BLOQUE IV · ESTADÍSTICA Y PROBABILIDAD

Unidad 14

Estadística bidimensional

Repaso de media y desviación típica · nube de puntos · covarianza · coeficiente de correlación · recta de regresión · estimaciones y su fiabilidad · siete ejercicios resueltos, quince propuestos y un ejercicio de síntesis



Por qué esta unidad. En ciencias sociales casi nada depende de una sola cosa, y casi ninguna relación es exacta. Un piso más grande suele ser más caro, y aun así dos pisos iguales se alquilan a precios distintos. La estadística bidimensional mide cuánto se parecen dos variables a una relación lineal y aprovecha ese parecido para hacer estimaciones, diciendo además cuánto se puede confiar en ellas.

Qué vas a saber hacer al terminar

Apartado	Ejercicio tipo
1. Repaso	calcular la media y la desviación típica de una variable
2. Nube de puntos	reconocer el tipo de relación entre dos variables
3. Covarianza	calcularla y leer su signo
4. Correlación	hallar el coeficiente r e interpretarlo
5. Recta de regresión	escribir la recta de Y sobre X
6. Estimación	predecir un valor y valorar si la predicción es fiable

1. Teoría

1.1 · Repaso: media y desviación típica

PARA MEMORIZAR

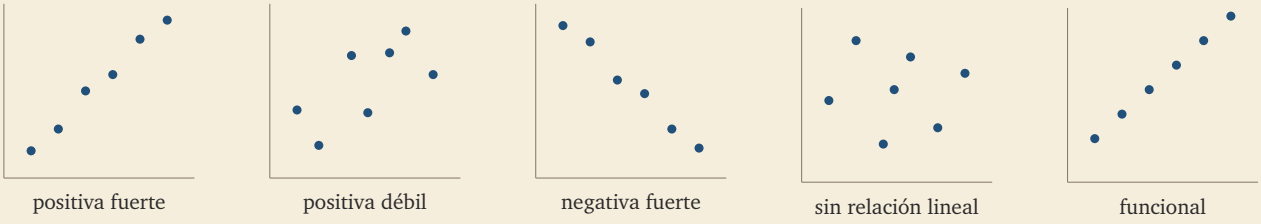
Para una variable con n datos $x_1, x_2, \ldots, x_n$:

$$\text{media: } \bar{x} = \frac{\sum x_i}{n} \qquad \text{varianza: } \sigma_x^2 = \frac{\sum x_i^2}{n} - \bar{x}^2 \qquad \text{desviación típica: } \sigma_x = \sqrt{\sigma_x^2}$$

La media dice dónde está el centro de los datos. La desviación típica dice cuánto se separan de él.

1.2 · Variables bidimensionales y nube de puntos

Una **variable bidimensional** $(X,\ Y)$ recoge dos datos de cada individuo: la superficie y el alquiler de un piso, el precio y las unidades vendidas de un producto. Cada individuo es un punto $(x_i,\ y_i)$, y el conjunto de todos es la **nube de puntos** o diagrama de dispersión.



La relación es **positiva** o directa si al aumentar X tiende a aumentar Y, y **negativa** o inversa si tiende a disminuir. Es más **fuerte** cuanto más se aprietan los puntos alrededor de una recta. Si todos caen sobre ella, la relación es funcional.

1.3 · La covarianza

PARA MEMORIZAR

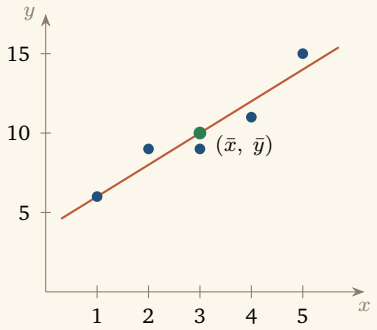
Covarianza de X e Y:

$$\sigma_{xy} = \frac{\sum x_i\, y_i}{n} - \bar{x} \cdot \bar{y}$$

Su **signo** indica el sentido de la relación: positiva si $\sigma_{xy} > 0$, negativa si $\sigma_{xy} < 0$. Su tamaño depende de las unidades, así que por sí sola no dice si la relación es fuerte.

Ejemplo. Cinco pares de datos, con las columnas que hacen falta para todos los cálculos.

x_i	y_i	x_i^2	y_i^2	$x_i\, y_i$
1	6	1	36	6
2	9	4	81	18
3	9	9	81	27
4	11	16	121	44
5	15	25	225	75
15	**50**	**55**	**544**	**170**



$\bar{x} = \dfrac{15}{5} = 3$, $\bar{y} = \dfrac{50}{5} = 10$, $\sigma_{xy} = \dfrac{170}{5} - 3 \cdot 10 = 34 - 30 = 4$. Positiva: relación directa.

1.4 · El coeficiente de correlación lineal

PARA MEMORIZAR

Coeficiente de correlación lineal de Pearson:

$$r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} \qquad \text{siempre } -1 \le r \le 1$$

No tiene unidades. Su signo es el de la covarianza, y su valor absoluto mide la fuerza de la relación lineal.

$\lvert r\rvert$ cerca de 1	relación lineal fuerte: los puntos se aprietan junto a una recta
$\lvert r\rvert$ intermedio	relación débil: hay tendencia, con mucha dispersión
r cerca de 0	sin relación lineal
$r = 1$ o $r = -1$	relación funcional: todos los puntos están sobre una recta

Ejemplo. Con los datos anteriores: $\sigma_x^2 = \dfrac{55}{5} - 3^2 = 2$, $\sigma_x = 1{,}414$; $\sigma_y^2 = \dfrac{544}{5} - 10^2 = 8{,}8$, $\sigma_y = 2{,}966$.

$r = \dfrac{4}{1{,}414 \cdot 2{,}966} = 0{,}954$. Relación lineal positiva y fuerte.



AVISO · CORRELACIÓN NO SIGNIFICA CAUSA

Un coeficiente alto dice que dos variables se mueven juntas. No dice que una provoque la otra. Puede existir una tercera variable que mueva a las dos, o tratarse de una coincidencia. Para hablar de causa hace falta algo más que un número.

1.5 · La recta de regresión

PARA MEMORIZAR

Recta de regresión de Y sobre X: la recta que mejor se ajusta a la nube para estimar Y a partir de X.

$$y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x})$$

Pasa por el punto $(\bar{x},\ \bar{y})$ y su pendiente es $\dfrac{\sigma_{xy}}{\sigma_x^2}$: la covarianza entre la **varianza** de X.

Ejemplo. Pendiente: $\dfrac{4}{2} = 2$. Recta: $y - 10 = 2(x - 3) \Rightarrow y = 2x + 4$. Por cada unidad que aumenta X, la recta estima 2 unidades más de Y.



TRUCO · ES UNA RECTA PUNTO-PENDIENTE

La fórmula es la misma de la recta tangente de la unidad 12, $y - y_0 = m(x - x_0)$, con otro punto y otra pendiente. El punto es el de las medias. La pendiente sale de la covarianza.

1.6 · Estimaciones y fiabilidad

PARA MEMORIZAR

Para **estimar** Y con un valor de X, se sustituye en la recta. La estimación es más fiable cuando se cumplen dos condiciones:

- $|r|$ está cerca de 1.
- El valor de x está dentro del intervalo de los datos, o muy cerca. Fuera de él se está **extrapolando**, y nada garantiza que la relación se mantenga.

Ejemplo. Para $x = 3{,}5$: $y = 2 \cdot 3{,}5 + 4 = 11$. Es fiable: $r = 0{,}954$ y $3{,}5$ está entre 1 y 5.



CONEXIÓN · LA HERRAMIENTA DE LA ECONOMÍA APLICADA

Cuando un informe dice que cada año adicional de estudios se asocia con un salario más alto, detrás hay una recta de regresión. La econometría que se estudia en la universidad empieza exactamente aquí, con la pendiente $\frac{\sigma_{xy}}{\sigma_x^2}$.

2. Ejercicios resueltos

Siete ejercicios tipo, de menos a más. Cada solución sigue el mismo orden: planteamiento, desarrollo, resultado y conclusión. Los resultados se dan con tres decimales.

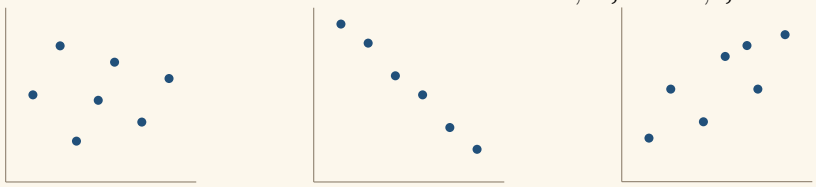
RESUELTO 1 · Repaso: media y desviación típica

Las notas de un alumno en cinco exámenes son 4, 6, 7, 8 y 5. Calcula la media y la desviación típica.

Desarrollo. $\bar{x} = \dfrac{4+6+7+8+5}{5} = \dfrac{30}{5} = 6$. $\quad \sum x_i^2 = 16+36+49+64+25 = 190$.
$\sigma_x^2 = \dfrac{190}{5} - 6^2 = 38 - 36 = 2$. $\quad \sigma_x = \sqrt{2} = 1{,}414$.
Conclusión. La nota media es **6** y la desviación típica, **1,414** puntos.

RESUELTO 2 · Leer una nube de puntos

Asocia cada nube con su coeficiente de correlación: $r = -0{,}95$, $r = 0{,}6$, $r = 0{,}05$.



Planteamiento. El signo de r dice si la nube sube o baja. El valor absoluto, si los puntos están apretados.
Desarrollo. B baja y los puntos están casi en línea: $r = -0{,}95$. C sube, con dispersión: $r = 0{,}6$. A no muestra tendencia: $r = 0{,}05$.
Conclusión. A, $\mathbf{0{,}05}$; B, $\mathbf{-0{,}95}$; C, $\mathbf{0{,}6}$.

RESUELTO 3 · Covarianza

Una papelería prueba cinco precios para un cuaderno y anota las unidades vendidas cada semana.

Precio X (€)	1	2	3	4	5
Unidades Y	15	14	12	12	7

Calcula la covarianza e interpreta su signo.

Desarrollo. $\bar{x} = \dfrac{15}{5} = 3$. $\quad \bar{y} = \dfrac{60}{5} = 12$.
$\sum x_i\, y_i = 15 + 28 + 36 + 48 + 35 = 162$. $\quad \sigma_{xy} = \dfrac{162}{5} - 3 \cdot 12 = 32{,}4 - 36 = -3{,}6$.
Conclusión. $\sigma_{xy} = \mathbf{-3{,}6}$. Es negativa: cuando el precio sube, las unidades vendidas tienden a bajar.

RESUELTO 4 · Coeficiente de correlación

Con los datos del ejercicio anterior, calcula el coeficiente de correlación lineal e interprétalo.

Planteamiento. Hacen falta las dos desviaciones típicas y la covarianza, que ya se conoce.
Desarrollo. $\sum x_i^2 = 55$. $\sigma_x^2 = \dfrac{55}{5} - 3^2 = 2$. $\sigma_x = 1{,}414$.
$\sum y_i^2 = 225 + 196 + 144 + 144 + 49 = 758$. $\sigma_y^2 = \dfrac{758}{5} - 12^2 = 151{,}6 - 144 = 7{,}6$. $\sigma_y = 2{,}757$.
$r = \dfrac{-3{,}6}{1{,}414 \cdot 2{,}757} = \dfrac{-3{,}6}{3{,}898} = -0{,}923$.
Conclusión. $r = \mathbf{-0{,}923}$. La relación lineal es negativa y fuerte: el precio explica bien las ventas.

RESUELTO 5 · Recta de regresión

Con los mismos datos, halla la recta de regresión de Y sobre X e interpreta su pendiente.

Desarrollo. Pendiente: $\dfrac{\sigma_{xy}}{\sigma_x^2} = \dfrac{-3{,}6}{2} = -1{,}8$. Punto: $(\bar{x},\ \bar{y}) = (3,\ 12)$.
$y - 12 = -1{,}8\,(x - 3) \Rightarrow y = -1{,}8x + 5{,}4 + 12 \Rightarrow y = -1{,}8x + 17{,}4$.
Comprobación. La recta pasa por el punto de las medias: $-1{,}8 \cdot 3 + 17{,}4 = 12$.
Conclusión. $\mathbf{y = -1{,}8x + 17{,}4}$. Por cada euro que sube el precio, la recta estima 1,8 cuadernos menos a la semana.

RESUELTO 6 · Estimar y valorar

Con la recta anterior, estima las unidades que se venderían a 2,50 € y a 10 €. Valora las dos estimaciones.

Desarrollo. $x = 2{,}5$: $y = -1{,}8 \cdot 2{,}5 + 17{,}4 = 12{,}9$. $x = 10$: $y = -1{,}8 \cdot 10 + 17{,}4 = -0{,}6$.
Conclusión. A 2,50 € se venderían unos **13 cuadernos**. La estimación es fiable: $|r| = 0{,}923$ y el precio está entre 1 € y 5 €. A 10 € la recta da un número negativo de cuadernos, que no tiene sentido: el precio queda muy lejos de los datos y la recta deja de servir.

RESUELTO 7 · A partir de las sumas

En una empresa se han anotado los años de antigüedad, X, y el salario mensual en cientos de euros, Y, de 10 trabajadores. Se obtiene: $\sum x_i = 50$, $\sum y_i = 200$, $\sum x_i^2 = 290$, $\sum y_i^2 = 4.250$, $\sum x_i\, y_i = 1.080$.
a) Calcula el coeficiente de correlación. b) Halla la recta de regresión de Y sobre X. c) Estima el salario de un trabajador con 6 años de antigüedad.

Planteamiento. El enunciado ya da las sumas de la tabla. Se calculan medias, varianzas y covarianza, en ese orden.
Desarrollo. a) $\bar{x} = \dfrac{50}{10} = 5$. $\bar{y} = \dfrac{200}{10} = 20$.
$\sigma_x^2 = \dfrac{290}{10} - 5^2 = 4$, $\sigma_x = 2$. $\sigma_y^2 = \dfrac{4.250}{10} - 20^2 = 25$, $\sigma_y = 5$. $\sigma_{xy} = \dfrac{1.080}{10} - 5 \cdot 20 = 8$.

$r = \dfrac{8}{2 \cdot 5} = 0{,}8$.
b) Pendiente: $\dfrac{8}{4} = 2$. $y - 20 = 2\,(x - 5) \Rightarrow y = 2x + 10$.
c) $y = 2 \cdot 6 + 10 = 22$.
Resultado. $r = 0{,}8$; $y = 2x + 10$; $y(6) = 22$.
Conclusión. La relación es positiva y bastante fuerte. Cada año de antigüedad se asocia con 200 € más de salario, y a un trabajador con 6 años le corresponde un salario estimado de **2.200 €**.

Errores que hacen perder puntos en esta unidad

- Olvidar restar el cuadrado de la media en la varianza, o el producto de las medias en la covarianza.
- Dividir la covarianza entre las varianzas para hallar r. En r van las desviaciones típicas.
- Dividir la covarianza entre σ_x para hallar la pendiente. En la pendiente va la varianza, σ_x^2.
- Dar un coeficiente r mayor que 1 o menor que -1 sin revisar los cálculos.
- Redondear demasiado pronto: las desviaciones típicas se llevan con tres decimales.
- Estimar sin decir nada sobre la fiabilidad.
- Afirmar que una variable causa la otra porque r es alto.

3. Ejercicios propuestos

Soluciones en el apartado 5, con la resolución completa de los ejercicios 13 y 14.

NIVEL 1 · PRACTICAR

1. Calcula la media y la desviación típica de los datos 2, 4, 4, 6 y 9.

2. Indica si esperas una correlación positiva, negativa o casi nula entre: a) la renta de una familia y su gasto en ocio; b) el precio de un producto y las unidades vendidas; c) el número de pie de una persona adulta y su salario.

3. De cinco pares de datos se sabe que $\sum x_i\, y_i = 260$, $\bar{x} = 4$ e $\bar{y} = 12$. Calcula la covarianza.

4. Calcula r sabiendo que $\sigma_{xy} = 6$, $\sigma_x = 2$ y $\sigma_y = 4$.

5. Escribe la recta de regresión de Y sobre X sabiendo que $\bar{x} = 4$, $\bar{y} = 10$, $\sigma_{xy} = 3$ y $\sigma_x^2 = 2$.

NIVEL 2 · APLICAR

6. Calcula la covarianza de estos datos:

X	1	2	3	4	5
Y	5	6	10	8	11

7. Calcula el coeficiente de correlación de los datos del ejercicio 6 e interprétalo.

8. Halla la recta de regresión de Y sobre X con los datos del ejercicio 6.

9. Con la recta del ejercicio 8, estima el valor de Y para $x = 6$ y comenta su fiabilidad.

10. La recta de regresión de Y sobre X de una distribución es $y = 0{,}8x + 12$, y la media de X es 10. ¿Cuál es la media de Y?

NIVEL 3 · PENSAR

11. Entre el precio de un artículo y sus ventas se ha obtenido $r = -0{,}92$. Razona si estas afirmaciones son verdaderas o falsas: a) la relación es débil porque r es negativo; b) la pendiente de la recta de regresión es negativa; c) las estimaciones hechas con la recta dentro del rango de precios son fiables.

12. De 8 pares de datos se sabe: $\sum x_i = 40$, $\sum y_i = 120$, $\sum x_i^2 = 232$, $\sum y_i^2 = 1.928$, $\sum x_i\, y_i = 648$. Calcula r y la recta de regresión de Y sobre X.

13. Una comunidad de vecinos anota durante cinco semanas la temperatura media exterior, X, en grados, y el consumo de gas de la calefacción, Y, en m^3.

Temperatura X	7	9	10	11	13
Consumo Y	46	40	38	42	34

a) Calcula el coeficiente de correlación e interprétalo. b) Halla la recta de regresión de Y sobre X. c) Estima el consumo de una semana con 12 grados de media.

14. De una distribución bidimensional se conoce la recta de regresión de Y sobre X, $y = 2x + 1$, y además $\bar{x} = 4$, $\sigma_x = 1{,}5$ y $\sigma_y = 4$. Halla la media de Y, la covarianza y el coeficiente de correlación.

15. En los municipios de una provincia, el número de farmacias y el número de bares tienen un coeficiente de correlación de $0{,}94$. ¿Se puede concluir que abrir farmacias hace que abran más bares? Explica qué puede estar ocurriendo.

4. Ejercicio de síntesis

SÍNTESIS · Alquileres en el barrio

Una agencia inmobiliaria anota la superficie, X, en m^2, y el alquiler mensual, Y, en euros, de cinco pisos de un mismo barrio.

Superficie X (m^2)	50	60	70	80	90
Alquiler Y (€)	700	800	1.000	950	1.050

a) Calcula las medias y las desviaciones típicas de las dos variables.
b) Calcula la covarianza y el coeficiente de correlación. Interprétalo.
c) Halla la recta de regresión de Y sobre X. ¿Qué significa su pendiente?
d) Estima el alquiler de un piso de 75 m^2 y el de uno de 150 m^2. ¿Cuál de las dos estimaciones es más fiable?
e) Un piso de 70 m^2 se anuncia por 1.150 €. ¿Qué le dirías al cliente?

5. Soluciones

Ejercicios propuestos

1. $\bar{x} = \dfrac{25}{5} = 5$; $\sigma_x^2 = \dfrac{153}{5} - 25 = 5{,}6$; $\sigma_x = 2{,}366$.

2. a) Positiva. b) Negativa. c) Casi nula.

3. $\sigma_{xy} = \dfrac{260}{5} - 4 \cdot 12 = 52 - 48 = 4$.

4. $r = \dfrac{6}{2 \cdot 4} = 0{,}75$.

5. Pendiente $\dfrac{3}{2} = 1{,}5$. $y - 10 = 1{,}5\,(x - 4) \Rightarrow y = 1{,}5x + 4$.

6. $\bar{x} = 3$, $\bar{y} = 8$, $\sum x_i\, y_i = 5 + 12 + 30 + 32 + 55 = 134$. $\sigma_{xy} = 26{,}8 - 24 = 2{,}8$.

7. $\sigma_x^2 = 2$, $\sigma_x = 1{,}414$. $\sum y_i^2 = 346$, $\sigma_y^2 = 69{,}2 - 64 = 5{,}2$, $\sigma_y = 2{,}280$. $r = \dfrac{2{,}8}{1{,}414 \cdot 2{,}280} = 0{,}868$. Relación positiva y bastante fuerte.

8. Pendiente $\dfrac{2{,}8}{2} = 1{,}4$. $y - 8 = 1{,}4\,(x - 3) \Rightarrow y = 1{,}4x + 3{,}8$.

9. $y = 1{,}4 \cdot 6 + 3{,}8 = 12{,}2$. El valor $x = 6$ queda fuera de los datos, aunque muy cerca, y $r = 0{,}868$: la

estimación es aceptable, con cierta cautela.
10. La recta pasa por $(\bar{x},\ \bar{y})$: $\bar{y} = 0{,}8 \cdot 10 + 12 = 20$.
11. a) Falsa: la fuerza la da el valor absoluto, $0{,}92$, que es alto. El signo solo indica el sentido. b) Verdadera: la pendiente tiene el signo de la covarianza, el mismo que r. c) Verdadera: $|r|$ está cerca de 1 y no se extrapola.
12. $\bar{x} = 5$, $\bar{y} = 15$. $\sigma_x^2 = 29 - 25 = 4$, $\sigma_x = 2$. $\sigma_y^2 = 241 - 225 = 16$, $\sigma_y = 4$. $\sigma_{xy} = 81 - 75 = 6$. $r = \dfrac{6}{8} = 0{,}75$. Pendiente $\dfrac{6}{4} = 1{,}5$: $y - 15 = 1{,}5\,(x - 5) \Rightarrow y = 1{,}5x + 7{,}5$.
13. $r = -0{,}85$; $y = -1{,}7x + 57$; 36,6 m^3 (resolución completa a continuación).
14. $\bar{y} = 9$; $\sigma_{xy} = 4{,}5$; $r = 0{,}75$ (resolución completa a continuación).
15. No se puede concluir. Las dos variables crecen con una tercera, la población del municipio: donde vive más gente hay más farmacias y también más bares. La correlación es alta sin que una cosa cause la otra.

Resolución completa del ejercicio 13

Planteamiento. Se monta la tabla con las columnas x_i^2, y_i^2 y $x_i\,y_i$, y se suman.

x_i	y_i	x_i^2	y_i^2	$x_i\,y_i$
7	46	49	2.116	322
9	40	81	1.600	360
10	38	100	1.444	380
11	42	121	1.764	462
13	34	169	1.156	442
50	**200**	**520**	**8.080**	**1.966**

Desarrollo. a) $\bar{x} = 10$, $\bar{y} = 40$. $\sigma_x^2 = \dfrac{520}{5} - 100 = 4$, $\sigma_x = 2$. $\sigma_y^2 = \dfrac{8.080}{5} - 1.600 = 16$, $\sigma_y = 4$.
$\sigma_{xy} = \dfrac{1.966}{5} - 10 \cdot 40 = 393{,}2 - 400 = -6{,}8$. $r = \dfrac{-6{,}8}{2 \cdot 4} = -0{,}85$.
b) Pendiente: $\dfrac{-6{,}8}{4} = -1{,}7$. $y - 40 = -1{,}7\,(x - 10) \Rightarrow y = -1{,}7x + 57$.
c) $y = -1{,}7 \cdot 12 + 57 = 36{,}6$.
Resultado. $r = -0{,}85$; $y = -1{,}7x + 57$; $y(12) = 36{,}6$.
Conclusión. La relación es negativa y bastante fuerte: cuanto más calor hace fuera, menos gas se gasta. Cada grado más supone 1,7 m^3 menos a la semana. Con 12 grados se estima un consumo de **36,6 m^3**, y la estimación es fiable porque 12 está entre 7 y 13.

Resolución completa del ejercicio 14

Planteamiento. La recta de regresión guarda dos datos: pasa por $(\bar{x},\ \bar{y})$ y su pendiente es $\dfrac{\sigma_{xy}}{\sigma_x^2}$.
Desarrollo. Media de Y: $\bar{y} = 2 \cdot 4 + 1 = 9$.
Covarianza: la pendiente es 2 y $\sigma_x^2 = 1{,}5^2 = 2{,}25$. $\dfrac{\sigma_{xy}}{2{,}25} = 2 \Rightarrow \sigma_{xy} = 4{,}5$.
Correlación: $r = \dfrac{4{,}5}{1{,}5 \cdot 4} = \dfrac{4{,}5}{6} = 0{,}75$.
Resultado. $\bar{y} = 9$; $\sigma_{xy} = 4{,}5$; $r = 0{,}75$.
Conclusión. Con la recta y las dos desviaciones típicas se recupera todo lo demás. La relación es positiva, de fuerza media-alta.

Ejercicio de síntesis

a) $\sum x_i = 350$, $\bar{x} = 70$ m^2. $\sum y_i = 4.500$, $\bar{y} = 900$ €.
$\sum x_i^2 = 25.500$: $\sigma_x^2 = 5.100 - 4.900 = 200$, $\sigma_x = 14{,}14$ m^2.
$\sum y_i^2 = 4.135.000$: $\sigma_y^2 = 827.000 - 810.000 = 17.000$, $\sigma_y = 130{,}38$ €.
b) $\sum x_i\, y_i = 35.000 + 48.000 + 70.000 + 76.000 + 94.500 = 323.500$. $\sigma_{xy} = 64.700 - 63.000 = 1.700$.
$r = \dfrac{1.700}{14{,}14 \cdot 130{,}38} = 0{,}922$. Relación positiva y fuerte: a más superficie, más alquiler, con pocos pisos fuera de la tendencia.
c) Pendiente: $\dfrac{1.700}{200} = 8{,}5$. $y - 900 = 8{,}5\,(x - 70) \Rightarrow y = 8{,}5x + 305$. Cada metro cuadrado más se asocia con 8,50 € más de alquiler al mes.
d) $x = 75$: $y = 637{,}5 + 305 = 942{,}50$ €. $x = 150$: $y = 1.275 + 305 = 1.580$ €.
La primera es fiable: r es alto y 75 m^2 está entre 50 y 90. La segunda es una extrapolación muy lejana: un piso de 150 m^2 no se parece a los cinco de la muestra.
e) La recta estima $\bar{y} = 900$ € para 70 m^2. El piso se anuncia 250 € por encima, casi dos desviaciones típicas del alquiler. Es caro para lo habitual en el barrio, salvo que tenga algo que la superficie no recoge.
Conclusión. La superficie explica bien el alquiler en este barrio: $r = 0{,}922$ y la recta $y = 8{,}5x + 305$ permite estimar con confianza entre 50 y 90 m^2. La recta describe lo habitual, y sirve también para detectar precios que se salen de ello.

6. Compruebo que sé...

- ☐ Calcular la media, la varianza y la desviación típica de una variable.
- ☐ Reconocer en una nube de puntos el sentido y la fuerza de la relación.
- ☐ Montar la tabla de cálculos con las columnas x_i^2, y_i^2 y $x_i\,y_i$.
- ☐ Calcular la covarianza e interpretar su signo.
- ☐ Calcular el coeficiente de correlación y decir qué significa.
- ☐ Escribir la recta de regresión de Y sobre X y explicar su pendiente.
- ☐ Comprobar que la recta pasa por el punto de las medias.
- ☐ Estimar un valor y valorar la fiabilidad con r y con el rango de los datos.
- ☐ Distinguir correlación de causalidad.

7. Hoja de fórmulas

UNIDAD 14 · TODO EN MEDIA CARA

Medias	$\bar{x} = \dfrac{\sum x_i}{n}$ $\quad \bar{y} = \dfrac{\sum y_i}{n}$	
Varianzas	$\sigma_x^2 = \dfrac{\sum x_i^2}{n} - \bar{x}^2$ $\quad \sigma_y^2 = \dfrac{\sum y_i^2}{n} - \bar{y}^2$	desviación típica: la raíz
Covarianza	$\sigma_{xy} = \dfrac{\sum x_i\,y_i}{n} - \bar{x}\cdot\bar{y}$	su signo da el sentido de la relación
Correlación	$r = \dfrac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$ $\quad -1 \le r \le 1$	$\lvert r\rvert$ cerca de 1: relación fuerte
Recta de regresión	$y - \bar{y} = \dfrac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x})$	pasa por $(\bar{x},\ \bar{y})$
Estimación	sustituir x en la recta · fiable si $\lvert r\rvert$ es alto y x está dentro de los datos	