

REGRESIÓN: DISTRIBUCIONES BIDIMENSIONALES

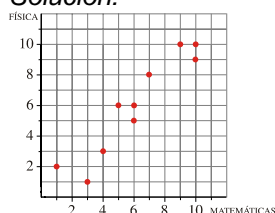
NUBES DE PUNTOS Y COEFICIENTES DE CORRELACIÓN

EJERCICIO 1 : Las notas de 10 alumnos y alumnas de una clase en Matemáticas y en Física han sido las siguientes:

| | | | | | | | | | | |
|-------------|---|---|---|---|----|----|---|---|----|---|
| Matemáticas | 7 | 6 | 4 | 5 | 9 | 10 | 3 | 1 | 10 | 6 |
| Física | 8 | 6 | 3 | 6 | 10 | 9 | 1 | 2 | 10 | 5 |

Representa los datos mediante una nube de puntos y di cuál de estos valores te parece más apropiado para el coeficiente de correlación: 0,23; 0,94; ☐ 0,37; ☐ 0,94.

Solución:



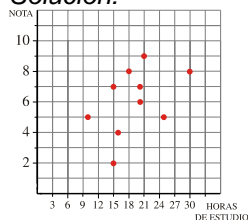
Viendo la representación, observamos que el coeficiente de correlación es positivo y alto. Por tanto, $r = 0,94$.

EJERCICIO 2 : Un grupo de 10 amigos se ha presentado a una prueba de oposición. Anotaron el número de horas que dedicaron a estudiar la semana antes del examen y la nota obtenida en la prueba. La información se recoge en la siguiente tabla:

| | | | | | | | | | | |
|------------------|----|----|----|----|----|----|----|----|----|----|
| Horas de estudio | 21 | 15 | 10 | 15 | 20 | 30 | 18 | 20 | 25 | 16 |
| Nota | 9 | 7 | 5 | 2 | 7 | 8 | 8 | 6 | 5 | 4 |

Representa los datos mediante una nube de puntos e indica cuál de estos valores te parece más apropiado para el coeficiente de correlación: 0,92; ☐ 0,44; ☐ 0,92; 0,44.

Solución:



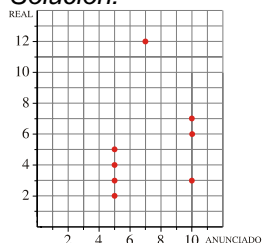
Observando la representación, vemos que el coeficiente de correlación es positivo y bajo. Por tanto, $r = 0,44$.

EJERCICIO 3 : En una empresa de televenta se ha anotado el plazo de entrega, en días, que anunciaban en los productos y el plazo real, también en días, de entrega de estos, obteniendo la siguiente tabla:

| | | | | | | | | | | |
|-----------------|---|----|----|---|---|---|----|---|----|----|
| Plazo anunciado | 5 | 10 | 7 | 5 | 5 | 5 | 10 | 5 | 10 | 10 |
| Plazo real | 3 | 3 | 12 | 3 | 5 | 2 | 3 | 4 | 7 | 6 |

Representa los datos mediante una nube de puntos e indica cuál de estos números te parece más apropiado para el coeficiente de correlación: 0,87; 0,2; ☐ 0,87; ☐ 0,2.

Solución:



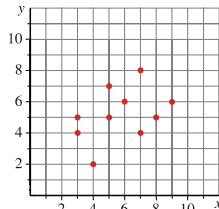
Vemos que la relación entre las variables es ligeramente positiva, pero muy baja. Por tanto, $r = 0,2$.

EJERCICIO 4 : Considera la siguiente distribución:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| x | 9 | 5 | 4 | 3 | 7 | 5 | 6 | 7 | 8 | 3 |
| y | 6 | 7 | 2 | 4 | 8 | 5 | 6 | 4 | 5 | 5 |

Representa los datos mediante una nube de puntos y di cuál de estos valores te parece más apropiado para el coeficiente de correlación: 0,99; \square 0,4; \square 0,83; 0,4.

Solución:



Vemos que hay una relación positiva entre las variables, pero es baja. Por tanto, $r = 0,4$.

COVARIANZA, COEFICIENTE DE CORRELACIÓN

EJERCICIO 5 : En un reconocimiento médico a los niños de un colegio, se les ha pesado, en kilogramos, y se les ha medido, en centímetros. Aquí tienes los datos de los primeros seis niños:

| | | | | | | |
|----------|-----|-----|-----|-----|-----|-----|
| Estatura | 120 | 110 | 140 | 130 | 125 | 115 |
| Peso | 25 | 30 | 35 | 25 | 20 | 20 |

Calcula la covarianza y el coeficiente de correlación. ¿Cómo es la relación entre las dos variables?

Solución:

| x_i | y_i | x_i^2 | y_i^2 | $x_i y_i$ |
|-------|-------|---------|---------|-----------|
| 120 | 25 | 14400 | 625 | 3000 |
| 110 | 30 | 12100 | 900 | 3300 |
| 140 | 35 | 19600 | 1225 | 4900 |
| 130 | 25 | 16900 | 625 | 3250 |
| 125 | 20 | 15625 | 400 | 2500 |
| 115 | 20 | 13225 | 400 | 2300 |
| 740 | 155 | 91850 | 4175 | 19250 |

• Medias: $\bar{x} = \frac{740}{6} = 123,33$

$\bar{y} = \frac{155}{6} = 25,83$

• Desviaciones típicas:

$\sigma_x = \sqrt{\frac{91850}{6} - 123,33^2} = \sqrt{98,04} = 9,90$

$\sigma_y = \sqrt{\frac{4175}{6} - 25,83^2} = \sqrt{28,64} = 5,35$

• Covarianza: $\sigma_{xy} = \frac{19250}{6} - 123,33 \cdot 25,83 = 22,72 \rightarrow \sigma_{xy} = 22,72$

• Coeficiente de correlación: $r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{22,72}{9,90 \cdot 5,35} = 0,43 \rightarrow r = 0,43$

• La relación entre las variables es positiva, pero débil.

EJERCICIO 6 : En seis modelos de zapatillas deportivas se ha estudiado el peso, en gramos, que tiene (para el número 42) y su precio, en euros. La información obtenida se recoge en esta tabla:

| | | | | | | |
|--------|-----|-----|-----|-----|-----|-----|
| Peso | 620 | 645 | 655 | 640 | 630 | 610 |
| Precio | 60 | 35 | 95 | 75 | 30 | 75 |

Calcula la covarianza y el coeficiente de correlación. ¿Cómo es la relación entre las dos variables?

Solución:

| x_i | y_i | x_i^2 | y_i^2 | $x_i y_i$ |
|-------|-------|---------|---------|-----------|
| 620 | 60 | 384400 | 3600 | 37200 |
| 645 | 35 | 416025 | 1225 | 22575 |
| 655 | 95 | 429025 | 9025 | 62225 |
| 640 | 75 | 409600 | 5625 | 48000 |
| 630 | 30 | 396900 | 900 | 18900 |
| 610 | 75 | 372100 | 5625 | 45750 |
| 3800 | 370 | 2408050 | 26000 | 234650 |

$\bar{x} = \frac{3800}{6} = 633,33$

• Medias: $\bar{y} = \frac{370}{6} = 61,67$

• Desviaciones típicas:

$$\sigma_x = \sqrt{\frac{2408,050}{6} - 633,33^2} = \sqrt{234,78} = 15,32$$

$$\sigma_y = \sqrt{\frac{26000}{6} - 61,67^2} = \sqrt{530,14} = 23,02$$

$$\bullet \text{ Covarianza: } \sigma_{xy} = \frac{234650}{6} - 633,33 \cdot 61,67 = 50,87 \rightarrow \sigma_{xy} = 50,87$$

$$\bullet \text{ Coeficiente de correlación: } r = \frac{50,87}{15,32 \cdot 23,02} = 0,14 \rightarrow r = 0,14$$

- La relación entre las variables es muy débil. Podemos decir que no están relacionadas.

EJERCICIO 7 : Se ha medido la potencia (en kW) y el consumo (litros/100 km) de 6 modelos distintos de coches, obteniéndose los siguientes resultados:

| | | | | | | |
|----------|-----|------|-----|-----|------|-----|
| Potencia | 81 | 85 | 66 | 85 | 104 | 83 |
| Consumo | 7,5 | 10,6 | 8,2 | 9,2 | 10,7 | 8,7 |

Halla la covarianza y el coeficiente de correlación. ¿Cómo es la relación entre las dos variables?

Solución:

| x_i | y_i | x_i^2 | y_i^2 | $x_i y_i$ |
|-------|-------|---------|---------|-----------|
| 81 | 7,5 | 6561 | 56,25 | 607,5 |
| 85 | 10,6 | 7225 | 112,36 | 901 |
| 66 | 8,2 | 4356 | 67,24 | 541,2 |
| 85 | 9,2 | 7225 | 84,64 | 782 |
| 104 | 10,7 | 10816 | 114,49 | 1112,8 |
| 83 | 8,7 | 6889 | 75,69 | 722,1 |
| 504 | 54,9 | 43072 | 510,67 | 4666,6 |

$$\bar{x} = \frac{504}{6} = 84$$

$$\bullet \text{ Medias: } \bar{y} = \frac{54,9}{6} = 9,15$$

$$\sigma_x = \sqrt{\frac{43072}{6} - 84^2} = \sqrt{122,67} = 11,08$$

$$\bullet \text{ D.T: } \sigma_y = \sqrt{\frac{510,67}{6} - 9,15^2} = \sqrt{1,39} = 1,18$$

$$\bullet \text{ Covarianza: } \sigma_{xy} = \frac{4666,6}{6} - 84 \cdot 9,15 = 9,17 \rightarrow \sigma_{xy} = 9,17$$

$$\bullet \text{ Coeficiente de correlación: } r = \frac{9,17}{11,08 \cdot 1,18} = 0,70 \rightarrow r = 0,70$$

- Hay una relación positiva y relativamente alta entre las variables.

EJERCICIO 8 : 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 | 3 | 5 | 4 | 6 | 5 | 4 |
| Nº de habitaciones | 2 | 3 | 4 | 4 | 3 | 3 |

Halla la covarianza y el coeficiente de correlación. ¿Cómo es la relación entre las dos variables?

Solución:

| x_i | y_i | x_i^2 | y_i^2 | $x_i y_i$ |
|-------|-------|---------|---------|-----------|
| 3 | 2 | 9 | 4 | 6 |
| 5 | 6 | 25 | 36 | 30 |
| 4 | 4 | 16 | 16 | 16 |
| 6 | 4 | 36 | 16 | 24 |
| 5 | 3 | 25 | 9 | 15 |
| 4 | 3 | 16 | 9 | 12 |
| 27 | 19 | 127 | 63 | 88 |

$$\bar{x} = \frac{27}{6} = 4,5$$

$$\bullet \text{ Medias: } \bar{y} = \frac{19}{6} = 3,17$$

$$\sigma_x = \sqrt{\frac{127}{6} - 4,5^2} = \sqrt{0,92} = 0,96$$

$$\bullet \text{ D.T: } \sigma_y = \sqrt{\frac{63}{6} - 3,17^2} = \sqrt{0,45} = 0,67$$

$$\bullet \text{ Covarianza: } \sigma_{xy} = \frac{88}{6} - 4,5 \cdot 3,17 = 0,40 \rightarrow \sigma_{xy} = 0,40$$

$$\bullet \text{ Coeficiente de correlación: } r = \frac{0,40}{0,96 \cdot 0,67} = 0,62 \rightarrow r = 0,62$$

- Hay una relación positiva, aunque no demasiado fuerte, entre las variables.

EJERCICIO 9 : Se han realizado unas pruebas de habilidad (puntuán de 0 a 5) en un grupo de alumnos. Las siguientes puntuaciones corresponden a las obtenidas por seis alumnos en dos de ellas:

| | | | | | | |
|-----------|---|---|---|---|---|---|
| 1ª Prueba | 5 | 5 | 4 | 3 | 2 | 4 |
| 2ª Prueba | 4 | 3 | 4 | 4 | 3 | 2 |

Calcula la covarianza y el coeficiente de correlación. ¿Cómo es la relación entre las variables?

Solución:

| x_i | y_i | x_i^2 | y_i^2 | $x_i y_i$ |
|-------|-------|---------|---------|-----------|
| 5 | 4 | 25 | 16 | 20 |
| 5 | 3 | 25 | 9 | 15 |
| 4 | 4 | 16 | 16 | 16 |
| 3 | 4 | 9 | 16 | 12 |
| 2 | 3 | 4 | 9 | 6 |
| 4 | 2 | 16 | 4 | 8 |
| 23 | 20 | 95 | 70 | 77 |

$$\begin{aligned} \bar{x} &= \frac{23}{6} = 3,83 \\ \bar{y} &= \frac{20}{6} = 3,33 \\ \sigma_x &= \sqrt{\frac{95}{6} - 3,83^2} = \sqrt{1,16} = 1,08 \\ \sigma_y &= \sqrt{\frac{70}{6} - 3,33^2} = \sqrt{0,58} = 0,76 \end{aligned}$$

- Covarianza: $\sigma_{xy} = \frac{77}{6} - 3,83 \cdot 3,33 = 0,079 \rightarrow \sigma_{xy} = 0,079$
- Coeficiente de correlación: $r = \frac{0,079}{1,08 \cdot 0,76} = 0,096 \rightarrow r = 0,096$
- La relación entre las variables es prácticamente nula.

RECTAS DE REGRESIÓN, ESTIMACIONES

EJERCICIO 10 : Se ha estudiado en distintas marcas de yogures naturales el porcentaje de grasa que contenían, así como las kilocalorías por envase. Estos son los resultados obtenidos en seis de ellos:

| | | | | | | |
|----------------|-----|----|-----|-----|----|----|
| X: Grasa (%) | 2,2 | 2 | 1,9 | 3,1 | 3 | 2 |
| Y: Kcal/envase | 64 | 55 | 58 | 79 | 65 | 52 |

a) Halla la recta de regresión de Y sobre X.

b) Calcula $\hat{y}(2,5)$ e $\hat{y}(10)$. ¿Son válidas estas estimaciones? (Sabemos que $r = 0,85$).

Solución:

a)

| x_i | y_i | x_i^2 | $x_i y_i$ |
|-------|-------|---------|-----------|
| 2,2 | 64 | 4,84 | 140,8 |
| 2 | 55 | 4 | 110 |
| 1,9 | 58 | 3,61 | 110,2 |
| 3,1 | 79 | 9,61 | 244,9 |
| 3 | 65 | 9 | 195 |
| 2 | 52 | 4 | 104 |
| 14,2 | 373 | 35,06 | 904,9 |

$$\begin{aligned} \bar{x} &= \frac{14,2}{6} = 2,37 \\ \bar{y} &= \frac{373}{6} = 62,17 \\ \sigma_x^2 &= \frac{35,06}{6} - 2,37^2 = 0,23 \\ \sigma_{xy} &= \frac{904,9}{6} - 2,37 \cdot 62,17 = 3,47 \end{aligned}$$

- Coeficiente de regresión: $m_{yx} = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{3,47}{0,23} = 15,1$
- Ecuación de la recta de regresión de Y sobre X: $y = 62,17 + 15,1(x - 2,37) \rightarrow y = 15,1x + 26,38$
- b) $\hat{y}(2,5) = 15,1 \cdot 2,5 + 26,38 = 64,13 \text{ kcal}$; $\hat{y}(10) = 15,1 \cdot 10 + 26,38 = 177,38 \text{ kcal}$

Como la correlación es alta, $r = 0,85$, es razonable hacer estimaciones dentro del intervalo de datos. Para un porcentaje del 2,5 de grasa, las kilocalorías serán, aproximadamente, 64,13. Sin embargo, la segunda estimación no es válida porque $x = 10$ está muy alejado del intervalo de datos que hemos considerado.

EJERCICIO 11 : 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:

| | | | | | | |
|------------|-----|-----|-----|-----|-----|-----|
| X: Volumen | 97 | 102 | 94 | 107 | 92 | 98 |
| Y: Peso | 6,9 | 7,1 | 6,7 | 7,4 | 5,8 | 6,1 |

a) Halla la recta de regresión de Y sobre X.

b) Calcula $\hat{y}(120)$. ¿Es fiable esta estimación? (Sabemos que $r = 0,79$).

| x_i | y_i | x_i^2 | $x_i y_i$ |
|-------|-------|---------|-----------|
| 97 | 6,9 | 9409 | 669,3 |
| 102 | 7,1 | 10404 | 724,2 |
| 94 | 6,7 | 8836 | 629,8 |
| 107 | 7,4 | 11449 | 791,8 |
| 92 | 5,8 | 8464 | 533,6 |
| 98 | 6,1 | 9604 | 597,8 |
| 590 | 40 | 58166 | 3946,5 |

$$\bar{x} = \frac{590}{6} = 98,33$$

• Medias: $\bar{y} = \frac{40}{6} = 6,67$

• Coeficiente de regresión: $m_{yx} = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{1,89}{25,54} = 0,07$

• Ecuación de la recta de regresión de y sobre x : $y = 6,67 + 0,07(x - 98,33) \rightarrow y = 0,07x - 0,21$

b) $\hat{y}(120) = 0,07 \cdot 120 - 0,21 = 8,19$

Como $x = 120$ está alejado del intervalo que estamos considerando, la estimación no es fiable.

EJERCICIO 12 : En distintos modelos de aspiradores se ha medido el peso, en kilogramos, y la capacidad útil de la bolsa, en litros, obteniendo los siguientes resultados:

| | | | | | | |
|--------------|-----|-----|-----|-----|-----|-----|
| X: Peso | 6,1 | 7 | 5,8 | 5,4 | 7 | 6,4 |
| Y: Capacidad | 1,9 | 4,3 | 1,5 | 1,7 | 2,9 | 3,2 |

a) Halla la recta de regresión de y sobre x .

b) Calcula $\hat{y}(6)$. ¿Es fiable esta estimación? (Sabemos que $r = 0,85$).

Solución:

a)

| x_i | y_i | x_i^2 | $x_i y_i$ |
|-------|-------|---------|-----------|
| 6,1 | 1,9 | 37,21 | 11,59 |
| 7 | 4,3 | 49 | 30,1 |
| 5,8 | 1,5 | 33,64 | 8,7 |
| 5,4 | 1,7 | 29,16 | 9,18 |
| 7 | 2,9 | 49 | 20,3 |
| 6,4 | 3,2 | 40,96 | 20,48 |
| 37,7 | 15,5 | 238,97 | 100,35 |

$$\bar{x} = \frac{37,7}{6} = 6,28$$

• Medias: $\bar{y} = \frac{15,5}{6} = 2,58$

• Varianza de x : $\sigma_x^2 = \frac{238,97}{6} - 6,28^2 = 0,39$

• Covarianza: $\sigma_{xy} = \frac{100,35}{6} - 6,28 \cdot 2,58 = 0,52$

• Coeficiente de regresión: $m_{yx} = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{0,52}{0,39} = 1,33$

• Ecuación de la recta de regresión de y sobre x : $y = 2,58 + 1,33(x - 6,28) \rightarrow y = 1,33x - 5,77$

b) $\hat{y}(6) = 1,33 \cdot 6 - 5,77 = 2,21$

Sí es fiable, puesto que la correlación es fuerte, $r = 0,85$, y $x = 6$ está dentro del intervalo de datos que estamos considerando. Para un peso de 6 kg la capacidad de la bolsa será, aproximadamente, de 2,21 litros.

EJERCICIO 13 : En seis institutos de la misma zona se ha estudiado la nota media de los estudiantes de 1º de bachillerato en Matemáticas y en Inglés, obteniéndose la información que se recoge en la siguiente tabla:

| | | | | | | |
|----------------|-----|-----|---|-----|-----|---|
| X: Matemáticas | 6,5 | 5,2 | 6 | 6,5 | 7 | 6 |
| Y: Inglés | 7 | 5 | 5 | 6 | 7,5 | 5 |

a) Halla la recta de regresión de y sobre x .

b) Calcula $\hat{y}(5,5)$. ¿Es fiable esta estimación? (Sabemos que $r = 0,87$).

| x_i | y_i | x_i^2 | $x_i y_i$ |
|-------|-------|---------|-----------|
| 6,5 | 7 | 42,25 | 45,5 |
| 5,2 | 5 | 27,04 | 26 |
| 6,0 | 5 | 36 | 30 |
| 6,5 | 6 | 42,25 | 39 |
| 7 | 7,5 | 49 | 52,5 |
| 6 | 5 | 36 | 30 |
| 37,2 | 35,5 | 232,54 | 223 |

$$\bar{x} = \frac{37,2}{6} = 6,2$$

• Medias:

$$\bar{y} = \frac{35,5}{6} = 5,92$$

• Varianza de x : $\sigma_x^2 = \frac{232,54}{6} - 6,2^2 = 0,32$

• Covarianza: $\sigma_{xy} = \frac{223}{6} - 6,2 \cdot 5,92 = 0,46$

• Coeficiente de regresión: $m_{yx} = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{0,46}{0,32} = 1,44$

• Ecuación de la recta de regresión de y sobre x : $y = 5,92 + 1,44(x - 6,2) \rightarrow y = 1,44x - 3$

b) $\hat{y}(5,5) = 1,44 \cdot 5,5 - 3 = 4,92$

Si es fiable la estimación, puesto que la correlación es fuerte, $r = 0,87$, y $x = 5,5$ está dentro del intervalo de valores que estamos considerando. Por tanto, estimamos que si la nota de Matemáticas es 5,5, la de Inglés será muy probablemente 4,9.

EJERCICIO 14 : Un grupo de seis atletas ha realizado pruebas de salto de longitud y de altura. Los resultados obtenidos han sido los siguientes:

| | | | | | | |
|-------------|---|---|---|---|---|---|
| X: Longitud | 5 | 4 | 5 | 4 | 4 | 3 |
| Y: Altura | 4 | 4 | 5 | 3 | 4 | 3 |

a) Halla las dos rectas de regresión y represéntalas.

b) Observando el grado de proximidad entre las dos rectas, ¿cómo crees que será la correlación entre las dos variables?

Solución:

| x_i | y_i | x_i^2 | y_i^2 | $x_i y_i$ |
|-------|-------|---------|---------|-----------|
| 5 | 4 | 25 | 16 | 20 |
| 4 | 4 | 16 | 16 | 16 |
| 5 | 5 | 25 | 25 | 25 |
| 4 | 3 | 16 | 9 | 12 |
| 4 | 4 | 16 | 16 | 16 |
| 3 | 3 | 9 | 9 | 9 |
| 25 | 23 | 107 | 91 | 98 |

$$\bar{x} = \frac{25}{6} = 4,17$$

• Medias:

$$\bar{y} = \frac{23}{6} = 3,83$$

$$\sigma_x = \sqrt{\frac{107}{6} - 4,17^2} = \sqrt{0,44} = 0,67$$

• D.T.:

$$\sigma_y = \sqrt{\frac{91}{6} - 3,83^2} = \sqrt{0,498} = 0,71$$

• Covarianza: $\sigma_{xy} = \frac{98}{6} - 4,17 \cdot 3,83 = 0,36$

• Coeficientes de regresión: y sobre $x \rightarrow m_{yx} = \frac{0,36}{0,44} = 0,82$

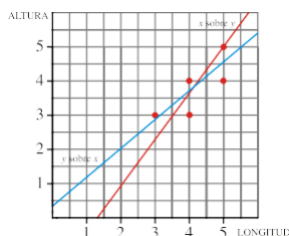
x sobre $y \rightarrow m_{xy} = \frac{0,36}{0,498} = 0,72$

• Rectas de regresión:

$$y \text{ sobre } x \rightarrow y = 3,83 + 0,82(x - 4,17) \rightarrow y = 0,82x + 0,41$$

$$x \text{ sobre } y \rightarrow x = 4,17 + 0,72(y - 3,83) \rightarrow x = 0,72y + 1,41 \Rightarrow y = \frac{x - 1,41}{0,72} \rightarrow y = 1,39x - 1,96$$

• Representación:



b) La correlación entre las dos variables no es demasiado fuerte, pues las dos rectas no están muy

próximas. Comprobamos que el coeficiente de correlación es: $r = \frac{0,36}{0,67 \cdot 0,71} = 0,76$

EJERCICIO 15 : La estatura, en centímetros, de seis chicos de la misma edad y la de sus padres viene recogida en la siguiente tabla:

| | | | | | | |
|----------|-----|-----|-----|-----|-----|-----|
| X: Hijo | 160 | 150 | 160 | 170 | 180 | 170 |
| Y: Padre | 180 | 170 | 175 | 185 | 180 | 175 |

a) Halla las dos rectas de regresión y represéntalas.

b) Observando el grado de proximidad entre las dos rectas, ¿cómo crees que será la correlación entre las dos variables?

Solución:

a)

| x_i | y_i | x_i^2 | y_i^2 | $x_i y_i$ |
|-------|-------|---------|---------|-----------|
| 160 | 180 | 25600 | 32400 | 28800 |
| 150 | 170 | 22500 | 28900 | 25500 |
| 160 | 175 | 25600 | 30625 | 28000 |
| 170 | 185 | 28900 | 34225 | 31450 |
| 180 | 180 | 32400 | 32400 | 32400 |
| 170 | 175 | 28900 | 30625 | 29750 |
| 990 | 1065 | 163900 | 189175 | 175900 |

$$\bar{x} = \frac{990}{6} = 165$$

$$\bullet \text{ Medias: } \bar{y} = \frac{1065}{6} = 177,5$$

$$\sigma_x = \sqrt{\frac{163900}{6} - 165^2} = \sqrt{91,67} = 9,57$$

$$\bullet \text{ D.T.: } \sigma_y = \sqrt{\frac{189175}{6} - 177,5^2} = \sqrt{22,92} = 4,79$$

$$\bullet \text{ Covarianza: } \sigma_{xy} = \frac{175900}{6} - 165 \cdot 177,5 = 29,17$$

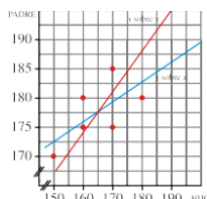
$$\bullet \text{ Coeficientes de regresión: } y \text{ sobre } x \rightarrow m_{yx} = \frac{29,17}{91,67} = 0,32 \quad x \text{ sobre } y \rightarrow m_{xy} = \frac{29,17}{22,92} = 1,27$$

• Rectas de regresión:

$$y \text{ sobre } x \rightarrow y = 177,5 + 0,32(x - 165) \rightarrow y = 0,32x + 124,7$$

$$x \text{ sobre } y \rightarrow x = 165 + 1,27(y - 177,5) \Rightarrow x = 1,27y - 60,43 \Rightarrow y = \frac{x + 60,43}{1,27} \rightarrow y = 0,79x + 47,58$$

• Representación:



b) La correlación entre las variables no es demasiado fuerte, pues las dos rectas no están muy

próximas. Comprobamos que el coeficiente de correlación es: $r = \frac{29,17}{9,57 \cdot 4,79} = 0,636$

EJERCICIO 16 : Se ha preguntado en seis familias por el número de hijos y el número medio de días que suelen ir al cine cada mes. Las respuestas han sido las siguientes:

| | | | | | | |
|--------------|---|---|---|---|---|---|
| X: Hijos | 2 | 1 | 3 | 4 | 2 | 3 |
| Y: Días cine | 3 | 4 | 4 | 2 | 1 | 4 |

a) Halla las dos rectas de regresión y represéntalas.

b) Observando el grado de proximidad entre las dos rectas, ¿cómo crees que será la correlación entre las dos variables?

Solución:

a)

| x_i | y_i | x_i^2 | y_i^2 | $x_i y_i$ |
|-------|-------|---------|---------|-----------|
| 2 | 3 | 4 | 9 | 6 |
| 1 | 4 | 1 | 16 | 4 |
| 3 | 4 | 9 | 16 | 12 |
| 4 | 2 | 16 | 4 | 8 |
| 2 | 1 | 4 | 1 | 2 |
| 3 | 4 | 9 | 16 | 12 |
| 15 | 18 | 43 | 62 | 44 |

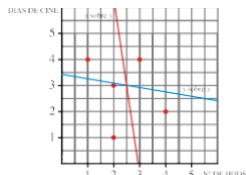
$$\bar{x} = \frac{15}{6} = 2,5$$

$$\bullet \text{ Medias: } \bar{y} = \frac{18}{6} = 3$$

$$\sigma_x = \sqrt{\frac{43}{6} - 2,5^2} = \sqrt{0,92} = 0,96$$

$$\bullet \text{ Desviaciones típicas: } \sigma_y = \sqrt{\frac{62}{6} - 3^2} = \sqrt{1,33} = 1,15$$

- Covarianza: $\sigma_{xy} = \frac{44}{6} - 2,5 \cdot 3 = -0,17$
- Coeficientes de regresión: $y \text{ sobre } x \rightarrow m_{yx} = \frac{-0,17}{0,92} = -0,18$ $x \text{ sobre } y \rightarrow m_{xy} = \frac{-0,17}{1,33} = -0,13$
- Rectas de regresión:
 $y \text{ sobre } x \rightarrow y = 3 - 0,18(x - 2,5) \rightarrow y = -0,18x + 3,45$
 $x \text{ sobre } y \rightarrow x = 2,5 - 0,13(y - 3) \Rightarrow x = -0,13y + 2,89 \Rightarrow 0,13y = 2,89 - x \Rightarrow$
 $y = \frac{-x + 2,89}{0,13} \rightarrow y = -7,69x + 22,23$
- Representación:



b) La correlación es prácticamente nula; las rectas son casi perpendiculares.

EJERCICIO 17 : Considera la siguiente distribución:

| X | 2 | 5 | 7 | 3 | 5 | 4 |
|---|---|----|----|----|----|----|
| Y | 5 | 12 | 15 | 10 | 15 | 10 |

- a) Halla las dos rectas de regresión y represéntalas.
b) Observando el grado de proximidad entre las dos rectas, ¿cómo crees que será la correlación entre las dos variables?

Solución:

a)

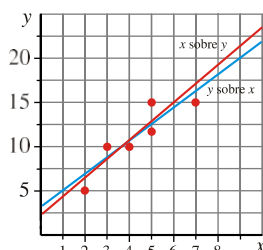
| x_i | y_i | x_i^2 | y_i^2 | $x_i y_i$ |
|-------|-------|---------|---------|-----------|
| 2 | 5 | 4 | 25 | 10 |
| 5 | 12 | 25 | 144 | 60 |
| 7 | 15 | 49 | 225 | 105 |
| 3 | 10 | 9 | 100 | 30 |
| 5 | 15 | 25 | 225 | 75 |
| 4 | 10 | 16 | 100 | 40 |
| 26 | 67 | 128 | 819 | 320 |

• Medias: $\bar{x} = \frac{26}{6} = 4,33$
 $\bar{y} = \frac{67}{6} = 11,17$

• Desviaciones típicas: $\sigma_x = \sqrt{\frac{128}{6} - 4,33^2} = \sqrt{2,58} = 1,61$
 $\sigma_y = \sqrt{\frac{819}{6} - 11,17^2} = \sqrt{11,73} = 3,43$

• Covarianza: $\sigma_{xy} = \frac{320}{6} - 4,33 \cdot 11,17 = 4,97$

- Coeficientes de regresión: $y \text{ sobre } x \rightarrow m_{yx} = \frac{4,97}{2,58} = 1,93$ $x \text{ sobre } y \rightarrow m_{xy} = \frac{4,97}{11,73} = 0,42$
- Rectas de regresión:
 $y \text{ sobre } x \rightarrow y = 11,17 + 1,93(x - 4,33) \rightarrow y = 1,93x + 2,81$
 $x \text{ sobre } y \rightarrow x = 4,33 + 0,42(y - 11,17) \Rightarrow x = 0,42y - 0,36 \Rightarrow y = \frac{x + 0,36}{0,42} \rightarrow y = 2,38x + 0,86$
- Representación:



b) La correlación es muy alta, puesto que las dos rectas están muy próximas, casi coinciden.

Comprobamos que el coeficiente de correlación es: $r = \frac{4,97}{1,61 \cdot 3,43} \approx 0,9$

OTROS

EJERCICIO 18 : Los gastos que una empresa tuvo en la publicidad de un determinado artículo en miles de euros y las ventas, también en miles de euros, de dicho artículo se recogen en la siguiente tabla:

| Gastos en publicidad (millones) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---------------------------------|----|----|----|----|----|----|----|----|
| Ventas (millones) | 15 | 16 | 14 | 17 | 20 | 18 | 18 | 19 |

Halla las medias, varianzas y desviaciones típicas de las dos variables, así como la covarianza de la distribución.

Solución:

| x_i | y_i | x_i^2 | y_i^2 | $x_i \cdot y_i$ |
|-------|-------|---------|---------|-----------------|
| 1 | 15 | 1 | 225 | 15 |
| 2 | 16 | 4 | 256 | 32 |
| 3 | 14 | 9 | 196 | 42 |
| 4 | 17 | 16 | 289 | 68 |
| 5 | 20 | 25 | 400 | 100 |
| 6 | 18 | 36 | 324 | 108 |
| 7 | 18 | 49 | 324 | 126 |
| 8 | 19 | 64 | 361 | 152 |
| 36 | 137 | 204 | 2375 | 643 |

Media de x : $\bar{x} = 36/8 = 4,5$ miles de euros

Media de y : $\bar{y} = 137/8 = 17,13$ miles de euros

Varianza de x : $\sigma_x^2 = \frac{204}{8} - 4,5^2 = 5,25$

Desviación típica de x : $\sigma_x = \sqrt{5,25} = 2,29$ miles de euros

Varianza de y : $\sigma_y^2 = \frac{2375}{8} - 17,13^2 = 3,61$

Desviación típica de y : $\sigma_y = \sqrt{3,61} = 1,9$ miles de euros

Covarianza: $\sigma_{xy} = \frac{643}{8} - 4,5 \cdot 17,13 = 3,29$

EJERCICIO 19 : La siguiente tabla recoge las medidas de los pesos en kg y las alturas en m de 20 alumnos:

| Nº de alumnos | 4 | 3 | 2 | 5 | 4 | 2 |
|------------------|------|------|------|------|------|------|
| Peso (X) en kg | 73 | 76 | 73 | 78 | 80 | 82 |
| Altura (Y) en m. | 1,65 | 1,68 | 1,70 | 1,72 | 1,76 | 1,80 |

Estima las medias, varianzas y desviaciones típicas de las variables estudiadas, así como la covarianza de ambas.

Solución:

| x_i | y_i | n_i | $x_i \cdot n_i$ | $y_i \cdot n_i$ | $x_i \cdot y_i \cdot n_i$ | $x_i^2 \cdot n_i$ | $y_i^2 \cdot n_i$ |
|-------|-------|-------|-----------------|-----------------|---------------------------|-------------------|-------------------|
| 73 | 1,65 | 4 | 292 | 6,60 | 481,80 | 21316 | 10,8900 |
| 76 | 1,68 | 3 | 228 | 5,04 | 383,04 | 17328 | 8,4672 |
| 73 | 1,7 | 2 | 146 | 3,40 | 248,20 | 10658 | 5,7800 |
| 78 | 1,72 | 5 | 390 | 8,60 | 670,80 | 30420 | 14,7920 |
| 80 | 1,76 | 4 | 320 | 7,04 | 563,20 | 25600 | 12,3904 |
| 82 | 1,8 | 2 | 164 | 3,60 | 295,20 | 13448 | 6,4800 |
| | | 20 | 1540 | 34,28 | 2642,24 | 118770 | 58,7996 |

Media de x : $\bar{x} = 1540/20 = 77$ Kg

Media de y : $\bar{y} = 34,28/20 = 1,714$ m

Varianza de x : $\sigma_x^2 = \frac{118770}{20} - 77^2 = 9,5$

Desviación típica de x : $\sigma_x = \sqrt{9,5} = 3,082$ Kg

Varianza de y : $\sigma_y^2 = \frac{58,7996}{20} - 1,714^2 = 0,002$

Desviación típica de y : $\sigma_y = \sqrt{0,002} = 0,045$

Covarianza: $\sigma_{xy} = \frac{2642,24}{20} - 77 \cdot 1,714 = 0,134$

EJERCICIO 20 : Halla los parámetros que caracterizan la distribución estadística de dos variables X e Y reflejadas en la tabla:

| | | Y | | |
|---|---|---|---|---|
| | | 0 | 2 | 4 |
| X | 1 | 2 | 1 | 2 |
| | 2 | 1 | 4 | 5 |
| | 3 | 3 | 2 | 0 |

Es decir: Medias, desviaciones y covarianza.

Solución:

| x_i | y_i | n_i | $x_i \cdot n_i$ | $y_i \cdot n_i$ | $x_i \cdot y_i \cdot n_i$ | $x_i^2 \cdot n_i$ | $y_i^2 \cdot n_i$ |
|-------|-------|-------|-----------------|-----------------|---------------------------|-------------------|-------------------|
| 1 | 0 | 2 | 2 | 0 | 0 | 2 | 0 |
| 2 | 0 | 1 | 2 | 0 | 0 | 4 | 0 |
| 3 | 0 | 3 | 9 | 0 | 0 | 27 | 0 |
| 1 | 2 | 1 | 1 | 2 | 2 | 1 | 4 |
| 2 | 2 | 4 | 8 | 8 | 16 | 16 | 16 |
| 3 | 2 | 2 | 6 | 4 | 12 | 18 | 8 |
| 1 | 4 | 2 | 2 | 8 | 8 | 2 | 32 |
| 2 | 4 | 5 | 10 | 20 | 40 | 20 | 80 |
| 3 | 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | 20 | 40 | 42 | 78 | 90 | 140 |

Media de x: $\bar{x} = 40/20 = 2$

Media de y: $\bar{y} = 42/20 = 2,1$

Varianza de x: $\sigma_x^2 = \frac{90}{20} - 2^2 = 0,5$

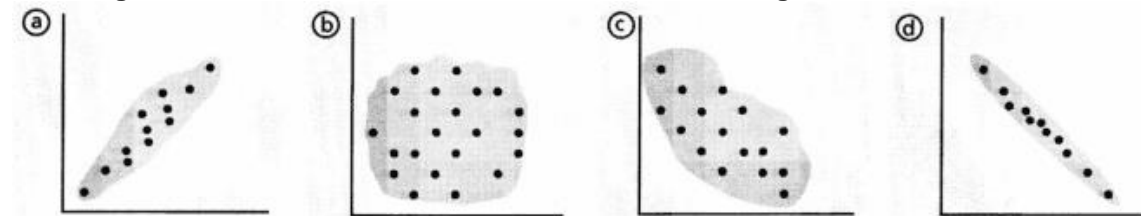
Desviación típica de x: $\sigma_x = \sqrt{0,5} = 0,707$

Varianza de y: $\sigma_y^2 = \frac{140}{20} - 2,1^2 = 2,59$

Desviación típica de y: $\sigma_y = \sqrt{2,59} = 1,609$

Covarianza: $\sigma_{xy} = \frac{78}{20} - 2 \cdot 2,1 = -0,3$

EJERCICIO 21 : Los números 0,1; 0,99; 0,6 y 0,89 son los valores absolutos del coeficiente de correlación de las distribuciones bidimensionales cuyas nubes de puntos adjuntamos. Asigna a cada diagrama su coeficiente de correlación cambiando el signo cuando sea necesario.



Solución:

a) $r = 0,89$ b) $r = 0,1$ c) $r = -0,6$ d) $r = -0,99$

EJERCICIO 22 : ¿Qué significa que en una distribución bidimensional el coeficiente de correlación sea?

a) $r = 1$ b) $r = -1$ c) $r = 0,75$ d) $r = 0$ e) $r = 0,1$ f) $r = 0,9$

Solución:

- a) $r = 1$, significa que existe dependencia funcional positiva.
- b) $r = -1$, significa que existe dependencia funcional negativa.
- c) $r = 0,75$; significa que existe dependencia aleatoria positiva fuerte.
- d) $r = 0$; significa que existe independencia aleatoria.
- e) $r = 0,1$; significa que existe independencia aleatoria.
- f) $r = 0,9$; significa que existe dependencia aleatoria positiva y muy fuerte.

EJERCICIO 23 : En una distribución bidimensional la recta de regresión de Y sobre X es $y = \bar{y}$ siendo \bar{y} la media de la distribución de la variable Y. ¿Cuál es la recta de X sobre Y? ¿Existe dependencia funcional entre Y y X? Razona la respuesta.

Solución:

- Si la recta de regresión de Y sobre X es $y = \bar{y}$ $\square m_{yx} = 0 \Rightarrow \sigma_{yx} = 0 \Rightarrow m_{xy} = 0$ y por tanto la recta de regresión de X sobre Y es $x = \bar{x} \Rightarrow x = \bar{x}$
- $r = \sqrt{m_{yx} \cdot m_{xy}} = 0 \Rightarrow$ No hay correlación, por tanto no hay dependencia funcional entre Y y X.

EJERCICIO 24 : Dada esta distribución bidimensional:

| X | 5 | 6,5 | 8 | 4 | 3 |
|---|-----|-----|-----|---|-----|
| Y | 4,5 | 7 | 7,5 | 5 | 3,5 |

- a) Calcula el coeficiente de correlación lineal, interpretando el resultado.
 b) Determina la recta de regresión de Y sobre X.
 c) Halla el punto donde se cortan las dos rectas de regresión.

Solución:

| x_i | y_i | x_i^2 | y_i^2 | $x_i \cdot y_i$ |
|-------|-------|---------|---------|-----------------|
| 5,0 | 4,5 | 25,00 | 20,25 | 22,5 |
| 6,5 | 7,0 | 42,25 | 49,00 | 45,5 |
| 8,0 | 7,5 | 64,00 | 56,25 | 60,0 |
| 4,0 | 5,0 | 16,00 | 25,00 | 20,0 |
| 3,0 | 3,5 | 9,00 | 12,25 | 10,5 |
| 26,5 | 27,5 | 156,25 | 162,75 | 158,5 |

Media de x: $\bar{x} = 26,5/5 = 5,3$

Media de y: $\bar{y} = 27,5/5 = 5,5$

Desviación típica de x: $\sigma_x = \sqrt{\frac{156,25}{5} - 5,3^2} = 1,78$

Desviación típica de y: $\sigma_y = \sqrt{\frac{162,75}{5} - 5,5^2} = 1,52$

Covarianza: $\sigma_{xy} = \frac{158,5}{5} - 5,3 \cdot 5,5 = 2,55$

- a) Coeficiente de correlación: $r = \frac{S_{xy}}{S_x \cdot S_y} = 0,94$

Al ser positivo y próximo a la unidad, la correlación es positiva (al aumentar X aumenta Y) y fuerte.

- b) Recta de regresión de Y sobre X:

$$m_x = \frac{S_{xy}}{S_x^2} = 0,81 \Rightarrow y - 5,5 = 0,81(x - 5,3) \Rightarrow y = 0,81x + 1,21$$

- c) El punto donde se cortan las dos rectas de regresión es: $\left(\begin{matrix} -- \\ |x, y| = (5,3; 5,5) \\ \end{matrix} \right)$

EJERCICIO 25 : Cinco niñas de 2, 3, 5, 7 y 8 años de edad pesan, respectivamente, 14, 20, 32, 42 y 44 kilos.

- a) Halla la ecuación de la recta de regresión de la edad sobre el peso.
 b) ¿Cuál sería el peso aproximado de una niña de 6 años?

Solución:

| x_i | y_i | x_i^2 | y_i^2 | $x_i \cdot y_i$ |
|-------|-------|---------|---------|-----------------|
| 2 | 14 | 4 | 196 | 28 |
| 3 | 20 | 9 | 400 | 60 |
| 5 | 32 | 25 | 1024 | 160 |
| 7 | 42 | 49 | 1764 | 294 |
| 8 | 44 | 64 | 1936 | 352 |
| 25 | 152 | 151 | 5320 | 894 |

Media de x: $\bar{x} = 25/5 = 5$ años

Media de y: $\bar{y} = 152/5 = 30,4$ Kg

Desviación típica de x: $\sigma_x = \sqrt{\frac{151}{5} - 5^2} = 2,28$ años

Desviación típica de y: $\sigma_y = \sqrt{\frac{5320}{5} - 30,4^2} = 11,83$ Kg

Covarianza: $\sigma_{xy} = \frac{894}{5} - 5 \cdot 30,4 = 26,8$

- a) Recta de regresión de X sobre Y: $m_y = \frac{S_{xy}}{S_y^2} = 0,192 \Rightarrow x - 5 = 0,192(y - 30,4) \Rightarrow x = 0,192y - 0,84$

- b) Recta de regresión de Y sobre X: $m_x = \frac{S_{xy}}{S_x^2} = 5,15 \Rightarrow y - 30,4 = 5,15(x - 5) \Rightarrow y = 5,15x + 4,65$

Para una niña cuya edad sea $x = 6$ años, se obtiene un peso de $y = 35,55$ kilos

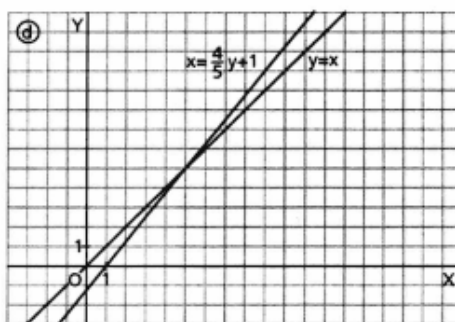
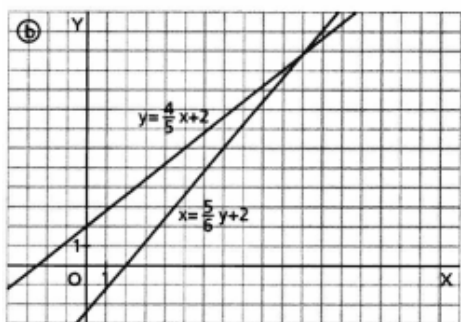
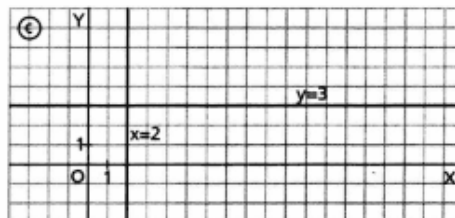
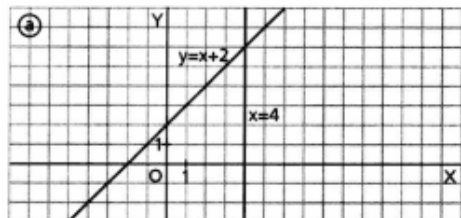
EJERCICIO 26 : Las rectas de regresión de cuatro distribuciones bidimensionales son las siguientes:

a) $y = x + 2$ $x = 4$ b) $y = \frac{4}{5}x + 2$ $x = \frac{5}{6}y + 2$

c) $y = 3$ $x = 2$ d) $y = x$ $x = \frac{4}{5}y + 1$

Di en qué casos es significativa la correlación lineal.

Solución: Basta con representar en un mismo diagrama los pares de rectas de cada apartado. Será más significativa la correlación lineal, cuanto menor sea el ángulo formado por las dos rectas de regresión.



Luego la correlación más significativa es la del apartado d), en segundo lugar b), seguida de a). Las rectas de regresión del apartado c) son perpendiculares, y por tanto, las variables están incorreladas.

EJERCICIO 27 : La media de los pesos de una población es de 65 kg y la de la estatura de 170 cm, mientras que las desviaciones típicas son de 5 kg y de 10 cm, respectivamente, y la covarianza de ambas variables es 40. Calcula la recta de regresión de los pesos respecto de las estaturas. ¿Cuánto se estima que pesará un individuo de 180 cm de estatura?

Solución: Del enunciado se obtienen los siguientes datos:

$$x = 65 \text{ kg}; y = 170 \text{ cm}; S_x = 5 \text{ kg}; S_y = 10 \text{ cm}; S_{xy} = 40$$

$$\text{Recta de regresión de los pesos sobre las alturas: } m_y = \frac{S_{xy}}{S_y^2} = 0,4 \Rightarrow x - 65 = 0,4(y - 170) \Rightarrow x = 0,4y - 3$$

Para estimar el peso de un individuo que mide $y = 180$ cm, basta con substituir dicho valor en la recta anterior, se tiene: $x = 0,4 \cdot 180 - 3 = 69$ kg

EJERCICIO 28 : Las estaturas y pesos de diez jugadores de baloncesto de un equipo son:

| Estatura (X) | 186 | 189 | 190 | 192 | 193 | 193 | 198 | 201 | 203 | 205 |
|--------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Pesos (Y) | 85 | 85 | 86 | 90 | 87 | 91 | 93 | 103 | 100 | 101 |

Teniendo en cuenta que $x \approx 195$ cm; $y \approx 92,1$ kg; $S_x \approx 6,07$; $S_y \approx 6,56$ y la covarianza $S_{xy} \approx 37,6$ se pide:

- Recta de regresión de Y sobre X.
- Halla el coeficiente de correlación.
- Si el equipo ficha a un jugador que mide 208 cm, ¿se puede predecir su peso? En caso afirmativo, obténlo.

Solución

- Con los datos suministrados, se tiene la siguiente recta de regresión de Y sobre X:

$$m_x = \frac{S_{xy}}{S_x^2} = 1,02 \Rightarrow y - 92,1 = 1,02(x - 195) \Rightarrow y = 1,02x - 106,8$$

b) El coeficiente de correlación es: $r = \frac{S_{xy}}{S_x \cdot S_y} = 0,94$ Correlación positiva y muy fuerte.

c) Substituyendo en la ecuación obtenida en el apartado a) el valor $x = 208$ cm, se tiene un peso $y = 105,36$ kg.

EJERCICIO 29 : Se ha observado una variable estadística bidimensional y se ha obtenido la siguiente tabla:

| | | X | | |
|---|----|-----|----|----|
| | | 100 | 50 | 25 |
| Y | 14 | 1 | 1 | |
| | 18 | 2 | 3 | |
| | 22 | | 1 | 2 |

Se pide:

- Calcula la covarianza.
- Obtén e interpreta el coeficiente de correlación lineal.
- Ecuación de la recta de regresión de Y sobre X.

Solución:

| X_i | y_i | n_i | $n_i \cdot X_i$ | $X_i^2 \cdot n_i$ | $n_i \cdot y_i$ | $y_i^2 \cdot n_i$ | $n_i \cdot X_i \cdot y_i$ |
|-------|-------|-------|-----------------|-------------------|-----------------|-------------------|---------------------------|
| 100 | 14 | 1 | 100 | 10000 | 14 | 196 | 1400 |
| 100 | 18 | 2 | 200 | 20000 | 36 | 648 | 3600 |
| 50 | 14 | 1 | 50 | 2500 | 14 | 196 | 700 |
| 50 | 18 | 3 | 150 | 7500 | 54 | 972 | 2700 |
| 50 | 22 | 1 | 50 | 2500 | 22 | 484 | 1100 |
| 25 | 22 | 2 | 50 | 1250 | 44 | 968 | 1100 |
| | | | 600 | 43750 | 184 | 3464 | 10600 |

Media de x : $\bar{x} = 600/10 = 60$

Media de y : $\bar{y} = 184/10 = 18,4$

Desv. típica de x :

$$\sigma_x = \sqrt{\frac{43750}{10} - 60^2} = 27,84$$

Desviación típica de y :

$$\sigma_y = \sqrt{\frac{3464}{10} - 18,4^2} = 2,8$$

$$\text{Covarianza: } \sigma_{xy} = \frac{10600}{10} - 60 \cdot 18,4 = -44$$

b) Coeficiente de correlación: $r = \frac{S_{xy}}{S_x \cdot S_y} = -0,56$

Se trata de una correlación negativa (al aumentar una variable, disminuye la otra) y débil ya que su valor absoluto está muy alejado de la unidad.

c) Recta de regresión de Y sobre X:

$$m_x = \frac{S_{xy}}{S_x^2} = -0,06 \Rightarrow y - 18,4 = -0,06(x - 60) \Rightarrow y = -0,06x + 22$$

EJERCICIO 30 : Un examen de cierta asignatura consta de dos partes, una teórica (x) y otra práctica (y). El profesor de la misma quiere ver si existe algún tipo de correlación entre las notas de teoría y práctica. Obtiene que la recta de regresión de y sobre x es $4x - 3y = 0$ y la de x sobre y es $3x - 2y = 1$.

- Calcular el coeficiente de correlación y decir si las variables están o no correlacionadas.
- Calcular la media de las notas de teoría y práctica.

Solución: x = Nota en teoría, y = Nota en práctica

a)

$$\text{La recta de regresión de } y \text{ sobre } x : 4x - 3y = 0 \Rightarrow y = \frac{4x}{3} \Rightarrow m_{yx} = 4/3$$

$$\text{La recta de regresión de } x \text{ sobre } y : 3x - 2y = 1 \Rightarrow x = \frac{2y + 1}{3} \Rightarrow m_{xy} = 2/3$$

El coeficiente de correlación $r = \sqrt{m_{yx} \cdot m_{xy}} = \sqrt{\frac{4}{3} \cdot \frac{2}{3}} = \sqrt{\frac{8}{9}} = 0,94 \sim 1$

Como el coeficiente de correlación es cercano a 1, hay correlación fuerte y positiva, por tanto si un alumno saca buena nota en teoría también la saca en práctica.

b) Para calcular la media de las notas de teoría y práctica, resolvemos el sistema de las dos rectas de

$$\text{regresión: } \begin{cases} 4x - 3y = 0 \\ 3x - 2y = 1 \end{cases} \Rightarrow x = 3, y = 4$$

Por tanto la nota media en teoría es de 3 y la nota media en práctica es de un 4.

EJERCICIO 31 : Un jugador de baloncesto juega una media de 22,5 minutos por partido, con una desviación típica de 5 minutos, obteniendo una media de 17,5 puntos, con una desviación típica de 6,5 puntos. El coeficiente de correlación entre minutos jugados y puntos conseguidos es 0,7. Estimar el número de puntos conseguidos si jugara en un partido 18 minutos.

Solución: Sea x = Tiempo(en minutos), y = Puntos conseguidos

Los datos: $\bar{x} = 22,5'$ $\sigma_x = 5'$ $\bar{y} = 17,5$ pto $\sigma_y = 6,5$ pto, $r = 0,7$

Como queremos hallar el número de puntos conseguidos, calcularemos la recta de regresión de Y sobre X:

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

Conocemos todo menos σ_{xy} : $r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} \Rightarrow \sigma_{xy} = r \cdot \sigma_x \cdot \sigma_y = 0,7 \cdot 5 \cdot 6,5 = 22,75$

Sustituyendo en la recta de regresión de Y sobre X: $y - 17,5 = \frac{22,75}{5^2} (18 - 22,5) \Rightarrow y = 13,405$ ptos

Marcará, aproximadamente: 13,4 pto

EJERCICIO 32 : Se ha hecho un test a 100 atletas sobre sus marcas en 100 metros y 400 metros. Se ha obtenido que la marca media en 100 metros es de 12,2 segundos con una desviación típica de 0,5 segundos, mientras que la marca media en 400 metros es de 61,3 segundos con una desviación típica de 1 segundo. Si el coeficiente de correlación lineal entre ambas pruebas es de 0,9,

a) ¿Podemos asegurar que los corredores que son mejores en 100 metros lo son también en 400 metros? (Justifica tu respuesta)

b) ¿Qué marca en 400 metros puede esperarse de un atleta que corre los 100 metros en 11 segundos?

Solución:

a) Si porque como $r = 0,9 > 0$, la correlación es positiva, aunque no sea demasiado buena.

b) Sea x = Marca en 100 m (en segundos), y = Marca en 400 m (en segundos)

Los datos: $\bar{x} = 12,2''$ $\sigma_x = 0,5''$ $\bar{y} = 61,3''$ $\sigma_y = 1''$, $r = 0,9$

Como queremos hallar la marca en 400 m, calcularemos la recta de regresión de Y sobre X:

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

Conocemos todo menos σ_{xy} : $r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} \Rightarrow \sigma_{xy} = r \cdot \sigma_x \cdot \sigma_y = 0,9 \cdot 0,5 \cdot 1 = 0,45$

Sustituyendo en la recta de regresión de Y sobre X: $y - 61,3 = \frac{0,45}{0,5^2} (11 - 12,2) \Rightarrow y = 59,14''$

Su marca aproximada en 400 metros será de 59,14''.