el cálculo de la media intervienen todos los valores entre la década de 1950 y la del 2000, incluido el valor de 1990 que nos falta. Bastará con aplicar la fórmula del cálculo de la media para su obtención:

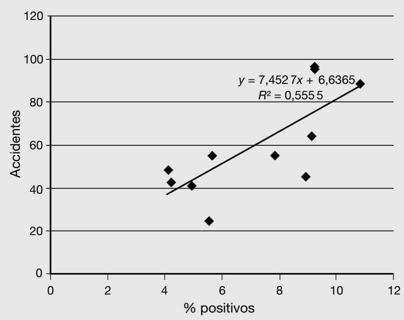
$$\bar{x} = \frac{\sum_i x_i \cdot n_i}{N}$$

$$2{,}43 = \frac{1{,}25 + 1{,}82 + 1{,}95 + 2{,}86 + x_{1990} + 3{,}58}{6}$$

$$14{,}58 = 11{,}46 + x_{1990}$$

$$x_{1990} = 3{,}12$$

3. a)



X (% positivos)	Y (n.º de accidentes)	$X \cdot Y$	X^2	Y^2
9,1	65	591,5	82,81	4 225
4,1	49	200,9	16,81	2 401
8,9	46	409,4	79,21	2 116
9,2	97	892,4	84,64	9 409
5,5	25	137,5	30,25	625
9,2	96	883,2	84,64	9 216
7,8	56	436,8	60,84	3 136
4,2	43	180,6	17,64	1 849
5,6	56	313,6	31,36	3 136
4,9	42	205,8	24,01	1 764
10,8	89	961,2	116,64	7 921
$\sum = 79{,}3$	$\sum = 664$	$\sum = 5\,212{,}9$	$\sum = 628{,}85$	$\sum = 45\,798$

Medias aritméticas:

$$\bar{x} = \frac{79{,}3}{11} = 7{,}209 \qquad \bar{y} = \frac{664}{11} = 60{,}364$$

Covarianza:

$$\sigma_{xy} = \frac{\sum_i x_i \cdot y_i}{N} - \bar{x} \cdot \bar{y} = \frac{5\,212{,}9}{11} - 7{,}209 \cdot 60{,}364 =$$
$$= 38{,}736$$

Desviaciones típicas:

$$\sigma_x = \sqrt{\frac{628{,}85}{11} - 7{,}209^2} = 2{,}280$$

$$\sigma_y = \sqrt{\frac{45\,798}{11} - 60{,}364^2} = 22{,}796$$

Coeficiente de correlación lineal:

$$r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \frac{38{,}736}{2{,}280 \cdot 22{,}796} = 0{,}745$$

b) Sustituyendo $x = 10$ en la recta de regresión:

$$y = 7{,}452\,7\,x + 6{,}636\,5$$

$$y = 7{,}452\,7 \cdot 10 + 6{,}636\,5 = 81{,}163\,5$$

Por lo que cabría esperar 81 accidentes.

4. Respuesta abierta. En función de las comunidades autónomas seleccionadas.