ediciones sm

Análisis estadístico de dos variables

1. La tabla siguiente muestra los resultados de diez pares de observaciones:

X	7	8	9	6	5	8	10	12	8	7
Y	82	78	86	72	65	80	95	99	89	74

a) Halla las rectas de regresión y represéntalas junto con el diagrama de dispersión.

b) ¿Qué grado de relación lineal existe entre ambas variables?

2. La publicidad anuncia para un determinado coche un consumo medio de 6,5 litros por cada 100 km. Durante 10 días se realizan mediciones sobre los litros consumidos y los kilómetros, recogidas en la siguiente tabla:

km	100	80	50	100	10	100	70	120	150	220
Litros	6,5	6	3	6	1	7	5,5	7,5	10	15

a) ¿Cuál es la diferencia entre el consumo medio obtenido y el que anunciaba el fabricante?

b) Imagina que se analiza el consumo medio según la tabla contando sólo los recorridos no inferiores a 100 km. Si la fiabilidad de la medición se estima en ± 0,1 litros/100 km, ¿hay motivo para quejarse al fabricante?

c) Determina el coeficiente de correlación lineal y la ecuación de la recta de regresión de Y sobre X.

d) Para realizar un viaje de 1 000 km, ¿cuánto combustible será necesario? Si el litro cuesta 0,75 euros y el peaje de la autopista 0,06 euros/km, ¿cuál será la cantidad de dinero necesaria para realizar el viaje?

3. Calcula el coeficiente de correlación lineal para los datos de la siguiente tabla:

X	2	4	5	6	8	11
Y	18	12	10	8	7	5

Halla el coeficiente de correlación de las variables: $Z = 2X + 6$; $T = 3Y - 15$. ¿Qué te sugiere el resultado?

4. Un colectivo de 50 alumnos de primero de bachillerato ha realizado tres exámenes de Matemáticas y tres de Filosofía. La clasificación de los alumnos por el número de exámenes aprobados está reflejada en la tabla siguiente:

F \ M	1	2	3
1	10	3	0
2	3	20	1
3	1	2	10

a) Calcula el número medio de exámenes de Filosofía aprobados por estos alumnos.

b) ¿Cómo es el grado de dependencia entre el número de exámenes aprobados en Filosofía y en Matemáticas?

c) Obtén la recta de regresión que explica la dependencia anterior.

5. Para una variable bidimensional se conocen $\bar{x} = 15{,}6$, $\bar{y} = 43{,}1$, $s_x = 3{,}9$, $s_y = 7{,}8$, $s_{xy} = -28{,}36$.

a) Calcula el coeficiente de correlación lineal y las pendientes de las rectas de regresión.

b) Escribe la recta de regresión de X sobre Y. Si $y = 45$, ¿qué valor tomará x?

6. Un conjunto de datos bidimensionales tiene coeficiente de correlación lineal $r = 0{,}8$. Además, $x = 3$, $y = 10$. Sin efectuar cálculos, razona por qué las siguientes ecuaciones no corresponden a la recta de regresión de Y sobre X: $y = -2x + 16$; $y = 1{,}5x + 1$; $y = -3{,}5x - 1$

SOLUCIONES

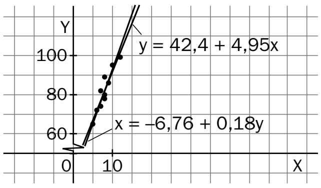
1. a) $\overline{x} = 8$; $\overline{y} = 82$; $s_x = 1{,}89$; $s_y = 9{,}97$; $s_{xy} = 17{,}7$

Recta de regresión de Y sobre X:

$$y - 82 = \frac{17{,}7}{1{,}89^2}(x - 8) \Leftrightarrow y = 42{,}4 + 4{,}95x$$

Recta de regresión de Y sobre X:

$$x - 8 = \frac{17{,}7}{9{,}97^2}(y - 82) \Leftrightarrow x = 6{,}76 + 0{,}18y$$



c) La relación lineal es positiva y muy fuerte.

2. a) $\dfrac{67{,}5}{1000} = \dfrac{6{,}75}{100}$

El consumo medio obtenido es 6,75 l/100 km frente a los 65 l/100 km anunciados.

b) $\dfrac{52}{790} = \dfrac{6{,}58}{100}$

El consumo medio en los recorridos no inferiores a 100 km es 6,5 l/100 km. Al ser la diferencia menor que 0,1 l/100 km, no hay motivo para quejarse al fabricante.

c) Si X representa los kilómetros e Y los litros:

$\overline{x} = 100$ km; $\overline{y} = 6{,}75$ l; $s_x = 54{,}04$ km;

$s_y = 3{,}59$ l; $s_{xy} = 192{,}5$ l;

$$r = \frac{192{,}5}{54{,}04 \cdot 3{,}59} = 0{,}992$$

La recta de regresión de Y sobre X es:

$$y - 6{,}75 = \frac{192{,}5}{54{,}04^2}(x - 100) \Leftrightarrow$$

$$\Leftrightarrow y = 0{,}1575 + 0{,}0659x$$

d) Combustible necesario: $y(1000) = 66$ l

Coste en euros: $6 \cdot 0{,}75 + 1000 \cdot 0{,}06 = 109{,}50$

3. $\overline{x} = \dfrac{36}{6} = 6$; $\overline{y} = 10$; $s_x = \sqrt{\dfrac{266}{6} - 6^2} = 2{,}9$

$$s_y = \sqrt{\frac{706}{6} - 10^2} = 4{,}2$$

$$s_{xy} = \frac{293}{6} - 6 \cdot 10 = -11{,}2;\ r = \frac{-11{,}2}{2{,}9 \cdot 4{,}2} = -0{,}92$$

z	t	z^2	t^2	zt
10	39	100	1521	390
14	21	196	441	294
16	15	256	225	240
18	9	324	81	162
22	6	484	36	132
28	0	784	0	0
108	90	2 144	2 304	1 218

De nuevo, $r = 0{,}92$, lo que sugiere que el coeficiente de correlación lineal no varía con transformaciones lineales.

4.

x_i (M)	y_i (F)	f_i
1	1	10
1	2	3
1	3	1
2	1	3
2	2	20
2	3	2
3	2	1
3	3	10

$\overline{x} = 1{,}94$; $\overline{y} = 2$

$s_x = 0{,}704$

$s_y = 0{,}721$

$s_{xy} = 0{,}38$

a) La respuesta es 2.

b) $r = \dfrac{0{,}38}{0{,}704 \cdot 0{,}721} = 0{,}7479$

luego la dependencia es positiva y fuerte.

c) $y - 2 = \dfrac{0{,}38}{0{,}704^2}(x - 1{,}94) \Rightarrow y = 0{,}51 + 0{,}77x$

5. a) $r = \dfrac{s_{xy}}{s_x \cdot s_y} = -0{,}932$

Pendiente de la recta de regresión de Y sobre X:

$$\frac{s_{xy}}{s_y^2} = -1{,}864$$

Pendiente de la recta de regresión de X sobre Y:

$$\frac{s_{xy}}{s_y^2} = -0{,}466$$

b) $x - 15{,}6 = -0{,}466\,(y - 3{,}1) \Rightarrow x = 35{,}68 - 0{,}47y$

$x = 35{,}68 - 0{,}47 \cdot 45 = 14{,}53$, estimación fiable, puesto que el valor de r está próximo a -1.

6. Al ser r positivo, la pendiente de la recta de regresión tiene que ser positiva, así, $y = -2x + 16$, e $y = -3{,}5x - 1$ están descartadas. Además, la recta de regresión tiene que pasar por $(\overline{x}, \overline{y}) = (3, 10)$, pero $y = 1{,}5x + 1$ no pasa por este punto.