

1. Se ha medido el peso, en kilogramos, y el volumen, en litros, de distintos tipos de maletas, obteniendo los resultados que se recogen en esta tabla:

Volumen	97	102	94	108	92	98
Peso	6'9	7'3	6'7	7'4	5'8	6'1

- a) Obtén el coeficiente de correlación, ¿cómo es la relación entre las dos variables? ($r=0'78598$)
 b) ¿Cuál será el peso estimado de una maleta cuyo volumen es 105 l.? ¿Es fiable esta estimación?
 c) ¿Cuál será el volumen estimado de una maleta que pesa 7,5 Kg.? ¿Es fiable esta estimación?

Solución:

Tabla:

	x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
	97	6,9	9409	47,610	669,30
	102	7,3	10404	53,290	744,60
	94	6,7	8836	44,890	629,80
	108	7,4	11664	54,760	799,20
	92	5,8	8464	33,640	533,60
	98	6,1	9604	37,210	597,80
Sumas	591	40,2	58.381	271,400	3974,3
Cuántos	6	6			

Cálculos,

$$\bar{x} = \frac{\sum x_i}{N} = \frac{591}{6} = 98'5 \qquad \bar{y} = \frac{\sum y_i}{N} = \frac{40'2}{6} = 6'7$$

$$\sigma_x = \sqrt{\frac{\sum x_i^2}{N} - \bar{x}^2} = \sqrt{\frac{58381}{6} - 98'5^2} = 5'28362$$

$$\sigma_y = \sqrt{\frac{\sum y_i^2}{N} - \bar{y}^2} = \sqrt{\frac{271'4}{6} - 6'7^2} = 0'58595$$

$$\sigma_{x,y} = \frac{\sum x_i y_i}{N} - \bar{x} \bar{y} = \frac{3974'3}{6} - 98'5 \cdot 6'7 = 2'43333$$

$$\rho = \frac{\sigma_{x,y}}{\sigma_x \sigma_y} = \frac{2'43333}{5'28362 \cdot 0'58595} = 0'78598$$

a) $\rho = 0'78598$, la relación entre las dos variables es positiva y muy fuerte

b) Sabiendo el volumen de la maleta (x) queremos calcular su peso (y), por lo que

debemos obtener la recta de regresión de y sobre x .

$$y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2} (x - \bar{x}) \quad y - 6'7 = \frac{2'43333}{(5'28362)^2} (x - 98'5)$$

$$\text{Para } x = 105, \quad y - 6'7 = \frac{2'43333}{(5'28362)^2} (105 - 98'5) \rightarrow y = 7'267$$

El peso estimado de la maleta de 105 l será de 7'3 kg.

Los valores de volumen estudiados van desde 92 l a 108 l, 105 l está entre los valores estudiados.

La relación entre las variables es muy fuerte.

Por lo tanto se cumplen las dos condiciones para que la estimación sea fiable.

La estimación obtenida es fiable.

- c) Sabiendo el peso de la maleta (y) queremos calcular su volumen (x), por lo que debemos obtener la recta de regresión de x sobre y .

$$x - \bar{x} = \frac{\sigma_{xy}}{\sigma_y^2} (y - \bar{y}) \quad x - 98'5 = \frac{2'43333}{(0'58595)^2} (y - 6'7)$$

$$\text{Para } y = 7'5, \quad x - 98'5 = \frac{2'43333}{(0'58595)^2} (7'5 - 6'7) \rightarrow x = 104'2$$

El volumen estimado de la maleta de 7'5 kg será de 104'2 l.

Los valores de peso estudiados van desde 5'8 kg a 7'4 kg, 7'5kg está fuera de los valores estudiados a una distancia del $\frac{7'5 - 7'4}{7'4 - 5'8} = 0'0625 = 6'25\% > 5\%$. No cumple la primera condición.

Por lo tanto la estimación no es fiable.

2. Se ha realizado una encuesta preguntando por el número de personas que habitan el hogar familiar y el número de habitaciones que tiene la casa. La tabla siguiente recoge la información obtenida:

Nº de personas	2	5	4	3	5	7
Nº de habitaciones	1	3	3	4	3	3

Halla la covarianza y el coeficiente de correlación. ¿Cómo es la relación entre las dos variables? ¿Cuántas habitaciones habrá en una casa en la que vivan 6 personas? ¿Cuántas personas vivirán en una casa con 6 habitaciones? ¿Son fiables las estimaciones?

Solución:

Tabla:

	x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
	2	1	4	1	2
	5	3	25	9	15
	4	3	16	9	12
	3	4	9	16	12
	5	3	25	9	15
	7	3	49	9	21
Sumas	26	17	128	53	77
Cuántos	6	6			

Cálculos,

$$\bar{x} = \frac{\sum x_i}{N} = \frac{26}{6} = 4'3333$$

$$\bar{y} = \frac{\sum y_i}{N} = \frac{17}{6} = 2'8333$$

$$\sigma_x = \sqrt{\frac{\sum x_i^2}{N} - \bar{x}^2} = \sqrt{\frac{128}{6} - 4'3333^2} = 1'5986$$

$$\sigma_y = \sqrt{\frac{\sum y_i^2}{N} - \bar{y}^2} = \sqrt{\frac{53}{6} - 2'8333^2} = 0'8975$$

$$\sigma_{xy} = \frac{\sum x_i y_i}{N} - \bar{x} \bar{y} = \frac{77}{6} - 4'3333 \cdot 2'8333 = 0'5556$$

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{0'5556}{1'5986 \cdot 0'8975} = 0'3872$$

$\rho = 0,3872$, la relación entre las dos variables es positiva y débil

Sabiendo las personas que viven en una casa (x) queremos calcular el nº de habitaciones (y), por lo que debemos obtener la recta de regresión de y sobre x .

$$y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2} (x - \bar{x}) \quad y - 2,8333 = \frac{0,5556}{(1,5986)^2} (x - 4,3333)$$

$$\text{Para } x = 6, \quad y - 2,8333 = \frac{0,5556}{(1,5986)^2} (6 - 4,3333) \rightarrow y = 3,196$$

El número estimado de habitaciones es 3 (aproximadamente).

La relación entre las variables es débil.

Por lo tanto no se cumplen una de las condiciones para que la estimación sea fiable. La estimación obtenida no es fiable.

Sabiendo el nº de habitaciones (y) queremos calcular el nº de personas (x), por lo que debemos obtener la recta de regresión de x sobre y .

$$x - \bar{x} = \frac{\sigma_{xy}}{\sigma_y^2} (y - \bar{y}) \quad x - 4,3333 = \frac{0,5556}{(0,8975)^2} (y - 2,8333)$$

$$\text{Para } y = 6, \quad x - 4,3333 = \frac{0,5556}{(0,8975)^2} (6 - 2,8333) \rightarrow x = 6,517$$

El número estimado de personas será de 7 (aproximadamente)

La relación entre las variables es débil.

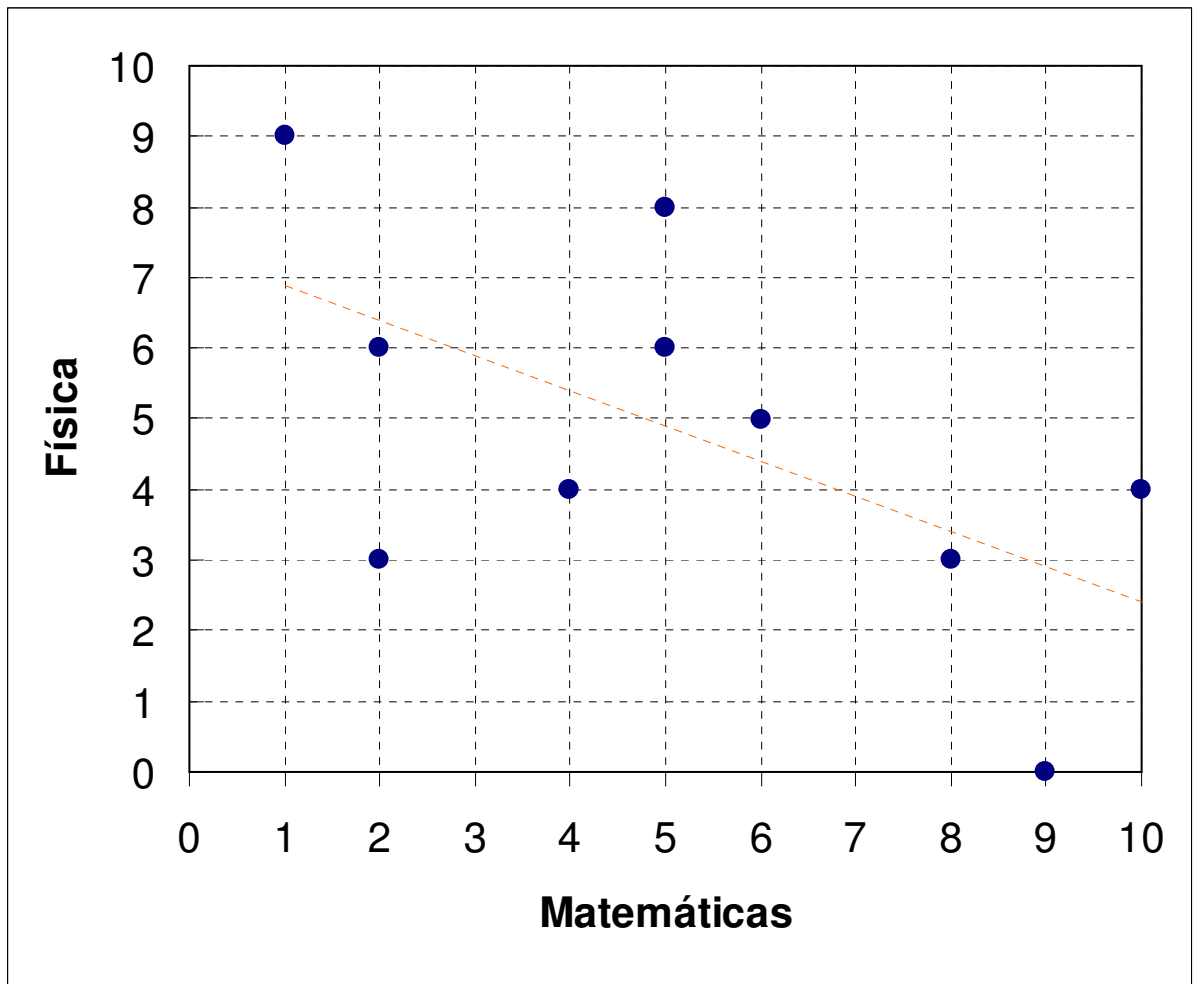
Por lo tanto no se cumplen una de las condiciones para que la estimación sea fiable. La estimación obtenida no es fiable.

3. Las notas de 10 alumnos y alumnas de una clase en Matemáticas y en Física han sido las siguientes:

Matemáticas	5	2	2	5	9	10	8	4	1	6
Física	8	6	3	6	0	4	3	4	9	5

Representa los datos mediante una nube de puntos y di, razonadamente y sin realizar cálculos, cuál de estos valores te parece más apropiado para el coeficiente de correlación: 0,58; 0,89; -0,58; -0,89.

Representamos la nube de puntos y trazamos, de forma aproximada, la recta de regresión:



La relación entre las notas de Matemáticas y Física es negativa, al aumentar la nota de Matemáticas disminuye la de Física, por lo que el coeficiente de correlación entre estas variables será negativo.

Como los puntos están alejados de la recta de regresión, el coeficiente de correlación será -0,58