

SI ENCUENTRAS ALGÚN ERROR COMUNÍCALO, POR FAVOR, AL CORREO DE LA PÁGINA WEB.



ESTADÍSTICA BIDIMENSIONAL

1. Calcula a partir de la tabla siguiente los parámetros estadísticos que has estudiado.

x_i	y_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$	$y_i \cdot f_i$	$y_i^2 \cdot f_i$	$x_i \cdot y_i \cdot f_i$
2	3	1	2	4	3	9	6
3	5	2	6	18	10	50	30
4	6	1	4	16	6	36	24
5	6	2	10	50	12	72	60
6	8	1	6	36	8	64	48
		7	28	124	39	231	168

· Media de x: $\bar{x} = \frac{\sum x_i \cdot f_i}{\sum f_i} = \frac{28}{7} = 4$

· Desviación típica de x: $\sigma_x = \sqrt{\frac{\sum x_i^2 \cdot f_i}{\sum f_i} - \bar{x}^2} = \sqrt{\frac{124}{7} - 4^2} = 1,31$

· Media de y: $\bar{y} = \frac{\sum y_i \cdot f_i}{\sum f_i} = \frac{39}{7} = 5,57$

· Desviación típica de y: $\sigma_y = \sqrt{\frac{\sum y_i^2 \cdot f_i}{\sum f_i} - \bar{y}^2} = \sqrt{\frac{231}{7} - 5,57^2} = 1,41$

· Covarianza: $\sigma_{xy} = \frac{\sum x_i \cdot y_i \cdot f_i}{\sum f_i} - \bar{x} \cdot \bar{y} = \frac{168}{7} - 4 \cdot 5,57 = 1,72$

· Coeficiente de correlación: $r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{1,72}{1,31 \cdot 1,41} = 0,93$. Correlación alta y positiva.

· Rectas de regresión:

De x sobre y: $x - \bar{x} = \frac{\sigma_{xy}}{\sigma_y^2} (y - \bar{y}); x - 4 = \frac{1,72}{1,41^2} (y - 5,57); x = 0,87y - 0,82$

Para calcular x sabiendo y.

De y sobre x: $y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2} (x - \bar{x}); y - 5,57 = \frac{1,72}{1,31^2} (x - 4); y = 1x + 1,56$

Para calcular y sabiendo x.

2. a. Calcula a partir de la tabla siguiente los parámetros estadísticos que has estudiado.
 b. ¿Qué valor toma x_i si $y_i = 3$
 c. ¿Qué valor toma y_i si $x_i = 3,7$

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
2	2	4	4	4
3	4	9	16	12
4	5	16	25	20
5	6	25	36	30
6	8	36	64	48
20	25	90	145	114

- Media de x : $\bar{x} = \frac{\sum x_i \cdot f_i}{\sum f_i} = \frac{20}{5} = 4$
- Desviación típica de x : $\sigma_x = \sqrt{\frac{\sum x_i^2 \cdot f_i}{\sum f_i} - \bar{x}^2} = \sqrt{\frac{90}{5} - 4^2} = 1,41$
- Media de y : $\bar{y} = \frac{\sum y_i \cdot f_i}{\sum f_i} = \frac{25}{5} = 5$
- Desviación típica de y : $\sigma_y = \sqrt{\frac{\sum y_i^2 \cdot f_i}{\sum f_i} - \bar{y}^2} = \sqrt{\frac{145}{5} - 5^2} = 2$
- Covarianza: $\sigma_{xy} = \frac{\sum x_i \cdot y_i \cdot f_i}{\sum f_i} - \bar{x}_i \cdot \bar{y}_i = \frac{114}{5} - 4 \cdot 5 = 2,8$
- Coeficiente de correlación: $r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{2,8}{1,41 \cdot 2} = 0,99$. Correlación alta y positiva.
- Rectas de regresión:
 De x sobre y : $x - \bar{x} = \frac{\sigma_{xy}}{\sigma_y^2} (y - \bar{y}); x - 4 = \frac{2,8}{2^2} (y - 5); x = 0,7 \cdot y + 0,5$

Para calcular x sabiendo y .

De y sobre x : $y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2} (x - \bar{x}); y - 5 = \frac{2,8}{1,41^2} (x - 4); y = 1,41 \cdot x - 0,63$

Para calcular y sabiendo x .

Si $y_i = 3 \rightarrow x - 4 = \frac{2,8}{2^2} (3 - 5) = 2,6$

Si $x_i = 3,7 \rightarrow y - 5 = \frac{2,8}{1,41^2} (3,7 - 4) = 4,58$

3. Dada la tabla siguiente, calcular el coeficiente de correlación.

$x_i \backslash y_i$	0	1	2	3	4	
0	17	22	6	4	1	50
1	15	14	8	2	0	39
2	13	6	10	5	1	35
3	5	4	2	6	2	19
4	3	1	0	3	4	11
	53	47	26	20	8	154

x_i	f_i	$x_i \cdot f_i$	x_i^2	$x_i^2 \cdot f_i$
0	53	0	0	0
1	47	47	1	47
2	26	52	4	104
3	20	60	9	180
4	8	32	16	128
	154	191		459

$$\begin{cases} \bar{x} = \frac{191}{154} = 1,26 \\ \sigma_x = \sqrt{\frac{459}{154} - 1,26^2} = 1,18 \end{cases}$$

y_i	f_i	$y_i \cdot f_i$	y_i^2	$y_i^2 \cdot f_i$
0	50	0	0	0
1	39	39	1	39
2	35	70	4	140
3	19	57	9	171
4	11	44	16	176
	154	210		526

$$\begin{cases} \bar{y} = \frac{210}{154} = 1,36 \\ \sigma_y = \sqrt{\frac{526}{154} - 1,36^2} = 1,25 \end{cases}$$

$$\sigma_{xy} = \frac{400}{154} - 1,36 \cdot 1,18 = 0,99$$

$$r = \frac{0,99}{1,25 \cdot 1,18} = 0,67$$

4. La estatura media de 100 escolares es de 155 cm. con una desviación típica de 15,5 cm. La recta de regresión de la estatura respecto al peso es $y = 80 + 1,5x$ (x : peso; y : estatura)

a) ¿Cuál es el peso medio de esos escolares?

b) ¿Cuál es el signo del coeficiente de correlación entre peso y estatura?

a. $155 = 80 + 1,5x$; $x = 50$ Kg.

b. Es de suponer que al aumentar la estatura también aumenta el peso, lo que significa una correlación positiva. Siendo, por tanto, el coeficiente de correlación positivo.

5. En un estudio realizado a los trabajadores de una cadena de fabricación de piezas de coches sobre su productividad quincenal, se relacionan las horas trabajadas (x) con las unidades producidas (y).

Sabemos que:

La recta de regresión de Y sobre X es: $y = 3,47x + 32,01$

La recta de regresión de X sobre Y es: $y = 3,81x + 5,36$

El intervalo de horas empleadas por los trabajadores es [60, 85].

a. Halla \bar{x} , \bar{y} y el coeficiente de correlación.

b. Si un operario trabaja 70 horas en una quincena, ¿cuántas unidades se estima que produzca? ¿Cómo de fiable es esta estimación? ¿Y si trabaja en total 40 horas? ¿Y si fueran 120 horas?

c. Si un empleado esta quincena ha llegado a producir 300 piezas, ¿cuántas horas se estima que ha trabajado?

Para hallar la media de x y la media de y , basta resolver el sistema formado por las rectas de regresión.

$$\begin{cases} y = 3,47x + 32,01 \\ y = 3,81x + 5,36 \end{cases} \rightarrow \begin{cases} \bar{x} = 78,38 \\ \bar{y} = 304 \end{cases}$$

Las pendientes de las rectas de regresión son:

$$\begin{cases} m_{yx} = \frac{\sigma_{xy}}{\sigma_y^2} \\ m_{xy} = \frac{\sigma_x^2}{\sigma_{xy}} \end{cases} \rightarrow m_{yx} : m_{xy} = \frac{\sigma_{xy}}{\sigma_y^2} : \frac{\sigma_x^2}{\sigma_{xy}} = \frac{\sigma_{xy}^2}{\sigma_y^2 \cdot \sigma_x^2} = r^2 \rightarrow r^2 = \frac{3,470}{3,81}; r = 0,95$$

b. Si $x = 70$; $y = 3,47 \cdot 70 + 32,01 = 274,91$ unidades producidas.

Si $x = 40$; $y = 3,47 \cdot 40 + 32,01 = 170,81$ unidades producidas.

Si $x = 120$; $y = 3,47 \cdot 120 + 32,01 = 448,41$ unidades producidas.

Las estimaciones más fiables son las realizadas teniendo en cuenta un número de horas comprendido entre 60 y 85.

c. Si $y = 300$; $300 = 3,81x + 5,36$; $x = \frac{300 - 5,36}{3,81} = 77,3$ h. Entre 77 y 78

6. La recta de regresión de Y sobre X de una cierta distribución bidimensional es $y = 1,6x - 3$. Sabemos que $\bar{x} = 10$ y $r = 0,8$.

a. Calcula \bar{y} .

b. Estima el valor de y para $x = 12$ y para $x = 50$. ¿Qué estimación te parece más fiable?

c. Halla la recta de regresión de X sobre Y.

a. $\bar{y} = 1,6 \cdot 10 - 3 = 13$

b. Si $x = 12$, $y = 1,6 \cdot 12 - 3 = 16,2$

Si $x = 50$, $y = 1,6 \cdot 50 - 3 = 77$

Es más fiable la primera pues $x = 12$ está más cerca de la media (10) que $x = 50$.

c. $m_{yx} : m_{xy} = \frac{\sigma_{xy}}{\sigma_y^2} : \frac{\sigma_x^2}{\sigma_{xy}} = \frac{\sigma_{xy}^2}{\sigma_y^2 \cdot \sigma_x^2} = r^2 \rightarrow m_{xy} = 2,5$

$y = 2,5x + c$; $13 = 2,5 \cdot 10 + c$; $c = -12$. $y = 2,5x - 12$

7. El consumo mensual de energía per cápita, y , en miles de kWh, y la renta per cápita, x , en miles de euros, de seis países son:

	A	B	C	D	E	F
x	11,1	8,5	11,3	4,5	9,9	6,5
y	5,7	5,0	5,1	2,7	4,6	3,1

a. Calcula la recta de regresión de Y sobre X.

b. Halla el coeficiente de correlación entre el consumo y la renta.

c. ¿Qué predicción podemos hacer sobre el consumo de energía per cápita de un país cuya renta per cápita es de 4400 €? (Recuerda que en la tabla se da la renta en miles de euros.)

d. Estima la renta per cápita que tendrá un país en el cual el consumo de energía per cápita ha sido de 9000 kWh.

e. ¿Cómo de fiables son estas estimaciones?

Igual que en los ejercicios uno y dos calculamos los siguientes parámetros:

$\bar{x} = 8,63$	$\bar{y} = 4,37$	$\sigma_x = 2,46$	$\sigma_y = 1,09$	$\sigma_{xy} = 2,51$
------------------	------------------	-------------------	-------------------	----------------------

a. $y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2} (x - \bar{x})$; $y - 4,37 = \frac{2,51}{2,46^2} (x - 8,63)$; $y = 0,8 + 0,41x$

5

b. Coeficiente de correlación: $r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{2,51}{2,46 \cdot 1,09} = 0,94$

c. Si $x = 4,4$; $y = 0,8 + 0,41 \cdot 4,4 = 2,604$

d. $x - \bar{x} = \frac{\sigma_{xy}}{\sigma_y^2} (y - \bar{y})$; $x - 8,63 = \frac{2,51}{1,09^2} (9 - 4,37) = 18,41$; 18410 €

e. En la primera estimación (apartado c), el valor $x = 4,4$ es próximo a los valores de la tabla. Como el coeficiente de correlación es alto (0,93), la estimación es razonablemente fiable. En la segunda estimación (apartado d), el valor $y = 9$ es lejano a los de la tabla. Por tanto, la estimación es poco fiable.