

Distribuciones bidimensionales

Ejercicio nº 1.-

Se ha medido el número medio de horas de entrenamiento a la semana de un grupo de 10 atletas y el tiempo, en minutos, que han hecho en una carrera, obteniendo los siguientes resultados:

Horas de entrenamiento	5	6	6	5	8	6	8	10	7	4
Tiempo carrera	30	23	24	24	22	21	24	20	23	28

Representa los datos mediante una nube de puntos y di cuál de estos valores te parece más apropiado para el coeficiente de correlación: 0,71; -0,71; 0,45; -0,32.

Ejercicio nº 2.-

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.

Ejercicio nº 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.

Ejercicio nº 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; -0,4; -0,83; 0,4.

Ejercicio nº 5.-

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.

Ejercicio nº 6.-

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?

Ejercicio nº 7.-

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?

Ejercicio nº 8.-

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?

Ejercicio nº 9.-

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?

Ejercicio nº 10.-

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?

Ejercicio nº 11.-

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 $r = 0,85$).

Ejercicio nº 12.-

Se ha analizado en distintos modelos de impresoras cuál es el coste por página (en céntimos de euro) en blanco y negro y cuál es el coste por página si esta es en color. La siguiente tabla nos da los seis primeros pares de datos obtenidos:

X: B Y N	8	11	17	21	14	10
Y: Color	33	49	95	106	58	53

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

b) ¿Cuánto nos costaría imprimir una página en color en una impresora en la que el coste por página en blanco y negro fuera de 12 céntimos de euro? ¿Es fiable la estimación? (Sabemos que $r = 0,97$).

Ejercicio nº 13.-

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 $r = 0,85$).

Ejercicio nº 14.-

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 $r=0,87$).

Ejercicio nº 15.-

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 $r=0,79$).

Ejercicio nº 16.-

En una academia para aprender a conducir se han estudiado las semanas de asistencia a clase de sus alumnos y las semanas que tardan en aprobar el examen teórico (desde que se apuntaron a la autoescuela). Los datos correspondientes a seis alumnos son:

X: Asistencia	6	1	4	3	5	8
Y: Aprobado	6	5	5	6	5	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?

Ejercicio nº 17.-

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?

Ejercicio nº 18.-

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?

Ejercicio nº 19.-

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?

Ejercicio nº 20.-

Un grupo de seis atletas ha realizado pruebas de salto de longitud y de altura. Las dos se han puntuado en una escala de 0 a 5. 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?

Soluciones

Distribuciones bidimensionales

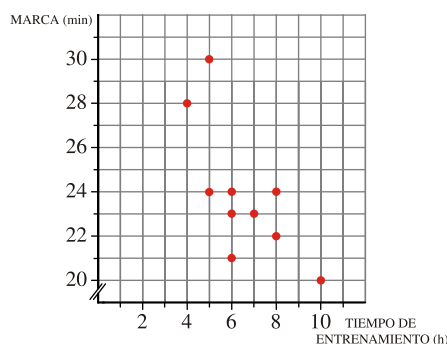
Ejercicio nº 1.-

Se ha medido el número medio de horas de entrenamiento a la semana de un grupo de 10 atletas y el tiempo, en minutos, que han hecho en una carrera, obteniendo los siguientes resultados:

Horas de entrenamiento	5	6	6	5	8	6	8	10	7	4
Tiempo carrera	30	23	24	24	22	21	24	20	23	28

Representa los datos mediante una nube de puntos y di cuál de estos valores te parece más apropiado para el coeficiente de correlación: 0,71; -0,71; 0,45; -0,32.

Solución:



A la vista de la representación, observamos que el coeficiente de correlación, r , es negativo y relativamente alto. Por tanto, $r = -0,71$.

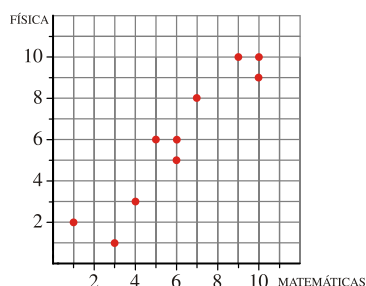
Ejercicio nº 2.-

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, $r = 0,94$.

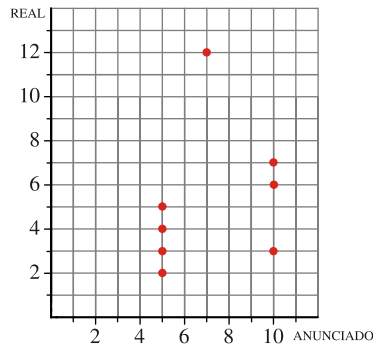
Ejercicio nº 3.-

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, $r = 0,2$.

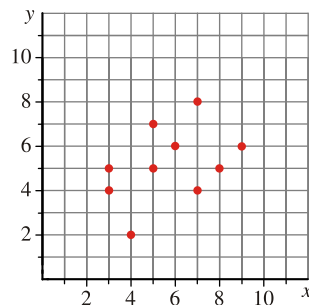
Ejercicio nº 4.-

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; -0,4; -0,83; 0,4.

Solución:



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

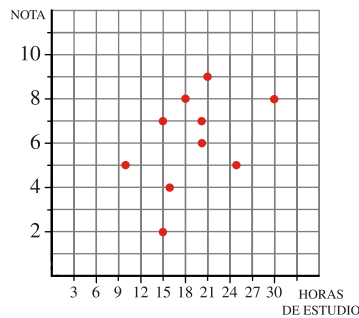
Ejercicio nº 5.-

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 nº 6.-

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	9	15
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

- Medias:

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

$$\bar{y} = \frac{19}{6} = 3,17$$

- Desviaciones típicas:

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

$$\sigma_y = \sqrt{\frac{63}{6} - 3,17^2} = \sqrt{0,45} = 0,67$$

- Covarianza:

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

- 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 nº 7.-

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

- Medias:

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

$$\bar{y} = \frac{20}{6} = 3,33$$

- Desviaciones típicas:

$$\sigma_x = \sqrt{\frac{95}{6} - 3,83^2} = \sqrt{1,16} = 1,08$$

$$\sigma_y = \sqrt{\frac{70}{6} - 3,33^2} = \sqrt{0,58} = 0,76$$

- 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.

Ejercicio nº 8.-

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 nº 9.-

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

- Medias:

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

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

- Desviaciones típicas:

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

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

- Covarianza:

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

- 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 nº 10.-

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

- Medias:

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

$$\bar{y} = \frac{54,9}{6} = 9,15$$

- Desviaciones típicas:

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

$$\sigma_y = \sqrt{\frac{510,67}{6} - 9,15^2} = \sqrt{1,39} = 1,18$$

- Covarianza:

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

- 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 nº 11.-

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

- Medias:

$$\bar{x} = \frac{14,2}{6} = 2,37$$

$$\bar{y} = \frac{373}{6} = 62,17$$

- Varianza de x:

$$\sigma_x^2 = \frac{35,06}{6} - 2,37^2 = 0,23$$

- Covarianza:

$$\sigma_{xy} = \frac{904,9}{6} - 2,37 \cdot 62,17 = 3,47$$

- 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 nº 12.-

Se ha analizado en distintos modelos de impresoras cuál es el coste por página (en céntimos de euro) en blanco y negro y cuál es el coste por página si esta es en color. La siguiente tabla nos da los seis primeros pares de datos obtenidos:

X: B Y N	8	11	17	21	14	10
Y: Color	33	49	95	106	58	53

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

b) ¿Cuánto nos costaría imprimir una página en color en una impresora en la que el coste por página en blanco y negro fuera de 12 céntimos de euro? ¿Es fiable la estimación? (Sabemos que $r = 0,97$).

Solución:

a)

x_i	y_i	x_i^2	$x_i y_i$
8	33	64	264
11	49	121	539
17	95	289	1615
21	106	441	2226
14	58	196	812
10	53	100	530
81	394	1211	5986

• Medias:

$$\bar{x} = \frac{81}{6} = 13,5$$

$$\bar{y} = \frac{394}{6} = 65,67$$

• Varianza de x:

$$\sigma_x^2 = \frac{1211}{6} - 13,5^2 = 19,58$$

• Covarianza:

$$\sigma_{xy} = \frac{5986}{6} - 13,5 \cdot 65,67 = 11,112$$

• Coeficiente de regresión:

$$m_{yx} = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{11,112}{19,58} = 5,68$$

• Ecuación de la recta de regresión de y sobre x :

$$y = 65,67 + 5,68(x - 13,5) \rightarrow y = 5,68x - 11,01$$

b) $\hat{y}(12) = 5,68 \cdot 12 - 11,01 \rightarrow \hat{y}(12) = 57,15$ céntimos de euro

Como la correlación es alta, $r = 0,97$, y $x = 12$ queda dentro del intervalo de valores que tenemos, la estimación sí es fiable. Si el coste de la página en blanco y negro es de 12 céntimos de euro, muy probablemente costará 57,15 céntimos de euro imprimirla en color.

Ejercicio nº 13.-

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

• Medias:

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

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

• 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 nº 14.-

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 $r = 0,87$).

Solución:

a)

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

- Medias:

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

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

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

Sí 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 nº 15.-

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 $r = 0,79$).

Solución:

a)

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

• Medias:

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

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

• Varianza de x:

$$\sigma_x^2 = \frac{58166}{6} - 98,33^2 = 25,54$$

• Covarianza:

$$\sigma_{xy} = \frac{3946,5}{6} - 98,33 \cdot 6,67 = 1,89$$

• 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 nº 16.-

En una academia para aprender a conducir se han estudiado las semanas de asistencia a clase de sus alumnos y las semanas que tardan en aprobar el examen teórico (desde que se apuntaron a la autoescuela). Los datos correspondientes a seis alumnos son:

X: Asistencia	6	1	4	3	5	8
Y: Aprobado	6	5	5	6	5	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	$x_i y_i$
6	6	36	36	36
1	5	1	25	5
4	5	16	25	20
3	6	9	36	18
5	5	25	25	25
8	10	64	100	80
27	37	151	247	184

- Medias:

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

$$\bar{y} = \frac{37}{6} = 6,17$$

- Desviaciones típicas:

$$\sigma_x = \sqrt{\frac{151}{6} - 4,5^2} = \sqrt{4,92} = 2,22$$

$$\sigma_y = \sqrt{\frac{247}{6} - 6,17^2} = \sqrt{3,1} = 1,76$$

- Covarianza:

$$\sigma_{xy} = \frac{184}{6} = 4,5 \cdot 6,17 = 2,9$$

- Coeficientes de regresión:

$$y \text{ sobre } x \rightarrow m_{yx} = \frac{2,9}{4,92} = 0,59$$

$$x \text{ sobre } y \rightarrow m_{xy} = \frac{2,9}{3,1} = 0,94$$

- Rectas de regresión:

$$y \text{ sobre } x \rightarrow y = 6,17 + 0,59(x - 4,5) \rightarrow y = 0,59x + 3,52$$

$$x \text{ sobre } y \rightarrow x = 4,5 + 0,94(y - 6,17)$$

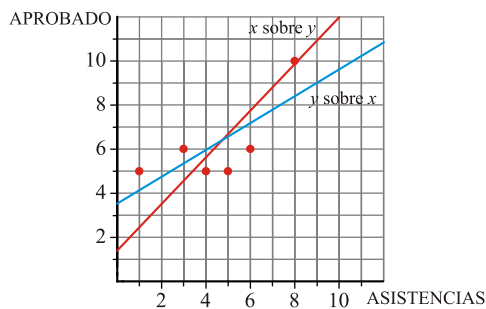
$$x = 4,5 + 0,94y - 5,80$$

$$x = 0,94y - 1,3$$

$$x + 1,3 = 0,94y$$

$$y = \frac{x + 1,3}{0,94} \rightarrow y = 1,06x + 1,38$$

- Representación:



- b) La correlación entre las variables no es demasiado fuerte, pues las dos rectas no están muy próximas. Con los datos obtenidos comprobamos que el coeficiente de correlación es: $r = 0,74$

Ejercicio nº 17.-

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

- Medias:

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

$$\bar{y} = \frac{1065}{6} = 177,5$$

- Desviaciones típicas:

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

$$\sigma_y = \sqrt{\frac{189175}{6} - 177,5^2} = \sqrt{22,92} = 4,79$$

- Covarianza:

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

- Coeficientes de regresión:

$$y \text{ sobre } x \rightarrow m_{yx} = \frac{29,17}{91,67} = 0,32$$

$$x \text{ sobre } y \rightarrow m_{xy} = \frac{29,17}{22,92} = 1,27$$

- Rectas de regresión:

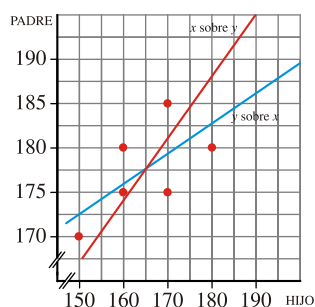
$$y \text{ sobre } x \rightarrow y = 177,5 + 0,32(x - 165) \rightarrow y = 0,32x + 124,7$$

$$x \text{ sobre } y \rightarrow x = 165 + 1,27(y - 177,5)$$

$$x = 1,27y - 60,43$$

$$y = \frac{x + 60,43}{1,27} \rightarrow y = 0,79x + 47,58$$

- 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 nº 18.-

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

- Medias:

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

$$\bar{y} = \frac{18}{6} = 3$$

- Desviaciones típicas:

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

$$\sigma_y = \sqrt{\frac{62}{6} - 3^2} = \sqrt{1,33} = 1,15$$

- Covarianza:

$$\sigma_{xy} = \frac{44}{6} - 2,5 \cdot 3 = -0,17$$

- Coeficientes de regresión:

$$y \text{ sobre } x \rightarrow m_{yx} = \frac{-0,17}{0,92} = -0,18$$

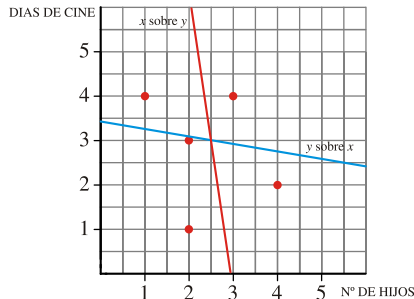
$$x \text{ sobre } y \rightarrow m_{xy} = \frac{-0,17}{1,33} = -0,13$$

- Rectas de regresión:

$$y \text{ sobre } x \rightarrow y = 3 - 0,18(x - 2,5) \rightarrow y = -0,18x + 3,45$$

$$\begin{aligned}
 x \text{ sobre } y &\rightarrow x = 2,5 - 0,13(y - 3) \\
 &x = -0,13y + 2,89 \\
 0,13y &= 2,89 - x \\
 y &= \frac{-x + 2,89}{0,13} \rightarrow y = -7,69x + 22,23
 \end{aligned}$$

• Representación:



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

Ejercicio nº 19.-

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:

$$\begin{aligned}
 \bar{x} &= \frac{26}{6} = 4,33 \\
 \bar{y} &= \frac{67}{6} = 11,17
 \end{aligned}$$

- 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:

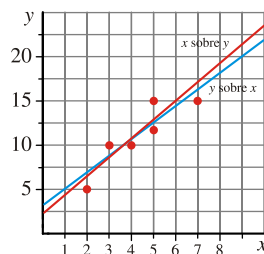
$$y \text{ sobre } x \rightarrow y = 11,17 + 1,93(x - 4,33) \rightarrow y = 1,93x + 2,81$$

$$x \text{ sobre } y \rightarrow x = 4,33 + 0,42(y - 11,17)$$

$$x = 0,42y - 0,36$$

$$y = \frac{x + 0,36}{0,42} \rightarrow y = 2,38x + 0,86$$

- Representación:



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

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

Ejercicio nº 20.-

Un grupo de seis atletas ha realizado pruebas de salto de longitud y de altura. Las dos se han puntuado en una escala de 0 a 5. 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:

a)

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

- Medias:

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

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

- Desviaciones típicas:

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

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

- Covarianza:

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

- Coeficientes de regresión:

$$y \text{ sobre } x \rightarrow m_{yx} = \frac{0,36}{0,44} = 0,82$$

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

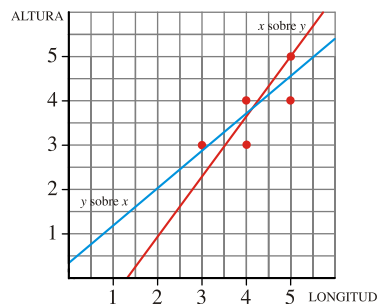
- Rectas de regresión:

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

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

$$y = \frac{x - 1,41}{0,72} \rightarrow y = 1,39x - 1,96$$

- 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: $r = \frac{0,36}{0,67 \cdot 0,71} = 0,76$