

### 3 Gráficos estadísticos

1. Construye la tabla de frecuencias absolutas y relativas de los datos recogidos en el gráfico de puntos de la Fig. 8.1

$x_i$	$f_i$	$n_i$
3	2	0,10
4	2	0,10
5	3	0,15
5,5	1	0,05
6	4	0,20
6,5	2	0,10
7	2	0,10
7,5	1	0,5
8	1	0,5
8,5	1	0,5
9	1	0,5
	<b>20</b>	<b>1</b>

2. Halla la tabla de frecuencias relativas de los datos de la Fig. 8.3.

Intervalo	$f_i$	$n_i$
[60,70)	18	0,15
[70,80)	36	0,30
[80,90)	48	0,40
[90,100)	12	0,10
[100,110)	6	0,05
	<b>120</b>	<b>1</b>

### 4 Medidas de centralización

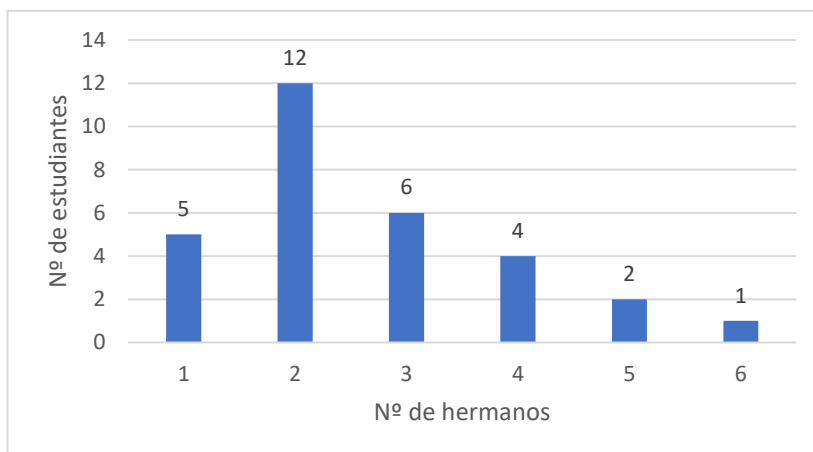
3. Construye la tabla de frecuencias relativas correspondiente al polígono de frecuencias de las Figura 8.4.

Tipo de carburante	$n_i$
Diésel	0,604
Gasolina	0,359
Híbrido	0,003
Eléctrico	0,000
Otros	0,001
Ns/Nc	0,033
	<b>1</b>

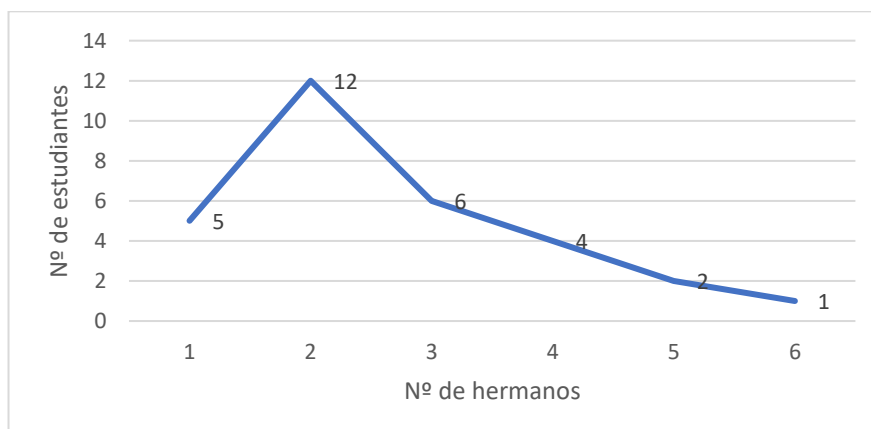
4. Usa los datos de los 30 estudiantes del ejemplo 3 y halla la tabla de frecuencias absolutas y relativas acumuladas. Después, dibuja el diagrama de barras de frecuencias absolutas y el polígono de frecuencias.

Nº de hermanos $x_i$	Fr. absolutas $f_i$	Fr. absolutas acumuladas $F_i$	Fr. relativas $n_i$	Fr. relativas acumuladas $N_i$
1	5	5	0,167	0,167
2	12	17	0,400	0,567
3	6	23	0,200	0,767
4	4	27	0,133	0,900
5	2	29	0,067	0,967
6	1	30	0,033	1
Total	30		1	

- El diagrama de barras de frecuencias absolutas es:



- Polígono de frecuencias absolutas



5. **Deporte y edad.** En un estudio sobre la edad de las personas que practican deporte semanalmente se obtuvieron los siguientes datos en 2015 y 2018. ¿Se puede afirmar que se ha modificado la edad media? Sugerencia: recuerda que debes utilizar la marca de clase de cada intervalo.

	Edad					Total
	[15,25)	[25,35)	[35,45)	[45,55)	[55,65)	
2015	30%	22%	20%	15%	13%	100
2018	26%	23%	19%	18%	14%	100

La edad media de las personas que practican deporte semanalmente se ha incrementado en algo más de un año en estos años.

6. **Peso.** Para evaluar el efecto de una dieta y ejercicio sobre la reducción de peso se eligieron dos grupos de diez personas, uno sin tratamiento y otro con él, durante seis meses. Los resultados del aumento o disminución de peso fueron estos:

Control	-7	-3	-4	-1	-2	5	-1	8	5	9	Estudiantes f <sub>i</sub>	Fr. absolutas Acumuladas F <sub>i</sub>
Tratamiento	-6	-3	5	4	-4	6	-7	3	3	-5		
											45	70
											20	90
											10	100
											Total	100

- a) Calcula la media y la mediana. ¿Qué sugieren ambas medidas en cuanto al efecto del tratamiento?
- b) ¿Cuántos kg redujo el 25% de las personas que más peso perdió en el grupo de tratamiento? ¿Y en el grupo de control?
- a) El grupo de control incrementó su peso medio en 0,9 kg, mientras que el grupo de tratamiento lo redujo en 0,4 kg. Por otra parte, la mitad de las personas del grupo de control perdieron como mucho 1 kg, mientras que en el grupo de tratamiento la mitad se mantuvo igual.
- b) El 25% de las personas que más peso redujo en el grupo de control fue de al menos 5 kg; sin embargo, en el grupo de tratamiento fue de al menos 4 kg.

7. **Adicción móviles.** Calcula e interpreta la mediana y los cuartiles del número de veces que ha consultado el móvil en la última hora una muestra de 100 estudiantes.

N.º de consultas	1	2	3	4
N.º de estudiantes	25	45	20	10

El 50% de los estudiantes han realizado como mucho dos consultas al móvil en la última hora. El 25 % de los estudiantes que menos han consultado el móvil en la última hora lo han hecho como mucho 1,5 veces. El 25 % de los estudiantes que más han consultado el móvil en la última hora lo han hecho al menos tres veces.

## 5 Medidas de dispersión

8. ¿Qué datos están más dispersos, los pesos o las alturas?

<b>Peso (kg)</b>	63	68	60	60	65	78
<b>Altura (cm)</b>	170	180	150	170	168	175

Los pesos están menos dispersos que las alturas.

9. La media y la varianza de una serie de datos son  $\bar{x}=12,3$  y  $s^2=2,5$ . Explica qué modificaciones sufren estos parámetros si se multiplican todos los datos por cinco.

La media transformada queda multiplicada por 5 y la varianza por 25.

10. Las edades, en años, de doce participantes en un curso de vela el verano pasado fueron: 16, 15, 17, 17, 17, 18, 16, 15, 17, 18, 16, 16.

a) Calcula la media de edad y la desviación típica de los participantes.

b) Estos participantes asistirán dentro de dos años a un segundo curso. Razona, sin hacer cálculos, cómo variarán la media de edad y la desviación típica.

a) La media de edad de los participantes en este curso de vela el pasado verano es: 16,5 años

Y la desviación típica: 0,96 años.

b) Si todos los participantes realizan el curso dentro de dos años, la media de edad también se incrementa en dos años. Sin embargo, la variabilidad de los datos, es decir, la desviación típica sigue siendo la misma.

## 6 Variables bidimensionales

11. **Averías de una máquina.** Un concesionario de automóviles vende tres modelos A, B y C. El número de averías de cada modelo durante el primer año de garantía viene reflejado en la siguiente tabla. Calcula la distribución de frecuencias relativas de la distribución conjunta:

Modelos \ Averías	Averías			
	0	1	2	3
A	30	14	8	4
B	12	4	1	0
C	16	8	1	0

Averías	0	1	2	3
Modelos				
A	0,31	0,14	0,08	0,04
B	0,12	0,04	0,01	0
C	0,16	0,08	0,01	0

12. **Control de calidad.** Dos inspectoras, A y B, expertas en el control de calidad de neumáticos, han valorado, de 1 a 3, una muestra de 60 neumáticos. Sea  $x$ , la valoración dada por el inspector A, y sea  $y$ , la valoración dada por el B. Dada la siguiente distribución conjunta de frecuencias relativas elabora la distribución conjunta de frecuencias absolutas.

$x \backslash y$	1	2	3
1	0,05	0,10	0,05
2	0,10	0,15	0,20
3	0,10	0,05	0,20

La tabla de la distribución conjunta de frecuencias absolutas es:

$x \backslash y$	1	2	3
1	3	6	3
2	6	9	12
3	6	3	12

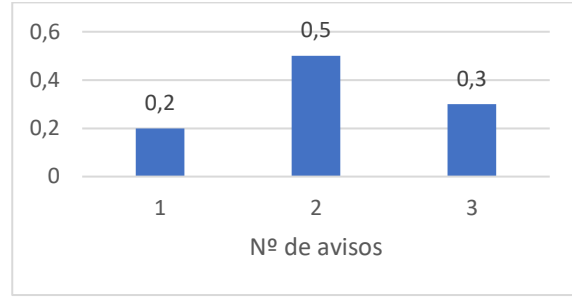
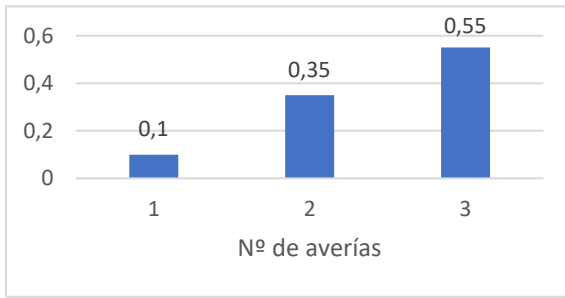
## 7 Distribuciones marginales

13. Sea  $x$  el número de veces que se avería en un día una máquina, e  $y$  el número de veces que se llama a un técnico para arreglarla. La distribución conjunta de frecuencias relativas viene dada por:

$x \backslash y$	1	2	3
1	0,05	0,05	0
2	0,05	0,10	0,20
3	0,10	0,35	0,10

- Calcula la distribución marginal de la variable  $x$ . ¿cuál es el número medio de averías de esta máquina?
  - Halla la distribución marginal de la variable  $y$ . Por término medio, ¿cuántas veces se avisa al técnico?
  - Construye un diagrama de barras comparativo de ambas variables.
  - ¿Dónde se produce mayor variabilidad, en el número de averías de la máquina o en el número de veces que se avisa al técnico?
- El número medio de averías es: 2,45
  - Por término medio se avisa al técnico: 2,1 veces

c) Los diagramas de barras de frecuencias relativas son:



d) Hay mayor variabilidad en el número de avisos al técnico porque tiene un coeficiente de variación mayor.

14. Las variables  $x$  e  $y$  tienen la siguiente distribución conjunta de frecuencias relativas:

$x \backslash y$	1	3	5
2	0,10	0,20	0,10
4	0,15	0,30	0,15

Sabiendo que  $N = 60$ , encuentra la distribución conjunta y las marginales de frecuencias absolutas.

La tabla de la distribución conjunta de frecuencias absolutas es:

$x \backslash y$	1	3	5
2	6	12	6
4	9	8	19

Las distribuciones marginales de frecuencias absolutas son:

$x \backslash y$	1	3	5	$f_{i.}$
2	6	12	6	24
4	9	8	19	36
$f_{.j}$	15	30	15	60

15. Con los datos de la Actividad 12 sobre las inspectoras expertas en control de calidad de neumáticos:
- Encuentra las distribuciones marginales de frecuencias relativas.
  - ¿Cuáles son las valoraciones medias de ambas inspectoras de control de calidad?
  - ¿Qué inspectora es más dispersa en sus puntuaciones?
- a) Las distribuciones marginales de frecuencias relativas de las valoraciones de ambas expertas son:

Experta A	Fr. relativas
$x_i$	$n_i$
1	0,20
2	0,45
3	0,35
Total	1

Experta B	Fr. relativas
$y_j$	$n_j$
1	0,25
2	0,30
3	0,45
Total	1

- b) La valoración media de la inspectora A es: 2,15

Y la valoración media de la inspectora B es: 2,2

- c) La inspectora B presenta más variabilidad que la de la inspectora A.

## 8 Distribución condicionada

16. **Tiempo y número de usuarios en una web.** Un servidor que alberga una determinada página web ha facilitado información sobre el tiempo (en segundos) de acceso ( $x$ ) y el número de usuarios que acceden simultáneamente ( $y$ ) a dicha página web.

$x \backslash y$	5	10	15	20
[0,5)	6	8	21	14
[5,10)	27	36	45	18
[10,15)	19	17	33	6

- Calcula el tiempo medio de acceso y el número medio de usuarios que acceden simultáneamente.
- Si el tiempo de acceso es inferior a 5 s, calcula la distribución del número de usuarios en simultaneidad.
- ¿Cuál es el tiempo medio de acceso, cuando hay 15 personas accediendo simultáneamente?

d) Encuentra la distribución de frecuencias relativas del número de usuarios que acceden simultáneamente, cuando el tiempo de acceso es de al menos 5 minutos.

a) El tiempo medio de acceso es de 8,02 segundos. El número medio de usuarios que acceden simultáneamente es 12,46.

b) La distribución del número de usuarios en simultaneidad, sabiendo que el tiempo de acceso es inferior a 5 segundos, es la distribución condicionada ( $y/x < 5$ ). Así:

Intervalo	$x_i$	Fr. absolutas $f_i$
[0,5)	2,5	49
[5,10)	7,5	126
[10,15)	12,5	75
Total		250

c) Si hay 15 usuarios accediendo simultáneamente, entonces el tiempo medio de acceso es de 8,1 segundos.

d) La distribución de frecuencias relativas del número de usuarios que acceden simultáneamente, cuando el tiempo de acceso es de al menos 5 minutos, es ( $y/x \geq 5$ )

Por tanto:

( $y/x \geq 5$ )	Fr. absolutas	Fr. relativas
5	46	0,23
10	53	0,26
15	78	0,39
20	24	0,12
Total	201	1

**17. Control de calidad.** Continuando con la valoración de las expertas en control de calidad de neumáticos, halla la distribución de la valoración de la segunda experta ( $y$ ), suponiendo que la experta A hizo una valoración igual a 2. ¿Cuál es la valoración media en este caso?

Suponiendo que la inspectora A ha realizado una puntuación de 2, entonces la puntuación media de la inspectora B es de 2,22.



18. **Tarjetas bancarias.** Se ha realizado un estudio sobre el tipo de tarjeta utilizada habitualmente y el gasto mensual en euros en un grupo de 302 personas. Los resultados han sido los siguientes:

Tipo tarjeta \ Gasto	Gasto			
	[0, 50)	[50, 100)	[100, 150)	[150, 200)
Débito	20	42	56	32
Crédito	30	16	28	16
Comercio	4	10	18	30

- a) ¿Cuál es el gasto medio con estas tarjetas? ¿Qué tipo de tarjeta utiliza la mayoría?
- b) Calcula la distribución del gasto mensual de las personas que utilizan tarjeta de crédito. ¿Cuál es el gasto medio en este caso?
- c) Halla la distribución del tipo de tarjeta condicionada a aquellas personas que gastan al menos 100€ al mes.
- a) El gasto medio con estas tarjetas es: 108,777 €. Una mayoría de 150 personas utilizan tarjeta de débito.
- b) El gasto medio de las personas que utilizan tarjeta de crédito es de 91,67 €
- c) La distribución del tipo de tarjeta condicionada a aquellas personas que gastan al menos 100 € al mes es  $(x/y \geq 100)$ . Por tanto:

$(x/y \geq 100)$	$f_j$
Débito	88
Crédito	44
Comercio	48
Total	180

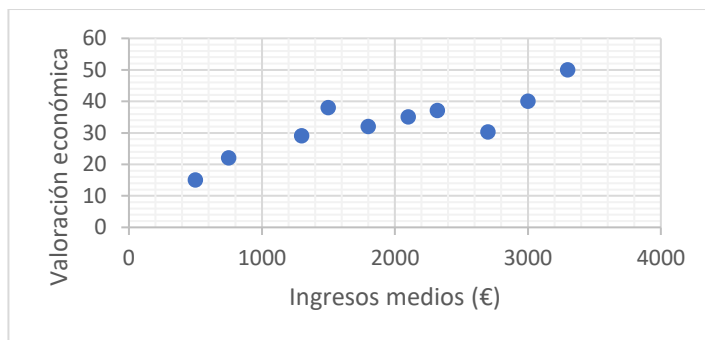
## 9 Diagrama de dispersión. Correlación

19. **Ingresos y valoración económica.** En un estudio sociológico sobre la valoración de la situación económica, se indica que esta viene determinada fundamentalmente por los ingresos brutos mensuales (en €) de cada persona. Para contrastar esta información se obtuvieron los datos relativos a una muestra de diez personas, cuyos ingresos y cuya valoración (en una escala de 0 a 100) de la situación económica aparecen reflejados en la tabla inferior:

- Construye el diagrama de dispersión. ¿Se deduce que, cuanto mayores son los ingresos, mayor es la valoración de la situación económica?
- Halla la covarianza y el coeficiente de correlación e interprétalos.

<b>Ingresos medios (x)</b>	750	500	1300	2320	3300	2100	1800	3000	2700	1500
<b>Valoración económica (y)</b>	22	15	29	37,1	50	35	32	40	30,3	38

a) El diagrama de dispersión es:



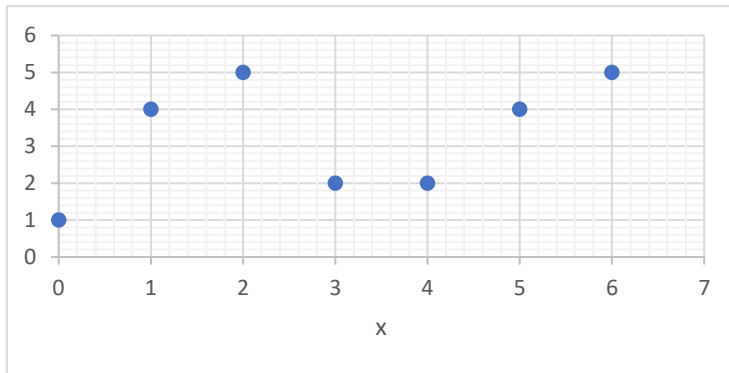
b) La covarianza es: 6985,52 · puntos €  
El coeficiente de correlación es: 0,85

20. Dados los siguientes pares de datos:

<b>x</b>	0	1	2	3	4	5	6
<b>y</b>	1	4	5	2	2	4	5

- Dibuja su diagrama de dispersión y analiza si observas una relación lineal entre ellos.
- Calcula las medias, desviaciones típicas y covarianza.
- Halla el coeficiente de correlación.
- ¿Consideras aceptable la suposición de relación lineal entre ambos pares de datos?

a) Su diagrama de dispersión es:



b) Las medias de ambos datos son: 3 y 3,29

La covarianza es: 1,27

Las desviaciones típicas son: 2 y 1,48

c) El coeficiente de correlación es: 0,43

d) Una vez hallados el diagrama de dispersión y el coeficiente de correlación no parece muy razonable aceptar la hipótesis de relación lineal entre ambos pares de datos.

## 10 Recta de regresión. Estimaciones

21. **Resultados de una inversión.** Se ha llevado a cabo un experimento para averiguar si existe una relación lineal entre el capital inicial ( $x$ ) y el capital final ( $y$ ) en euros de seis pequeños inversores, después de invertir en un fondo inmobiliario durante seis meses. En caso de que así sea, estima el capital final para una persona que invirtiera inicialmente 3 000 €. ¿Es fiable esta estimación? Razónalo.

Capital inicial ( $x$ )	Capital final ( $y$ )
1 000	1 350
2 000	2 600
2 500	3 450
3 200	4 750
4 100	6 300
5 600	8 400

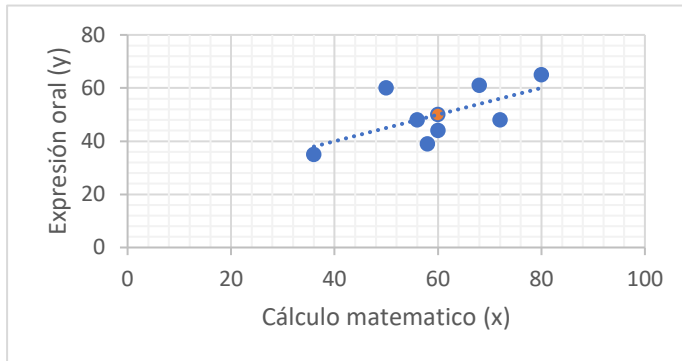
Si se invierte un capital inicial de 3000 € en este fondo de inversión inmobiliario se estima obtener un capital final de aproximadamente 3471 €.

22. **Resultado test.** Un test evalúa la capacidad de cálculo matemático ( $x$ ) y la expresión oral ( $y$ ) de una persona. Los resultados de ocho personas sometidas a este test han sido:

Cálculo matemático ( $x$ )	36	80	50	58	60	56	72	68
Expresión oral ( $y$ )	35	65	60	39	44	48	48	61

- Dibuja un diagrama de dispersión. Calcula las medias y represéntalas en el diagrama anterior.
- Determina la recta de regresión de  $y$  sobre  $x$ .
- Si una persona ha obtenido una valoración de 60 en capacidad de cálculo matemático, ¿qué valor cabe esperar que obtenga en expresión oral?

- a) Las capacidades medias de cálculo matemático y expresión oral de este grupo de ocho personas son, respectivamente, 60 y 50. Su diagrama dispersión en el que hemos incluido las medias respectivas es:



- b) La recta de regresión de  $y$  sobre  $x$  es:  $y = 0,50x + 19,9$
- c) Si una persona ha obtenido una valoración de 60 en capacidad de cálculo matemático, entonces su capacidad en expresión oral se aproxima a una valoración de 50.

## ➤ Actividades finales

### ✚ Variables unidimensionales, medidas de centralización y dispersión

23. El gasto semanal de diez jóvenes es: 14, 18, 18, 20, 35, 40, 40, 40, 40 y 500 €. ¿Qué medida de tendencia central representa mejor a estos datos? ¿Cuál descartarías?

Quizás deberíamos utilizar la moda ya que la mayoría (cuatro de los diez) gasta 40 euros a la semana.

Por otro lado, descartaríamos la media porque se ve muy afectada por el valor extremo de 500 euros y no representa a los gastos semanales de los diez jóvenes.

24. Se dispone de diez datos de una variable estadística. Si sabemos que su suma es 50 y que la suma sus cuadrados es 1025, ¿cuál es valor de la desviación típica? La desviación típica es: 8,8

25. La duración en horas (h) de 100 baterías de teléfonos móviles da como resultado una varianza de  $0,85 h^2$ . Si midiéramos la duración en minutos, ¿cuál sería el valor de la varianza?

$s^2 = 0,85h^2$  entonces para convertirlo a  $min^2$ . habrá que multiplicar por  $60^2$ . Luego,

$$s^2 = 0,85 \cdot 60^2 = 3060 min^2$$

26. En una muestra sobre el gasto en rebajas en una población dio como resultado que el 60% gastaba menos de 100€, mientras que el 25% gastaba por los menos 200€. Indica en términos de cuartiles y percentiles qué conocemos.

Si el gasto en rebajas en una población dio como resultado que el 60% gastaba menos de 100€ están informando del percentil 60. Por otro lado, el 25% de la población gastaba por los menos 200€ están informando del percentil 75, o lo que es equivalente, del cuartil tercero.

27. Calcula e interpreta la mediana del número de clientes que entran en un supermercado, de 10 a 12 de la mañana, durante un periodo de 60 días recogido en la tabla.

N.º de clientes $x_i$	0	1	2	3	4	5	6	7	8	Total
Fr. absolutas $f_i$	5	7	3	12	10	7	8	6	2	60

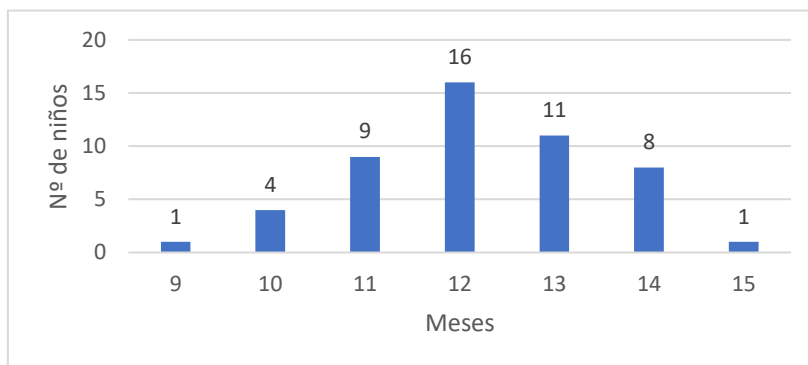
Entre las diez y las doce de la mañana, entran 4 clientes o menos.

28. Un pediatra ha seleccionado una muestra de 50 niños con el fin de realizar un estudio estadístico sobre el mes de vida en que comenzaron a caminar. Los datos son los siguientes:

Meses	9	10	11	12	13	14	15
N.º de niños	1	4	9	16	11	8	1

- Construye un diagrama de barras.
- Calcula la media, la mediana y la moda e interprétalas.
- Halla la desviación típica.
- ¿En qué mes empezó a caminar el 25% de los niños más precoces

- a) El diagrama de barras correspondiente es:



- Por término medio los niños comienzan a andar a los 12, 2 meses. La mayoría de los niños comienza a andar a los 12 meses. El 50 % de los niños comienza a andar como mucho a los 12 meses.
- La desviación típica es: 0,748 meses
- El 25% de los niños más precoces empezaron a andar a los 11 meses.

29. Un joven compró 1000 acciones de Picaresca S.A a 5€ la acción. Ahora la cotización ha bajado y su precio es de 4€. ¿Cuántas acciones debe comprar para que el precio medio de sus acciones sea de 4,2€?

Debe comprar 4000 acciones a un precio de 4 € para que el precio medio sea de 4,2 €.

30. Si de 20 datos numéricos se conoce la media y la desviación típica. Entonces, ¿las siguientes afirmaciones son ciertas o falsas?

- a) Al multiplicar todos los datos por 5, su media permanece inalterada.
- b) Al sumar a todos los datos 5 unidades, su desviación típica se incrementa en 5 unidades.
- c) Al multiplicar a todos los datos por 5, su desviación típica queda multiplicada por 5.

- a) Falsa. La media queda multiplicada por 5.
- b) Falsa. La desviación típica no cambia.
- c) Verdadera.

31. En un examen tres compañeros han obtenido una nota media de 8, una mediana de 9 y dos de ellos la misma nota. ¿Cuáles han sido sus notas?

Las notas de los tres compañeros son: 6, 9 y 9.

32. En un proceso de selección para un puesto de trabajo en una empresa, una de las pruebas consta de seis bloques de seis preguntas cada uno, tipo test. En la tabla figura el número de fallos de los aspirantes Carmen y Pablo. ¿A qué candidato seleccionarías? ¿Por qué?

Bloques	1	2	3	4	5	6
Pablo	5	6	2	1	2	0
Carmen	3	4	5	3	4	5

Se debería seleccionar a Pablo porque ha tenido unos resultados más uniformes al tener un coeficiente de variación mucho menor que el de Carmen.

33. Dos portales especializados en ofertas de empleo, A y B, han publicado una media de 10 y 20 ofertas de trabajo, respectivamente, con varianzas 4 y 9. ¿Qué portal de empleo presenta mayor dispersión absoluta? ¿Y relativa?

El portal A presenta mayor dispersión relativa.

34. En una población, el número medio de aulas por colegio es de 12, y el número medio de estudiantes por aula es de 30. ¿Cuál es el número medio de estudiantes por colegio?

El número medio de estudiantes por colegio en esa población es de 360.

35. Las notas medias y desviaciones típicas de dos grupos A y B, vienen dados en la tabla siguiente:

Grupo	Media	D. típica
A	6,4	0,8
B	5,9	0,9

Si Nicolás obtuvo un 7,3, en el grupo A, y Lucía un 7, en el grupo B, ¿cuál de los dos puede considerarse que obtuvo mejor nota dentro de su grupo?

Ten en cuenta: Se llama puntuación típica a la siguiente medida adimensional:

$$z = \frac{x - \bar{x}}{s}$$

, que mide cuánto se separa la desviación típica de la media.

Puede considerarse que Lucía obtuvo mejor nota dentro de su grupo.



## Variables bidimensionales. Distribución conjunta, marginal y condicional

36. Calcula las distribuciones marginales, y halla la distancia media recorrida y el tiempo medio empleado en ir al colegio de un grupo de 50 estudiantes recogidos en esta tabla. En ella se presenta la distancia que recorren diariamente (x), en kilómetros, y el tiempo empleado (y), en minutos.

t (min) (y) \ d (km) (x)	[0,10)	[10,20)	[20,30)	[30,40)
[0,2)	3	5	4	2
[2,4)	2	3	7	2
[4,6)	0	1	2	6
[6,8)	0	0	8	5

La distancia media recorrida diariamente en ir al colegio por este grupo de estudiantes es de 4,34 kilómetros.

El tiempo medio empleado en ir al colegio diariamente por este grupo de estudiantes es de 24,2 minutos.

## ➤ Diagrama de dispersión, covarianza y regresión

37. De dos variables x e y se conocen los siguientes datos. Calcula la covarianza y el coeficiente de correlación.

$$n = 30; \sum x_i = 41,2 ; \sum y_i = 63,8 ; \sum x_i y_i = 118,7 ; \sum x_i^2 = 188,2 ; \sum y_i^2 = 296,4 .$$

La covarianza es: 1,0386

El coeficiente de correlación es: 0,214

38. Sea  $y = 3x - 10$ . la recta de regresión de y sobre x. Sabiendo que  $\bar{x} = 4, s_x^2 = 9$ . y  $r = 1$ , encuentra la media y la varianza de x, la covarianza y la recta de regresión de x sobre y.

La recta de regresión de x sobre y es:  $x = \frac{s_{xy}}{s_y^2}(y - \bar{y}) + \bar{x} \Rightarrow x = \frac{3}{9}(y - 2) + 4 \Rightarrow x = \frac{1}{3}y + \frac{10}{3}$

39. De dos variables  $x$  e  $y$  se conocen los siguientes datos:

$$\bar{x}=12 ; \bar{y}=36 ; s_x=4. ; s_y=6 ; r=0,8$$

Para encontrar las rectas de regresión necesitamos hallar previamente la covarianza. Así:

$$r = \frac{s_{xy}}{s_x \cdot s_y} \Rightarrow 0,8 = \frac{s_{xy}}{4 \cdot 6} \Rightarrow s_{xy} = 19,2$$

40. Calcula las rectas de regresión de  $y$  sobre  $x$  y de  $x$  sobre  $y$

La recta de regresión de  $y$  sobre  $x$ :

$$y = \frac{s_{xy}}{s_x^2}(x - \bar{x}) + \bar{y}$$

$$y = \frac{19,2}{16}(x - 12) + 36$$

$$y = 1,2x + 21,6$$

La recta de regresión de  $x$  sobre  $y$  es:

$$x = \frac{s_{xy}}{s_y^2}(y - \bar{y}) + \bar{x}$$

$$x = \frac{19,2}{36}(y - 36) + 12$$

$$x = 0,53y - 7,2$$

## ➤ Aplicaciones

41. **Producción defectuosa.** En un estudio sobre una fábrica de producción reunimos una serie de datos para determinar si los artículos defectuosos producidos por los trabajadores dependían del turno de mañana, tarde o noche. Los datos obtenidos son:

		Artículos defectuosos	Artículos no defectuosos
Turno	Mañana	40	910
	Tarde	60	890
	Noche	85	850

a) Calcula las distribuciones marginales.

b) Halla la distribución de artículos defectuosos por turnos y analiza los resultados en términos de porcentajes.

a) Las distribuciones marginales de frecuencias absolutas de ambas variables las obtenemos al sumar las filas y las columnas.

		Fr. absolutas
Turno	Mañana	950
	Tarde	950
	Noche	935
	Total	2835

	Artículos defectuosos	Artículos no defectuosos	Total
Fr. absolutas	40	910	2835

b) Del total de artículos defectuosos proceden del turno de mañana el 22 %, del turno de tarde el 32 % y del de noche el 46 %. Como el número total de artículos producidos en cada turno es muy similar (algo menor en el de noche) parece razonable señalar al turno de noche como el mayor responsable de producir artículos defectuosos.



42. **Ingresos compañías de telefonía.** Según la Comisión Nacional de los Mercados y la Competencia (CNMC), los ingresos totales de telefonía móvil por operador, en millones de euros, durante el periodo 2012 - 2016, vienen dados en la siguiente tabla:

Ingresos totales de telefonía móvil por operador (millones de euros)					
	2012	2013	2014	2015	2016
<b>Movistar</b>	3 790,40	2 656,28	2 006,93	1 654,34	1 466,19
<b>Vodafone</b>	2 596,17	2 095,75	1 673,35	1 606,12	1 571,11
<b>Orange</b>	2 014,95	1 694,93	1 222,12	951,20	1 031,07
<b>Yoigo</b>	432,32	361,89	303,55	248,57	213,31
<b>Resto</b>	651,83	768,03	956,66	734,65	513,24
<b>Total</b>	<b>9 485,67</b>	<b>7 576,88</b>	<b>6 162,61</b>	<b>5 203,88</b>	<b>4 794,92</b>

- a) ¿Cuáles fueron los ingresos marginales de cada una de las compañías durante el período 2012-2016?
- b) ¿Qué porcentaje del total representa cada una de ellas?
- c) ¿Cuáles fueron los ingresos de cada una de las tres primeras compañías durante 2016? ¿En qué términos de distribución condicionada lo expresarías?
- d) ¿A qué crees que se debe la reducción de ingresos de las compañías durante este periodo?

	Total
Movistar	11.574,14
Vodafone	9.542,50
Orange	6.914,27
Yoigo	1.559,64
Resto	3.624,41
<b>Total</b>	<b>33.223,96</b>

- a) Los ingresos marginales de cada una de las compañías durante el período 2012-2016 han sido:
- b) El porcentaje que representa cada una de las compañías en este periodo viene dado en la siguiente tabla:
- c) Los ingresos totales de telefonía móvil para las tres primeras compañías (millones de euros) han sido: 4068,37
- d) Al aumento de la competencia, la fusión de empresas de telefonía, la reducción del precio de los contratos y la incorporación de ofertas de internet en casa, televisión por cable y móvil.

	Total (%)
Movistar	34,84
Vodafone	28,72
Orange	20,81
Yoigo	4,69
Resto	10,91
<b>Total</b>	<b>100</b>

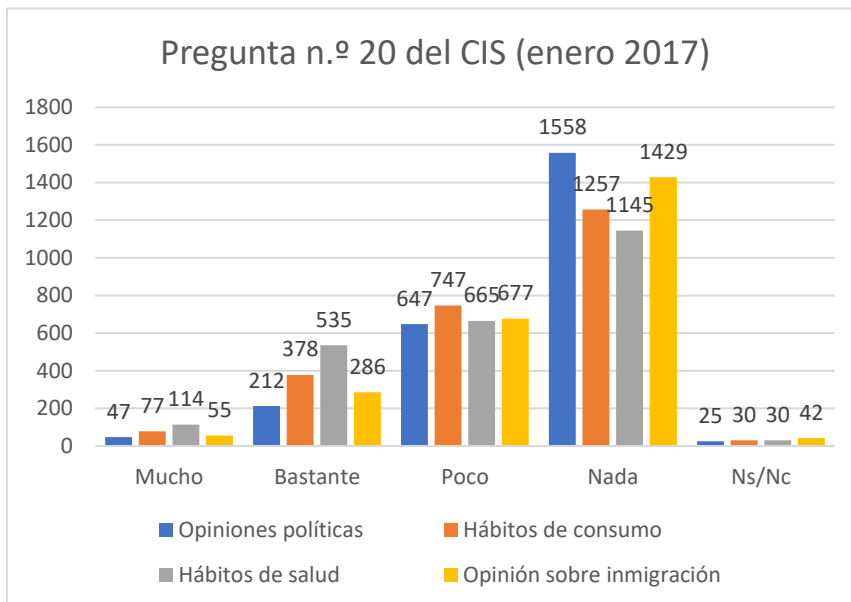
43. **Barómetro del CIS.** La pregunta nº 20: “¿En qué medida le han podido influir los resultados de las encuestas en su(s)...?”, del barómetro de enero de 2017 del CIS, se realizó con una muestra de 2489 personas y se obtuvieron estas respuestas (en porcentaje):

	Mucho	Bas- tante	Poco	Nada	Ns/Nc
<b>Opiniones políticas</b>	1,9	8,5	26	62,6	1
<b>Hábitos de consumo</b>	3,1	15,2	30	50,5	1,2
<b>Hábitos de salud</b>	4,6	21,5	26,7	46	1,2
<b>Opinión sobre inmigración</b>	2,2	11,5	27,2	57,4	1,7

- a) Esta tabla no es una tabla de doble entrada bidimensional, sino de cuatro variables (primera columna) cualitativas. Comenta estos resultados y calcula las frecuencias absolutas de estas cuatro variables.
- b) Construye un diagrama de barras comparativo.
- a) La tabla de frecuencias absolutas de estos resultados es la siguiente:

	Mucho	Bastante	Poco	Nada	Ns/Nc	Total
Opiniones políticas	47	212	647	1558	25	2489
Hábitos de consumo	77	378	747	1257	30	2489
Hábitos de salud	114	535	665	1145	30	2489
Opinión sobre inmigración	55	286	677	1429	42	2489

- b) Un diagrama de barras comparativo de frecuencias absolutas de las respuestas a las cuatro preguntas es:



44. **Estudio médico.** En una investigación sobre la relación entre la hipertensión y el hábito de fumar se tomaron estos datos de 187 personas:

- a) Calcula la distribución conjunta de frecuencias relativas y analízala en términos de porcentaje.
- b) Halla las distribuciones marginales de frecuencias relativas.

	Con hipertensión	Sin hipertensión
<b>No fumador</b>	20	49
<b>Fumador moderado</b>	36	24
<b>Fumador empedemido</b>	42	16

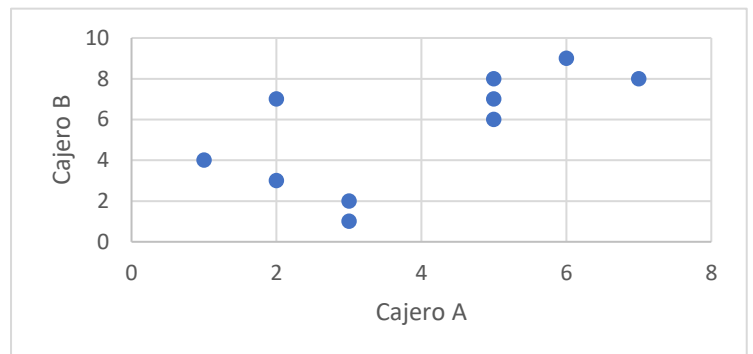
- a) La muestra de 187 personas, el 63,1 % son fumadores y el 52,41 % padecen hipertensión. De las personas con hipertensión casi el 80 % son fumadores, siendo este porcentaje del 45% entre los que no la padecen.
- b) Las distribuciones marginales de frecuencias relativas de ambas variables las obtenemos al sumar las filas y las columnas. Así:

	Fr, relativas		Con hipertensión	Sin hipertensión	Total
No fumador	0,3690				
Fumadores moderados	0,3208				
Fumadores empedernidos	0,3102	Fr. relativas	0,5241	0,4759	1
Total	1				

45. **Cajeros automáticos.** El número de clientes que sacan dinero de dos cajeros automáticos, A y B, en un centro comercial, de 19 a 20 horas, durante diez días laborables consecutivos ha sido:

Cajero A	2	3	1	6	5	2	5	3	7	5
Cajero B	7	1	4	9	6	3	7	2	8	8

- a) Dibuja el diagrama de dispersión y analiza si crees que hay algún tipo de relación entre estos datos.
  - b) Calcula e interpreta el coeficiente de correlación.
- a) Su diagrama de dispersión es:



- b) El coeficiente de correlación es: 0,673

46. **Salario.** El número de horas trabajadas semanalmente ( $x$ ) y el salario bruto semanal ( $y$ ) de un grupo de 40 personas que trabajan el sector hostelero se recoge en la tabla:

Salario ( $y$ ) (€)	[700, 1100)	[1100, 1500)	[1500, 1900)	[1900, 2300)
Horas ( $x$ )				
[31, 35)	5	4	2	1
[35, 37)	1	2	4	6
[37, 41)	0	3	4	8

- a) Halla el número medio de horas trabajadas semanalmente. ¿Cuánto cobra la mayoría de las personas mejor pagadas?
- b) Calcula el número de horas trabajadas por las personas que tienen un salario al de 1500 euros.

- a) La mayoría (el 37,5 %) cobra un salario entre 1900 y 2300 euros.
- b) El número de horas trabajadas por las personas que tienen un salario al menos de 1500 euros es la distribución condicionada ( $x/y \geq 1500$ ). Así:

$(x/y \geq 1500)$	Fr. absoluta
[31, 35)	3
[35, 37)	10
[37, 41)	12
Total	25

47. **Salarios.** Una empresa emergente (*startup*) de gran éxito tiene 19 trabajadores, incluyendo a su dueña, que tiene asignado un salario bruto anual de 650.000 euros. Los salarios del resto de empleados son:

80 000	85 000	123 000	143 000	83 000	147 000
84 000	184 000	163 000	195 000	142 000	81 000
92 000	157 000	153 000	285 000	315 000	99 000

- a) ¿Cuál es el salario medio de todos los empleados, incluida la dueña? ¿Es representativo?
- b) Calcula la desviación típica.
- c) Halla la mediana y los cuartiles e interprétalas. ¿Cuál es el recorrido intercuartílico?
- d) Debido a los buenos resultados del primer año de funcionamiento de la empresa, la dueña decide primar con 30000 € a todo el personal, incluida ella misma. ¿Cómo se modifican la media y la desviación típica con este incremento?
- a) El salario medio de todos los empleados, incluida la dueña, es: 171632 €
- b) La desviación típica es:  $s = 129679$  €.
- c) El 50 % de los empleados tiene un salario inferior o igual 143000 €. El 25 % de los empleados con los salarios más bajos ganan como mucho 85000 €. El 25 % de los empleados con los salarios más altos ganan al menos 184000 €.
- d) Si se incrementan todos los salarios en 30000 €, el salario medio se vería incrementado en dicha cantidad puesto que cada salario aumentaría en 30000 €. Sin embargo, la desviación típica permanece invariable puesto que mide la dispersión de los salarios con respecto a la media y seguiría siendo la misma.

48. **Antigüedad salarial.** Los salarios brutos mensuales ( $y$ ), en euros, y la antigüedad ( $x$ ), en años, de cuatro empleados de una empresa son:

<b>Antigüedad (<math>x</math>)</b>	0	1	3	4
<b>Salario (<math>y</math>)</b>	1000	1200	1500	1300

- a) Determina la recta de regresión que permita hallar el salario en función de la antigüedad.  
 b) Cuando recogieron los datos, había un empleado enfermo que tiene una antigüedad de 2 años, ¿qué salario puedes estimar que tiene?

a) La recta de regresión de  $y$  sobre  $x$ :  $y = 90x + 1070$

b) Se estima un salario de 1250 € para este empleado.

49. **Puntos de luz e intensidad luminosa.** Al analizar la relación entre el número de puntos de luz en diversas zonas de un edificio ( $x$ ) y su intensidad luminosa ( $y$ ), medida en lux, las medidas de cinco lugares han sido:

<b>Puntos de luz (<math>x</math>)</b>	1	2	3	4	5
<b>Intensidad luminosa (<math>y</math>)</b>	100	120	110	150	130

- a) Compara la variabilidad de ambas variables.  
 b) Realiza una regresión lineal para explicar la intensidad luminosa a partir de los puntos de luz. ¿Qué intensidad se esperaría en un lugar con 7 puntos de luz?  
 c) Calcula el coeficiente de correlación lineal y determina la bondad del ajuste. Interpreta los resultados.
- a) El número de puntos de luz es más disperso que la intensidad luminosa porque tiene un coeficiente de variación mayor.  
 b) Para siete puntos de luz, se estima una intensidad luminosa 158 de luz  
 c) El 54,76 % de la intensidad luminosa viene explicado mediante la recta de regresión por los puntos de luz y el resto por otras causas.

50. **Competición escolar.** En una prueba deportiva de mi instituto, estudiantes de tercero y cuarto de la ESO y primero de Bachillerato, han obtenido una puntuación media de 12, 15 y 10, respectivamente. Si participaron el doble de estudiantes de tercero que de cuarto, y el doble de cuarto que de primero de Bachillerato, ¿cuál es la puntuación media de las personas participantes?

La puntuación media de todos los participantes es 12,57.

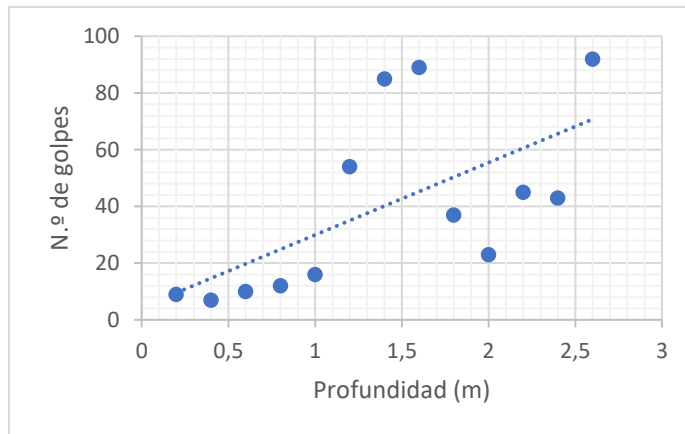
51. **Rentabilidad de las instalaciones renovables.** Una empresa de instalaciones renovables está analizando la rentabilidad y para ello mide la generación solar ( $x$ ) y eólica ( $y$ ) producida durante ocho días en una instalación, obteniendo los siguientes resultados (en  $kWh$ ):

$x$	54,6	28,4	45,3	49,8	34,2	41,6	64,1	33,8
$y$	16,2	52,3	45,3	18	39,1	29,9	13,5	43,1

- a) Estudia la correlación lineal de la generación eólica respecto a la generación solar.  
 b) Estima el valor previsto de generación eólica para una generación solar de 40  $kWh$ . ¿Es fiable esta estimación?
- a) Existe una correlación es negativa y fuerte y, por tanto, indica que al aumentar los  $kWh$  de energía solar disminuye los  $kWh$  de energía eólica en estas instalaciones renovables.
- b) La estimación sí es fiable porque 40  $kWh$  se encuentra dentro del rango de valores  $y$ , además, el coeficiente de determinación  $R^2 = 0,7744$  es bastante elevado.
52. **Características geotécnicas del terreno.** Para determinar las características geotécnicas de un terreno en el que se va a construir una vivienda unifamiliar, se ha realizado un ensayo dinámico con el fin de averiguar el número de golpes necesarios para que el penetrómetro se introduzca a una determinada profundidad. Los resultados han sido:
- a) Dibuja el diagrama de dispersión y coméntalo.  
 b) Calcula los coeficientes de correlación y determinación e interprétalos.  
 c) Determina la recta de regresión que permita estimar el número de golpes en función de la profundidad.  
 d) Explica si crees que es fiable estimar el número de golpes a tres metros de profundidad en este terreno.

Acta de ensayo de penetración dinámica continua	
Profundidad (m)	N.º de golpes
0,2	9
0,4	7
0,6	10
0,8	12
1,0	16
1,2	54
1,4	85
1,6	89
1,8	37
2,0	23
2,2	45
2,4	43
2,6	92

a) El diagrama de dispersión es:



b) El 40 % del número de golpes obtenidos por el penetrómetro viene explicado a través de la recta de regresión por los metros de profundidad, pero el 60% restante se deberá a otras causas.

c) La recta de regresión que permita estimar el número de golpes ( $y$ ) en función de los metros de profundidad ( $x$ ) es:  $y = 25,49x + 4,46$

d) La estimación no es fiable por un doble motivo: primero, porque 3 metros está fuera del rango valores (aunque próximo) y, segundo, debido a que el coeficiente de determinación es bajo.

53. **Equipo de baloncesto.** En una prueba para seleccionar jugadoras de baloncesto se ha medido la altura de 110 jóvenes, obteniéndose los siguientes resultados:

Altura (cm)	N.º de jugadoras
[175,180)	10
[180,185)	22
[185,190)	32
[190,195)	25
[195,200)	12
[200,205)	9
<b>Total</b>	<b>110</b>



Se cree que la mejor distribución es, de menor a mayor altura: un 20 % de las jugadoras en el puesto de base, un 50 % de aleros y un 30 % de pívots.

a) ¿Cuál es la altura máxima para jugar de base? ¿Y la altura mínima para jugar de pívot?

b) Una jugadora que tenga la altura en el intervalo más frecuente, ¿en qué puesto jugará?

- a) La altura máxima para jugar de base es de 183 cm. La altura mínima para jugar de pívot es de 193 cm.  
 b) Una jugadora que tenga una altura en el intervalo más frecuente jugará de alero.

### ➤ Un mundo matemático

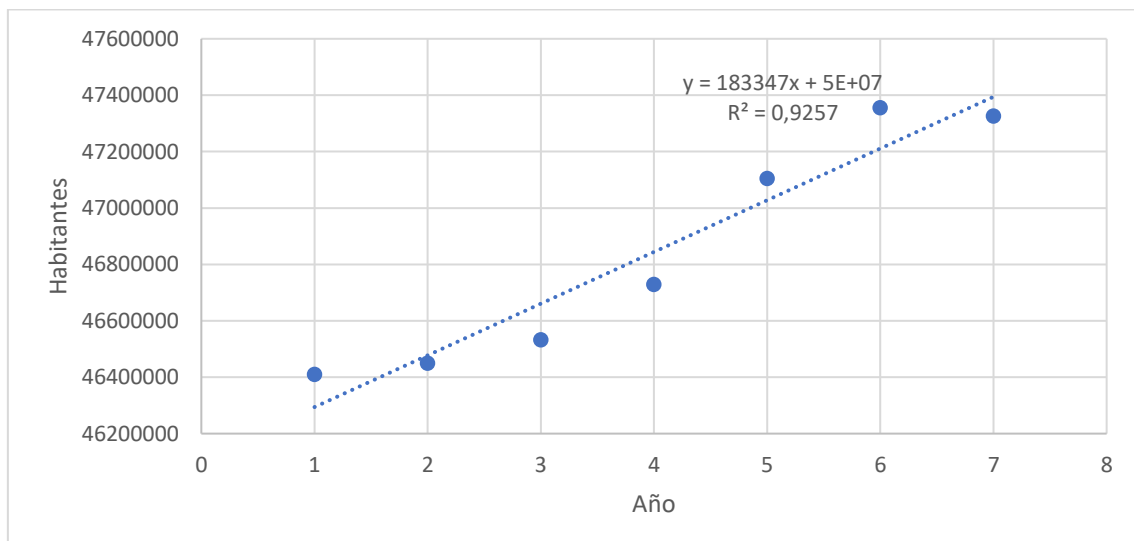
1. Construye una tabla con dos columnas a partir de la información obtenida.

Según el INE en el segundo trimestre de cada año la población española era la siguiente:

Año	Población total
2015	46410149
2016	46449874
2017	46532869
2018	46728814
2019	47105358
2020	47355685
2021	47326687

2. Dibuja un diagrama de dispersión de los datos, utilizando Excel, GeoGebra u otra hoja de cálculo o una calculadora gráfica. Para ello, considera como variable independiente (x), haciendo  $x = 1$  el año 2015,  $x = 2$  el año 2016 y así sucesivamente; para la población, considera como variable dependiente (y), en millones de habitantes.

Utilizando Excel obtenemos el siguiente diagrama de dispersión, la recta de regresión y el coeficiente de determinación.





3. Ajusta la recta de regresión a estos datos, obtén los coeficientes de correlación y de determinación e interprétalos. ¿Consideras satisfactorio este ajuste?

Por otra parte, el coeficiente de correlación lineal es:

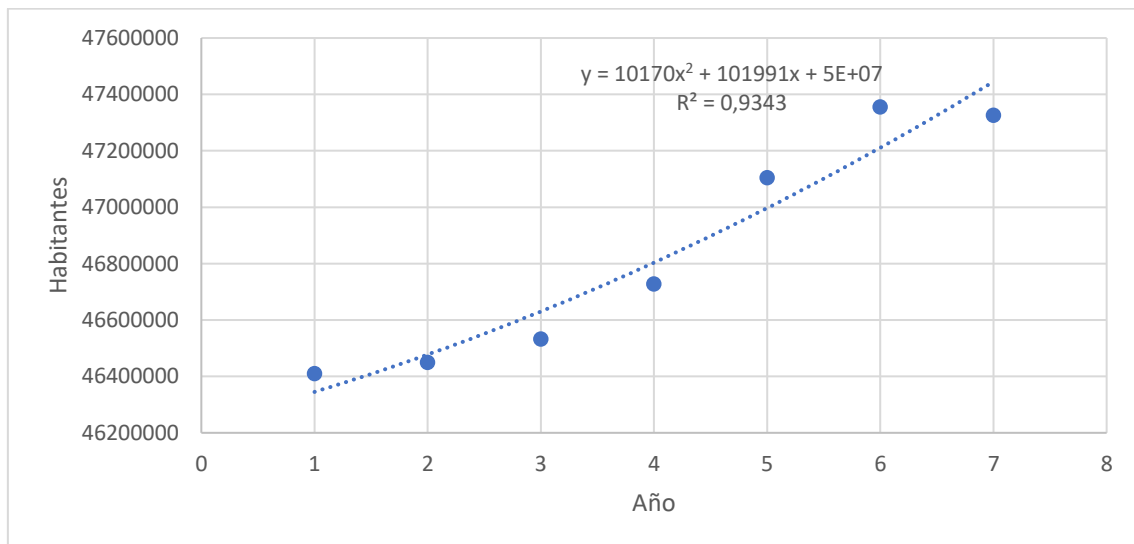
$$r = 0,9621$$

Por tanto, la correlación es positiva y fuerte, es decir, cada año se incrementa la población total española y, además, lo hace de forma lineal.

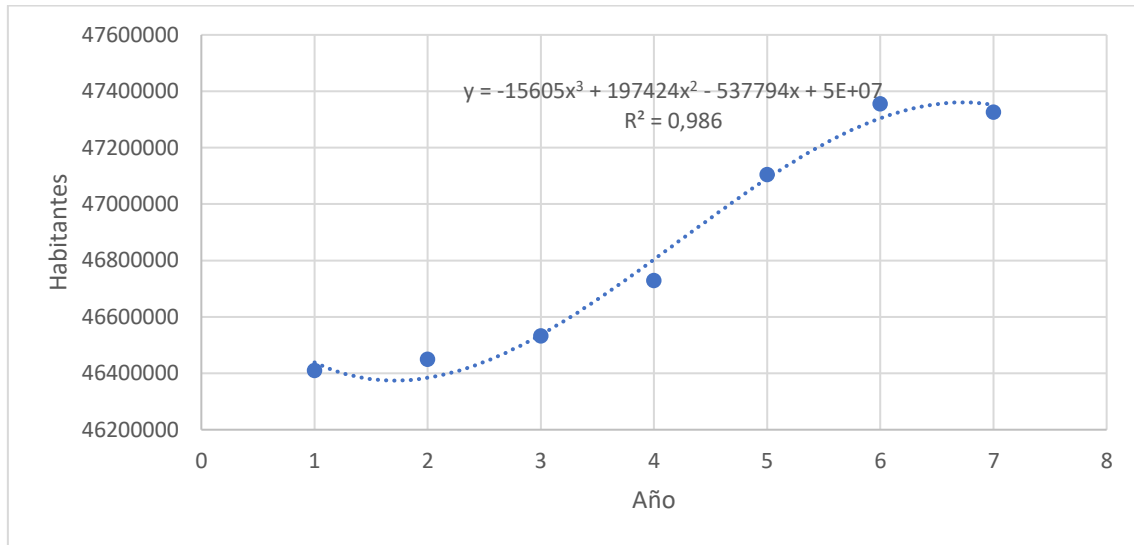
Como el coeficiente de determinación es  $R^2 = 0,9257$  significa que el 92,57 % de los datos de la población total son explicados por la relación lineal del año de estudio.

4. Analiza si un ajuste mediante una función polinómica de orden 2 o 3 sería más apropiado que el ajuste lineal obtenido en 3.

Los resultados del ajuste de un polinomio de grado 2 obtenidos con Excel son:



Mientras que los facilitados para un polinomio de grado tres son los siguientes:



Observamos que el ajuste mediante una función polinómica de orden 3 es el que mejor se ajusta a los datos puesto que tiene un coeficiente de determinación mayor. No obstante, es pequeña la diferencia obtenida entre los polinomios de orden 1, 2 y 3.

5. Utiliza la recta de regresión para estimar el número de habitantes en España al finalizar el año en curso. ¿Crees que es fiable esta estimación?

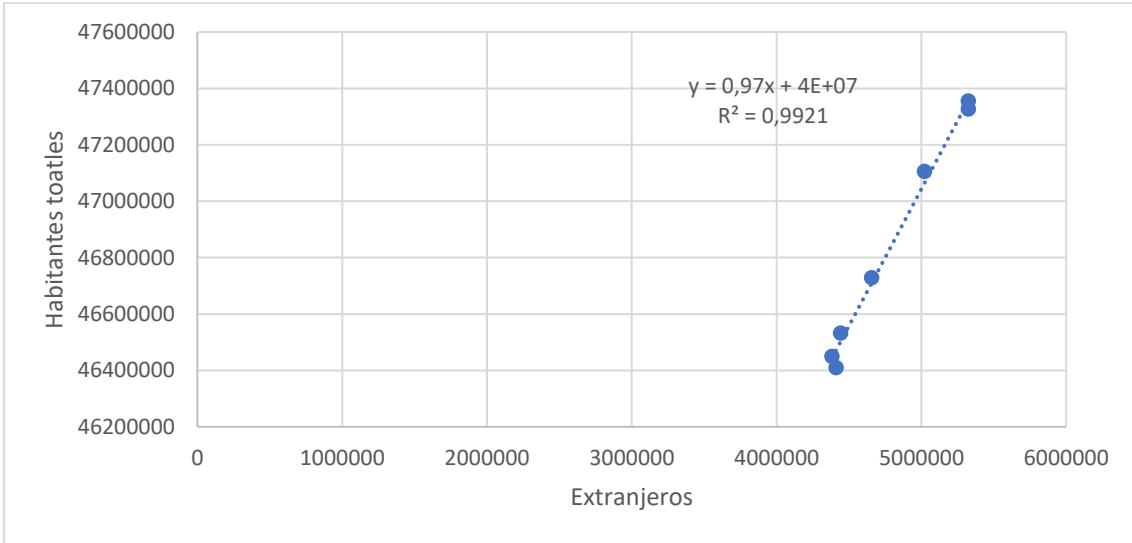
Se estima que en el segundo semestre del año 2022 haya una población total de 51466776.

6. Consulta de nuevo los datos del INE y construye una tabla con la población total como variable dependiente (y) y la población de extranjeros (x) también durante el mismo periodo de años. A continuación, dibuja el diagrama de dispersión correspondiente.

La tabla con los datos de población extranjera (x) y población total (y) es la siguiente:

Año	Población extranjera(x)	Población total (y)
2015	4522850	46410149
2016	4382618	46449874
2017	4443780	46532869
2018	4656333	46728814
2019	5021334	47105358
2020	5327965	47355685
2021	5375917	47326687

Utilizando Excel obtenemos el diagrama de dispersión, los dos ajustes mediante una función lineal y cuadrática, así como los coeficientes de determinación de ambas.



7. Ajusta a esta nube de puntos una recta y una parábola. Utiliza la más adecuada para estimar la población total para el próximo año si hubiera 6 millones de residentes extranjeros.

