PLANTILLA DE MÉTODO · CORRELACIÓN Y RECTA DE REGRESIÓN

Nombre: _____ Ejercicio: _____ Fecha: _____

La pregunta delante de una nube de puntos: **¿se parecen estos puntos a una recta, y cuánto?** El coeficiente r contesta. Si la respuesta es «mucho», la recta de regresión sirve para estimar.

A · LA TABLA DE CÁLCULOS

x_i	y_i	x_i^2	y_i^2	$x_i\,y_i$

Sumas

1 Número de datos: $n =$ _____

2 **Medias**

$\bar{x} =$ ——— $=$ _____

$\bar{y} =$ ——— $=$ _____

Unidades de X: _____ de Y: _____

B · DISPERSIÓN, COVARIANZA Y CORRELACIÓN

3 **Varianzas y desviaciones típicas**

$\sigma_x^2 = \frac{\text{———}}{n} - \bar{x}^2 =$ _____, $\sigma_x =$ _____ $\sigma_y^2 = \frac{\text{———}}{n} - \bar{y}^2 =$ _____, $\sigma_y =$ _____

4 **Covarianza** $\sigma_{xy} = \frac{\text{———}}{n} - \bar{x}\cdot\bar{y} =$ _____ $-$ _____ $=$ _____ Signo: ☐ $+$, directa ☐ $-$, inversa

5 **Correlación** $r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \frac{\text{———}}{\cdot} =$ _____ Interpretación: _____

C · RECTA Y ESTIMACIÓN

6 **Pendiente** $m = \frac{\sigma_{xy}}{\sigma_x^2} =$ ——— $=$ _____

7 **Recta de Y sobre X** $y -$ _____ $=$ _____ $\cdot (x -$ _____$)$ $\Rightarrow$ $y =$ _____ $x +$ _____

8 **Estimación** para $x =$ _____: $y =$ _____ $=$ _____

9 **¿Fiable?** $|r| =$ _____ ☐ alto ☐ bajo x dentro de los datos, entre _____ y _____: ☐ sí ☐ no

Repasa antes de entregar: ☐ he restado $\bar{x}^2$, $\bar{y}^2$ y $\bar{x}\cdot\bar{y}$ ☐ en r, desviaciones típicas ☐ en la pendiente, varianza de X ☐ r entre -1 y 1 ☐ la recta pasa por $(\bar{x},\ \bar{y})$ ☐ unidades ☐ frase de interpretación y de fiabilidad.

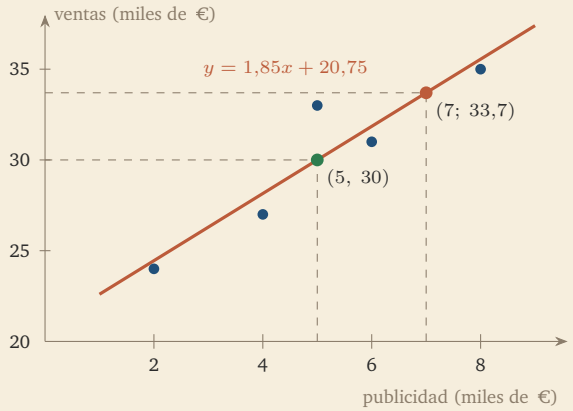
LA PLANTILLA RELLENA · PUBLICIDAD Y VENTAS (LECCIÓN 19)

Una tienda online anota durante cinco meses su gasto en publicidad, X, y sus ventas, Y, en miles de euros. a) Calcule el coeficiente de correlación lineal e interprételo (1,25 puntos). b) Halle la recta de regresión de Y sobre X (1 punto). c) Estime las ventas de un mes con 7.000 € de publicidad y valore la fiabilidad (0,75 puntos).

A · LA TABLA DE CÁLCULOS

x_i	y_i	x_i^2	y_i^2	$x_i\,y_i$
2	24	4	576	48
4	27	16	729	108
5	33	25	1.089	165
6	31	36	961	186
8	35	64	1.225	280
25	**150**	**145**	**4.580**	**787**

2 $\bar{x} = \dfrac{25}{5} = 5 \qquad \bar{y} = \dfrac{150}{5} = 30$



B · DISPERSIÓN, COVARIANZA Y CORRELACIÓN

3 $\sigma_x^2 = \dfrac{145}{5} - 5^2 = 29 - 25 = 4,\ \sigma_x = 2 \qquad \sigma_y^2 = \dfrac{4.580}{5} - 30^2 = 916 - 900 = 16,\ \sigma_y = 4$

4 $\sigma_{xy} = \dfrac{787}{5} - 5 \cdot 30 = 157{,}4 - 150 = 7{,}4.$ Positiva: relación directa.

5 $r = \dfrac{7{,}4}{2 \cdot 4} = \mathbf{0{,}925}.$ Relación lineal positiva y fuerte: los meses con más publicidad tienen más ventas.

C · RECTA Y ESTIMACIÓN

6 $m = \dfrac{7{,}4}{4} = 1{,}85.$ Cada 1.000 € más de publicidad se asocian con 1.850 € más de ventas.

7 $y - 30 = 1{,}85\,(x - 5) \Rightarrow \boldsymbol{y = 1{,}85x + 20{,}75}.$ Comprobación: $1{,}85 \cdot 5 + 20{,}75 = 30.$

8 Para $x = 7$: $y = 1{,}85 \cdot 7 + 20{,}75 = 12{,}95 + 20{,}75 = 33{,}7.$ Ventas estimadas: **33.700 €**.

9 Fiable: $r = 0{,}925$ está cerca de 1, y $x = 7$ queda dentro de los datos, entre 2 y 8.

GLOSARIO

Variable bidimensional. Dos datos medidos en cada individuo o caso, $(x_i,\ y_i)$.

Nube de puntos. Representación de todos los pares de datos en unos ejes.

Media. Suma de los datos entre el número de datos.

Varianza y desviación típica. Medidas de dispersión. La desviación típica es la raíz de la varianza.

Covarianza, σ_{xy}. Mide si las dos variables se mueven a la vez. Su signo da el sentido de la relación.

Coeficiente de correlación, r. Covarianza entre el producto de las desviaciones típicas. Va de -1 a 1 y mide la fuerza de la relación lineal.

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

Estimación. Valor de Y que da la recta para un valor de X.

Extrapolar. Estimar con un valor de X que queda fuera de los datos. Resta fiabilidad.

Causalidad. Que una variable provoque la otra. Una correlación alta no basta para afirmarla.