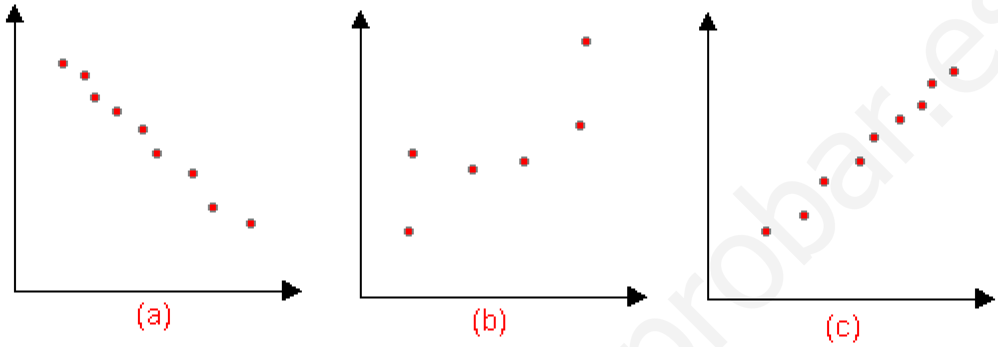
Resuelve tú (Pág ❸⓿❾)

Dibuja nubes de puntos que describan aceptablemente bien las relaciones entre las variables que se indican:

(a) Libros leídos y faltas de ortografía cometidas (inversa y muy fuerte).

(b) Afición al ciclismo en TV y práctica del mismo (muy débil).

(c) Consumo de grasas y colesterol (directa y fuerte).

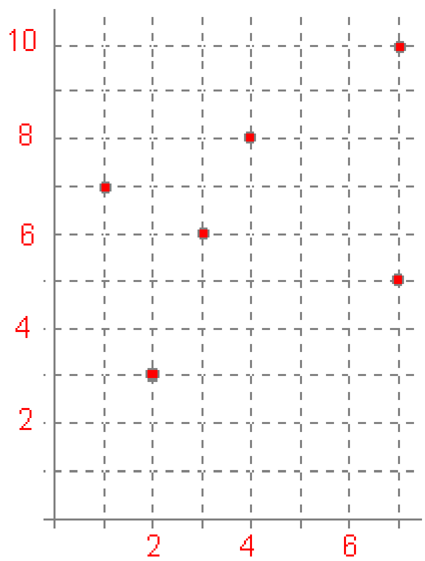


Resuelve tú (Pág ❸❶❶)

El número de problemas resueltos antes de una prueba y los aciertos obtenidos en ella, por 6 alumnos escogidos al azar, fueron:

Problemas resueltos, x:	2	3	1	7	4	7
Aciertos, y:	3	6	7	5	8	10

Representa gráficamente estos puntos. Halla el centro de gravedad de la nube de puntos obtenida y las desviaciones típicas marginales.



Para hallar las medias marginales (centro de gravedad) y las desviaciones típicas elaboramos la tabla base de los cálculos :

x_i	y_i	x_i^2	y_i^2
2	3	4	9
3	6	9	36
1	7	1	49
7	5	49	25
4	8	16	64
7	10	49	100
$\sum x_i = 24$	$\sum y_i = 39$	$\sum x_i^2 = 128$	$\sum y_i^2 = 283$

$$\bar{x} = \frac{\sum_{i=1}^{6} x_i}{n} = \frac{24}{6} = 4 \,;\, \bar{y} = \frac{\sum_{i=1}^{6} y_i}{n} = \frac{39}{6} = 6'5$$

Centro de gravedad = ($\bar{x}$, $\bar{y}$) = (4, 6'5)

$$s_x = \sqrt{\frac{\sum_{i=1}^{6} x_i^2}{n} - \bar{x}^2} = \sqrt{\frac{128}{6} - 4^2} = \sqrt{5'\hat{3}} = 2'31 ; s_y = \sqrt{\frac{\sum_{i=1}^{6} y_i^2}{n} - \bar{y}^2} = \sqrt{\frac{283}{6} - 6'5^2} = \sqrt{8'1\hat{6}} = 2'86$$

❄❄❄❁❁✪❁❁❄❄❄

Resuelve tú (Pág ❸❶❸)

Hallar el valor de la covarinza de la siguiente distribución, si la altura viniera expresada en metros. Comprobaras que es 100 veces menor.

Peso (kg), x:	45	55	43	47	51	60	63	58
Estatura (cm), y:	1,58	1,65	1,55	1,61	1,54	1,67	1,62	1,71

◊◊◊▣▣⊙▣▣◊◊◊

Usamos un programa de Estadística y obtenemos los resultados :

Media aritmética

X : 53.25 Y: 1.6175

Desviación típica

X : 6.456586 Y: 0.055396

Covariancia: 0.255625

❄❄❄❁❁✪❁❁❄❄❄

Resuelve tú (Pág 316)

A los alumnos del ejercicio de aplicación 4 se les preguntó también las horas que dedican a dormir diarimente. Contestaron :

7,5 8,5 8 9 7,5 8,5

calcula el coeficiente de correlación entre las horas dedicadas a ver TV y a dormir.

Elaboramos la tabla base para los cálculos :

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
12	7'5	144	56'25	90
10	8'5	100	72'25	85
15	8	225	64	120
20	9	400	81	180
28	7'5	784	56'25	210
8	8'5	64	72'25	68
$\sum x_i = 93$	$\sum y_i = 49$	$\sum x_i^2 = 1717$	$\sum y_i^2 = 402$	$\sum x_i \cdot y_i = 753$

Medias :

$$\bar{x} = \frac{\sum_{i=1}^{6} x_i}{n} = \frac{93}{6} = 15'5 \ ; \bar{y} = \frac{\sum_{i=1}^{6} y_i}{n} = \frac{49}{6} = 8'1\overset{\frown}{6}$$

Desviaciones típicas:

$$s_x = \sqrt{\frac{\sum_{i=1}^{6} x_i^2}{n} - \bar{x}^2} = \sqrt{\frac{1717}{6} - 15'5^2} = \sqrt{45'91\overset{\frown}{6}} = 6'78 \ ; s_y = \sqrt{\frac{\sum_{i=1}^{6} y_i^2}{n} - \bar{y}^2} = \sqrt{\frac{402}{6} - 8'1\overset{\frown}{6}^2} = \sqrt{0'31} = 0'55$$

Covarianza y coeficiente de correlación :

$$s_{xy} = \frac{\sum_{i=1}^{6} x_i y_i}{n} - \overline{x} \cdot \overline{y} = \frac{753}{6} - 15'5 \cdot 8'1\overset{\frown}{6} = -1'083 \qquad r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{-1'083}{6'78 \cdot 0'55} = -0'29$$

Resuelve tú (Pág ❸❶❾)

Halla la recta que mejor se ajusta, según el criterio de los mínimos cuadrados, a los datos :

x	1	3	4	5	5	7
y	2	3	5	5	7	8

◊◊◊▣▣⊙▣▣◊◊◊

Necesitamos hallar las medias, la desviación típica y la covarianza, luego elaboramos primero la tabla auxiliar :

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
1	2	1	4	2
3	3	9	9	9
4	5	16	25	20
5	5	25	25	25
5	7	25	49	35
6	8	36	64	48
$\sum x_i = 24$	$\sum y_i = 30$	$\sum x_i^2 = 112$	$\sum y_i^2 = 176$	$\sum x_i \cdot y_i = 139$

❂ **Medias :**

$$\bar{x} = \frac{\sum_{i=1}^{6} x_i}{n} = \frac{24}{6} = 4 \; ; \bar{y} = \frac{\sum_{i=1}^{6} y_i}{n} = \frac{30}{6} = 5$$

❂ **Desviación típica:**

$$s_x = \sqrt{\frac{\sum_{i=1}^{6} x_i^2}{n} - \bar{x}^2} = \sqrt{\frac{112}{6} - 4^2} = \sqrt{2'\widehat{6}} = 1'63$$

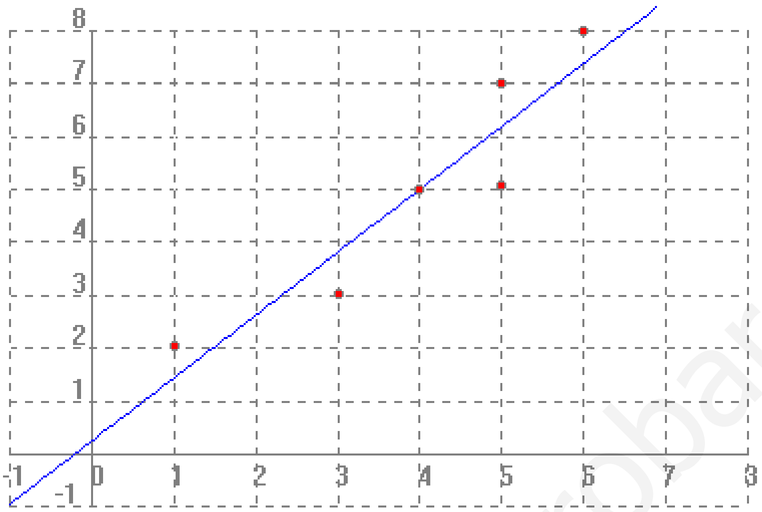
❂ **Covarianza :**

$$s_{xy} = \frac{\sum_{i=1}^{6} x_i y_i}{n} - \bar{x} \cdot \bar{y} = \frac{139}{6} - 4 \cdot 5 = 3'1\widehat{6}$$

❂ **Ecuación de la recta :**

$$y-\bar{y}=\frac{s_{xy}}{s_x^2}(x-\bar{x}) \rightarrow y-5=\frac{3'1\widehat{6}}{2'\widehat{6}}(x-4) \Leftrightarrow y-5=1'1875(x-4) \Leftrightarrow y=1'1875x+0'25$$

En el gráfico siguiente se muestra la nube de puntos y la representación de la ecuación de la recta hallada :



❄❄❄❁❁✪❁❁❄❄❄

Resuelve tú (Pág ❸❷❶)

Con los datos del Ejercicio de aplicación 4, halla la recta de regresión quc permite estimar las horas de estudio de un alumno sabiendo las hora; que dedica a ver la TV. ¿Cuánto estudiará, por término medio, ur alumno que ve la TV 14 h semanales?

◊◊◊▣▣⊙▣▣◊◊◊

Se trata de la recta de regresión de y sobre x :

$$y-\bar{y}=\frac{s_{xy}}{s_x^2}(x-\bar{x}) \rightarrow y-5'83=\frac{-16'53}{45'97}(x-15'5) \Leftrightarrow y-5'83=-0'36(x-15'5) \Leftrightarrow y=-0'36x+11'41$$

Sustituimos ahora en la ecuación obtenida la x = 14 y hallamos la y :

Si x = 14 h , y = -0’36 ·14 + 11’41 = 6’37 horas de estudio semanales.

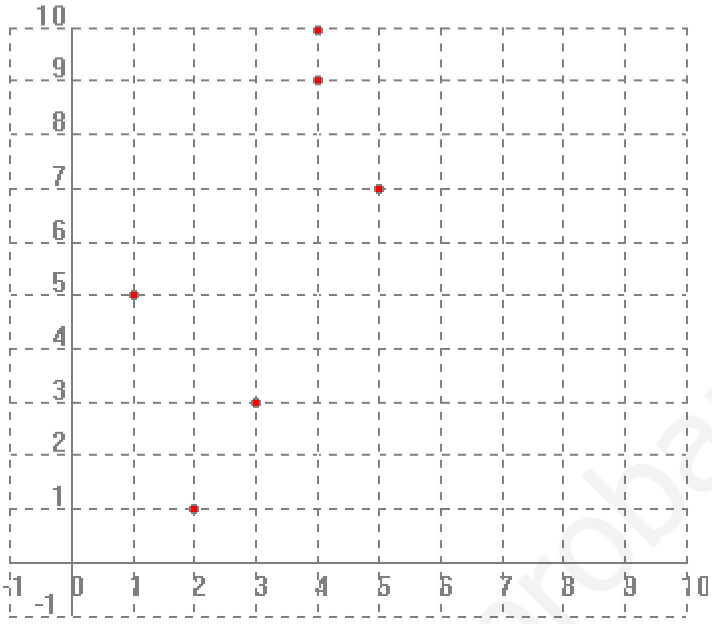
❄❄❄❁❁✪❁❁❄❄❄

PROBLEMAS PROPUESTOS

1 Representa la nube de puntos asociada a los datos:

x	3	2	5	4	1	9
y	3	1	7	10	5	4

¿Se observa correlación?



Poca correlación, hay gran dispersión.

2 Halla el centro de gravedad de la distribución anterior.

El centro de gravedad es el punto formado por las dos medias :

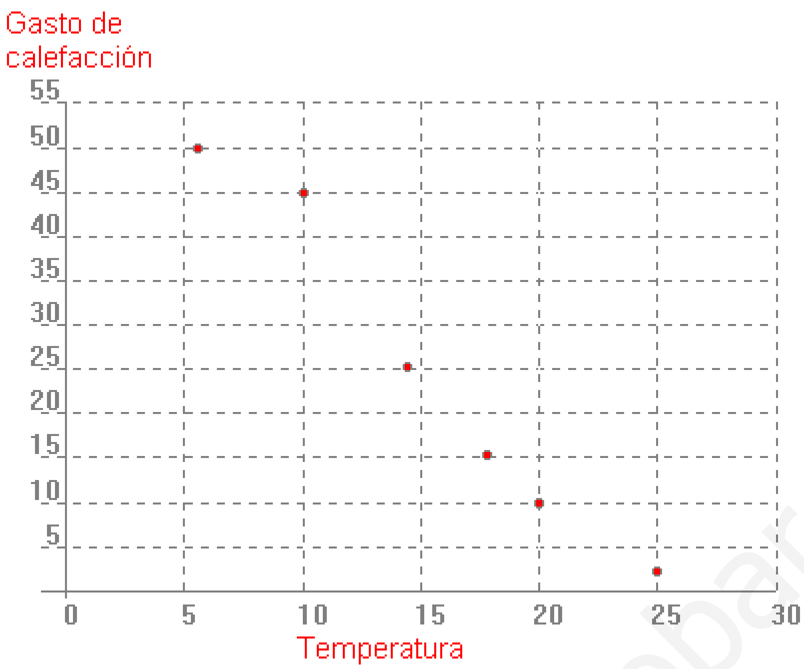
$$\bar{x} = \frac{3+2+5+4+1+9}{3} = \frac{24}{3} = 8 \qquad \bar{y} = \frac{3+1+7+10+5+4}{3} = \frac{30}{3} = 10$$

Centro de gravedad = ($\bar{x}$, $\bar{y}$) = (8, 10)

3 Representa un diagrama de dispersión en el que se observe una fuerte correlación negativa. Pon un ejemplo real que se ajuste aceptablemente bien a esa nube de puntos.

Representamos la temperatura media de varias ciudades y el gasto medio anual en calefacción por habitante (en miles de pesetas , que viene dado por la tabla :

Temperatura (ºC) x	6	10	14	18	20	25
Gasto en calefacción, y	50	45	25	15	10	2



4 Dos conjuntos de datos bidimensionales tienen coeficientes de correlación r = 0,5 y r = -0,8. Se pide:

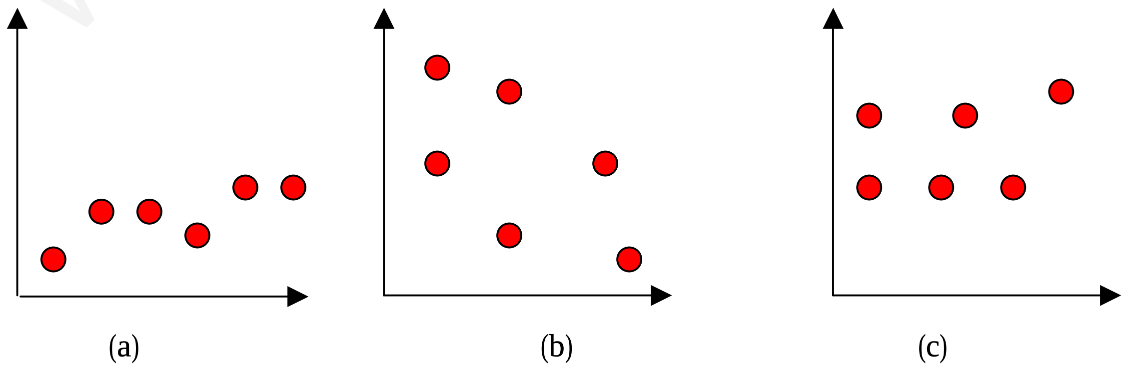
(a) ¿En cuál de los dos conjuntos se puede hacer una mejor estimación, mediante una recta, de una variable a partir de la otra?

(b) Representa dos conjuntos de puntos cuyas correlaciones se correspondan aproximadamente con los dados; sobre esos gráficos traza las rectas de regresión correspondientes.

(a) En el segundo caso en el que el coeficiente de correlación es mayor en valor absoluto

(b) Las correspondientes a los ejercicios (1) y (3) anteriores por ejemplo.

5 Asocia las rectas de regresión y=-x+6, y=x+5, y=0,5x+1 a las nubes de puntos siguientes:



y = - x + 6 que tiene pendiente negativa (con ángulo de 135º) ha de ser la **(b)** que es la única que tuiene esa pendiente.

De las dos de pendiente positiva una tiene el doble de pendiente que otra y mayor ordenada en el orígen, luego y = x + 5 ha de corresponder con la distribución **(c)** . y = 0'5x +1 es la **(a)** por exclusión.

6 Asigna los coeficientes de correlación lineal r = 0,7, r = - 0,5 y r = 0,9 a las nubes de puntos del problema anterior.

No hay duda que el valor negativo r = - 0'5 ha de ser la **(b)** que tiene correlación inversa. De los dos valores positivos el mayor r = 0'9 se corresponde con el diagrama **(a)** y r = 0'7 con el **(c).**

7 En la siguiente tabla se muestran las calificaciones de 8 alumnos en la asignatura de Física en dos pruebas:

Prueba 1	5	6	4	7	6	7	6	3
Prueba 2	7	5	4	9	8	8	9	3

(a) Representa la nube de puntos. Traza a ojo la recta de regresión. Estima su pendiente.

(b) Halla la recta de regresión. Compárala con tu estimación inicial.

(a) Ver figuar del final del apartad **(b)** . Recta estimada, en rojo.
(b) Para hallar los datos que necesitamos eleboramos al tabla auxiliar :

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
5	7	25	49	35
6	5	36	25	30
4	4	16	16	16
7	9	49	81	63
6	8	36	64	48
7	8	49	64	56
6	9	36	81	54
3	3	9	9	9
$\sum x_i = 44$	$\sum y_i = 53$	$\sum x_i^2 = 256$	$\sum y_i^2 = 389$	$\sum x_i \cdot y_i = 311$

❂ **Medias :**

$$\bar{x}=\frac{\sum_{i=1}^{6}x_i}{n}=\frac{44}{8}=5'5\;;\bar{y}=\frac{\sum_{i=1}^{6}y_i}{n}=\frac{53}{8}=6'625$$

❂ **Desviaciones típicas:**

$$s_x=\sqrt{\frac{\sum_{i=1}^{6}x_i^2}{n}-\bar{x}^2}=\sqrt{\frac{256}{8}-5'5^2}=\sqrt{1'75}=1'32\;;s_y=\sqrt{\frac{\sum_{i=1}^{6}y_i^2}{n}-\bar{y}^2}=\sqrt{\frac{389}{8}-6'625^2}=\sqrt{4'734}=2'18$$

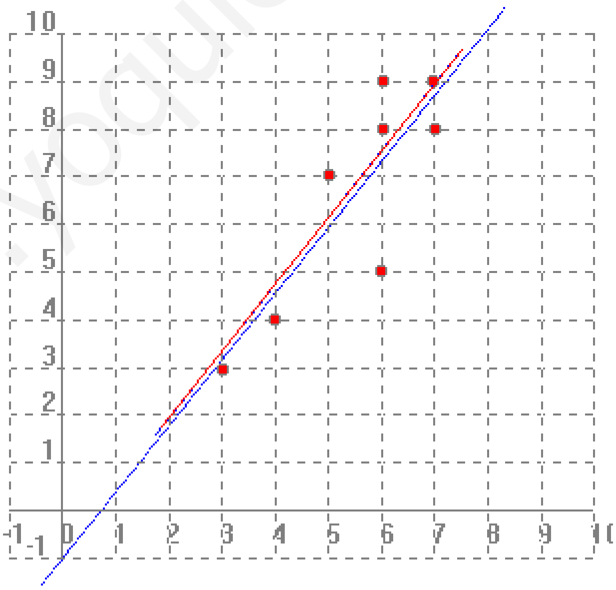
❂ **Covarianza :**

$$s_{xy}=\frac{\sum_{i=1}^{6}x_iy_i}{n}-\bar{x}\cdot\bar{y}=\frac{311}{8}-5'5\cdot 6'625=2'4375$$

❂ **Recta de regresión :**

$$y-\bar{y}=\frac{s_{xy}}{s_x^2}(x-\bar{x})\to y-6'625=\frac{2'4375}{1'75}(x-5'5)\Leftrightarrow y-6'625=1'39(x-5'5)\Leftrightarrow y=1'39x-1'02$$

Comparamos con la estimada (en rojo) la calculada (en azul) :



❈ ❈ ❈ ❁ ❁ ❂ ❁ ❁ ❈ ❈ ❈

8 Con los datos del problema anterior halla la recta de regresión de X sobre Y. ¿Qué nota habría sacado un alumno en el primer examen si obtuvo un 10 en el segundo?

◊◊◊▣▣⊙▣▣◊◊◊

❂ Recta de regresión de X sobre Y:

$$x-\bar{x}=\frac{s_{xy}}{s_y^2}(y-\bar{y}) \to x-5'5=\frac{2'4375}{4'734}(y-6'625) \Leftrightarrow x-5'5=0'515(y-6'625) \Leftrightarrow x=0'515y+2'09$$

La nota del primer examen estimada mediante la recta de regresión de X sobre Y sería :

x = 0’515 ·10 + 2’09 = 7’24

❄❄❄❁❁❂❁❁❄❄❄

9 Halla el coeficiente de correlación lineal y las rectas de regresión, de Y sobre X y de X sobre Y, del peso y la estatura de las alumnas consideradas en el Ejercicio de aplicación 3. ¿ Cuánto medirá una alumna si pesa 65 kg ? ¿ Cuánto pesará una alumna que mide 162 cm?

◊◊◊▣▣⊙▣▣◊◊◊

Peso (kg), x:	45	55	43	47	51	60	63	58
Estatura (cm), y:	1,58	1,65	1,55	1,61	1,54	1,67	1,62	1,71

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
45	1’58	2025	2’4964	71’1
55	1’65	3025	2’7225	90’75
43	1’55	1849	2’4025	66’65
47	1’61	2209	2’5921	75’67
51	1’54	2601	2’3716	78’54
60	1’67	3600	2’7889	100’2
63	1’62	3969	2’6244	102’06
58	1’71	3364	2’9241	99’18
$\sum x_i=422$	$\sum y_i=12'93$	$\sum x_i^2=22642$	$\sum y_i^2=20'9225$	$\sum x_i \cdot y_i=684'15$

❂ Medias :

$$\bar{x}=\frac{\sum_{i=1}^{6}x_i}{n}=\frac{422}{8}=52'75\,;\bar{y}=\frac{\sum_{i=1}^{6}y_i}{n}=\frac{12'93}{8}=1'62$$

❂ Desviaciones típicas:

$$s_x=\sqrt{\frac{\sum_{i=1}^{6}x_i^2}{n}-\bar{x}^2}=\sqrt{\frac{22642}{8}-52'75^2}=\sqrt{47'69}=6'91;s_y=\sqrt{\frac{\sum_{i=1}^{6}y_i^2}{n}-\bar{y}^2}=\sqrt{\frac{20'9225}{8}-1'62^2}=\sqrt{0'0031}=0'055$$

❂ Covarianza y coeficiente de correlación :

$$s_{xy}=\frac{\sum_{i=1}^{6}x_iy_i}{n}-\bar{x}\cdot\bar{y}=\frac{684'15}{8}-52'75\cdot 1'616=0'275 \qquad r=\frac{s_{xy}}{s_x\cdot s_y}=\frac{0'275}{6'91\cdot 0'055}=0'723$$

❂ **Recta de regresión de Y sobre X :**

$$y-\bar{y}=\frac{s_{xy}}{s_x^2}(x-\bar{x})\rightarrow y-1'616=\frac{0'275}{47'69}(x-52'75)\Leftrightarrow y-1'616=0'00577(x-52'75)\Leftrightarrow y=0'00577x+1'312$$

❂ **Recta de regresión de X sobre Y :**

$$x-\bar{x}=\frac{s_{xy}}{s_y^2}(y-\bar{y})\rightarrow x-52'75=\frac{0'275}{0'0031}(y-1'616)\Leftrightarrow x-52'75=88'71(y-1'616)\Leftrightarrow x=88'7y-90'60$$

❂ La alumna que pesa 65 kg medirá :

y = 0’00577·65 +1’312 = 1’69 m.

❂ La alumna que mida 1’62 m :

x = 88’7 · 1’62 – 90’6 = 53’09 kg.

❄❄❄❁❁✪❁❁❄❄❄

10 Se ha experimentado sobre la distancia de frenada de un coche dependiendo de su velocidad, obteniéndose los siguientes datos:

Velocidad (km/h) (x)	70	50	45	120	85	65
Distancia (m) (y)	32	18	19	43	35	34

Según estos datos, ¿qué distancia de seguridad habría que mantener con el vehículo que va delante si circulamos a 100 km/h?

◊◊◊▣▣⊙▣▣◊◊◊

Hay que hallar la recta de regresión de Y sobre X :

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
70	32	4900	1024	2240
50	18	2500	324	900
45	19	2025	361	855
120	43	14400	1849	5160
85	35	7225	1225	2975
65	34	4225	1156	2210
$\sum x_i=435$	$\sum y_i=181$	$\sum x_i^2=35275$	$\sum y_i^2=5939$	$\sum x_i\cdot y_i=14340$

❂ Medias :

$$\bar{x}=\frac{\sum_{i=1}^{6}x_i}{n}=\frac{435}{6}=72'5\;;\;\bar{y}=\frac{\sum_{i=1}^{6}y_i}{n}=\frac{181}{6}=30'1\widehat{6}$$

❂ Desviación típica :

$$s_x=\sqrt{\frac{\sum_{i=1}^{6}x_i^2}{n}-\bar{x}^2}=\sqrt{\frac{35275}{6}-72'5^2}=\sqrt{622'91\widehat{6}}=24'9583$$

❂ Covarianza :

$$s_{xy}=\frac{\sum_{i=1}^{6}x_iy_i}{n}-\bar{x}\cdot\bar{y}=\frac{14340}{6}-72'5\cdot 30'1\widehat{6}=202'92$$

❂ Recta de regresión de Y sobre X :

$$y-\bar{y}=\frac{s_{xy}}{s_x^2}(x-\bar{x})\rightarrow y-30'1\widehat{6}=\frac{202'92}{622'91\widehat{6}}(x-72'5)\Leftrightarrow y-30'1\widehat{6}=0'3257(x-72'5)\Leftrightarrow y=0'325x+6'549$$

La distancia de frenado esperada será :

y = 0’325 · 100 + 6’549 = 32’5 + 6’549 = 39’05 m

❄❄❄❁❁❂❁❁❄❄❄

11 A un grupo de 12 alumnos de ESO se le han realizado pruebas para determinar su: nivel de vocabulario (NV); nivel intelectual (NI); velocidad lectora (VL); agilidad de cálculo (AC). Los resultados han sido los siguientes:

NV	28	27	14	17	18	14	23	24	14	12	16	10
NI	43	30	18	21	24	20	23	19	22	14	10	18
VL	69	68	38	37	48	50	50	57	33	17	42	35
AC	28	22	15	19	20	14	25	19	20	9	7	10

Halla la correlación entre NV y NI. Interpreta el resultado

◊◊◊▣▣⊙▣▣◊◊◊

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
28	43	784	1849	1204
27	30	729	900	810
14	18	196	324	252
17	21	289	441	357
18	24	324	576	432
14	20	196	400	280
23	23	529	529	529
24	19	576	361	456
14	22	196	484	308
12	14	144	196	168
16	10	256	100	160
10	18	100	324	180
$\sum x_i = 217$	$\sum y_i = 262$	$\sum x_i^2 = 4319$	$\sum y_i^2 = 6484$	$\sum x_i \cdot y_i = 5136$

Medias :

$$\bar{x} = \frac{\sum_{i=1}^{6} x_i}{n} = \frac{217}{12} = 18'08\widehat{3} \; ; \; \bar{y} = \frac{\sum_{i=1}^{6} y_i}{n} = \frac{262}{12} = 21'8\widehat{3}$$

Desviaciones típicas:

$$s_x = \sqrt{\frac{\sum_{i=1}^{6} x_i^2}{n} - \bar{x}^2} = \sqrt{\frac{4319}{12} - 18'08\widehat{3}^2} = \sqrt{32'91} = 5'737 \; ; \; s_y = \sqrt{\frac{\sum_{i=1}^{6} y_i^2}{n} - \bar{y}^2} = \sqrt{\frac{6484}{12} - 21'8\widehat{3}^2} = \sqrt{63'64} = 7'977$$

Covarianza y coeficiente de correlación :

$$s_{xy} = \frac{\sum_{i=1}^{6} x_i y_i}{n} - \bar{x} \cdot \bar{y} = \frac{5136}{12} - 18'08\widehat{3} \cdot 21'8\widehat{3} = 33'1806 \qquad r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{33'18056}{5'737 \cdot 7'977} = 0'725$$

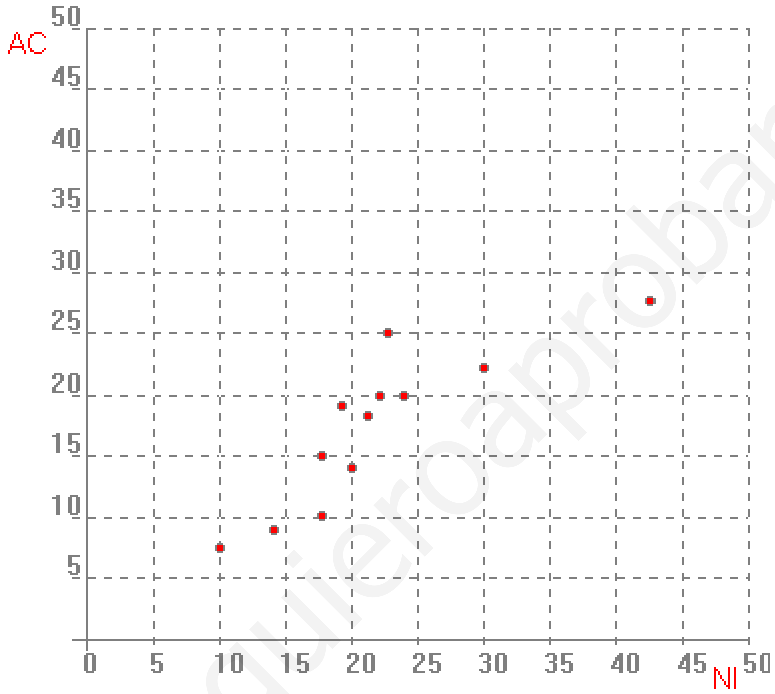
12 Con los datos del problema anterior:

(a) Representa la nube de puntos asociada a las variables NI (eje x) y AC (eje y).

(b) Halla la recta de regresión de la agilidad de cálculo sobre el nivel intelectual. ¿Qué AC cabe esperar para un alumno con NI 35?

◊◊◊▣▣⊙▣▣◊◊◊

(a)



(b)

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
43	28	1849	784	1204
30	22	900	484	660
18	15	324	225	270
21	19	441	361	399
24	20	576	400	480
20	14	400	196	280
23	25	529	625	575
19	19	361	361	361
22	20	484	400	440
14	9	196	81	126
10	7	100	49	70
18	10	324	100	180
$\sum x_i = 262$	$\sum y_i = 208$	$\sum x_i^2 = 6484$	$\sum y_i^2 = 4066$	$\sum x_i \cdot y_i = 5045$

Medias :

$$\bar{x}=\frac{\sum_{i=1}^{6}x_i}{n}=\frac{262}{12}=21{,}8\widehat{3}\;;\;\bar{y}=\frac{\sum_{i=1}^{6}y_i}{n}=\frac{208}{12}=17'\widehat{3}$$

Desviaciones típicas:

$$s_x=\sqrt{\frac{\sum_{i=1}^{6}x_i^2}{n}-\bar{x}^2}=\sqrt{\frac{6484}{12}-21'8\widehat{3}^2}=\sqrt{63'64}=7'98\;;\;s_y=\sqrt{\frac{\sum_{i=1}^{6}y_i^2}{n}-\bar{y}^2}=\sqrt{\frac{4066}{12}-17'\widehat{3}^2}=\sqrt{39'39}=6'2$$

Covarianza y coeficiente de correlación :

$$s_{xy}=\frac{\sum_{i=1}^{6}x_iy_i}{n}-\bar{x}\cdot\bar{y}=\frac{5045}{12}-21'8\widehat{3}\cdot17'\widehat{3}=41'97\qquad r=\frac{s_{xy}}{s_x\cdot s_y}=\frac{41{,}97}{6'2\cdot7'98}=0{,}85$$

Se nos pide la recta de regresión de x (AC) sobre y (NI):

Recta de regresión de Y sobre X :

$$y-\bar{y}=\frac{s_{xy}}{s_x^2}(x-\bar{x})\rightarrow y-17'\widehat{3}=\frac{41'97}{63{,}64}(x-21'8\widehat{3})\Leftrightarrow y-17'\widehat{3}=0{,}66(x-21'8\widehat{3})\Leftrightarrow y=0{,}66x+2{,}92$$

Para x = 35, y =0,66 · 35 + 2,92 = 26 de AC.

13 Para los datos del problema anterior halla la correlación entre velocidad lectora y agilidad de cálculo.

La velocidad lectora (VL) = x

La agilidad de cálculo (AC) = y

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
69	28	4761	784	1204
68	22	4624	484	660
38	15	1444	225	270
37	19	1369	361	399
48	20	2304	400	480
50	14	2500	196	280
50	25	2500	625	575
57	19	3249	361	361
33	20	1089	400	440
17	9	289	81	126
42	7	1764	49	70
35	10	1225	100	180
$\sum x_i = 544$	$\sum y_i = 208$	$\sum x_i^2 = 27188$	$\sum y_i^2 = 4066$	$\sum x_i \cdot y_i = 10151$

Medias :

$$\overline{x} = \frac{\sum_{i=1}^{6} x_i}{n} = \frac{544}{12} = 45'\widehat{3}\,;\ \overline{y} = \frac{\sum_{i=1}^{6} y_i}{n} = \frac{208}{12} = 17'\widehat{3}$$

Desviaciones típicas:

$$s_x = \sqrt{\frac{\sum_{i=1}^{6} x_i^2}{n} - \overline{x}^2} = \sqrt{\frac{27188}{12} - 45'\widehat{3}^2} = \sqrt{210'6} = 14'51\,;\ s_y = \sqrt{\frac{\sum_{i=1}^{6} y_i^2}{n} - \overline{y}^2} = \sqrt{\frac{4066}{12} - 17'\widehat{3}^2} = \sqrt{38'39} = 6'2$$

Covarianza y coeficiente de correlación :

$$s_{xy} = \frac{\sum_{i=1}^{6} x_i y_i}{n} - \overline{x}\cdot\overline{y} = \frac{10151}{12} - 45'\widehat{3}\cdot 17'\widehat{3} = 60{,}13 \qquad r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{60'13}{14'51 \cdot 6'2} = 0'67$$

14 Calcula el coeficiente de correlación lineal entre el número de profesores y alumnos de las cinco universidades españolas que se indican:

Universidad	Alumnos	Profesores
Alcalá de Henares	16 235	966
Carlos III	4 103	400
Extremadura	19 174	757
La Laguna	21 143	1 620
Murcia	29 389	1 391

Fuente: Consejo de Universidades. Datos de 1992.

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
16235	966	26357225	933156	15683010
4103	400	16834609	160000	1641200
19174	757	367642276	573049	14514718
21143	1620	447026449	2624400	34251660
29389	1391	863713321	1934881	40880099
$\sum x_i = 90044$	$\sum y_i = 5134$	$\sum x_i^2 = 1958791880$	$\sum y_i^2 = 6225486$	$\sum x_i \cdot y_i = 106970687$

Medias :

$$\bar{x} = \frac{\sum_{i=1}^{6} x_i}{n} = \frac{90044}{5} = 18008'8 \; ; \bar{y} = \frac{\sum_{i=1}^{6} y_i}{n} = \frac{5134}{5} = 1026'8$$

Desviaciones típicas:

$$s_x = \sqrt{\frac{\sum_{i=1}^{6} x_i^2}{n} - \bar{x}^2} = \sqrt{\frac{1958791880}{5} - 18008{,}8^2} = \sqrt{67441498{,}56} = 8212{,}28 \; ; s_y = \sqrt{\frac{\sum_{i=1}^{6} y_i^2}{n} - \bar{y}^2} = \sqrt{\frac{6225486}{5} - 1026{,}8^2} = \sqrt{190778{,}96} = 436'8$$

Covarianza y coeficiente de correlación :

$$s_{xy} = \frac{\sum_{i=1}^{6} x_i y_i}{n} - \overline{x \cdot y} = \frac{106970687}{5} - 18008{,}8 \cdot 1026'8 = 2902701{,}56 \qquad r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{2902701'56}{8212'281 \cdot 436'8} = 0'81$$

15 Para el diagnóstico de una cierta enfermedad es necesario saber la concentración de la sustancia A en líquido céfalo-raquídeo (LCR), cuya extracción es más molesta y costosa que la extracción de suero. Para un grupo de 6 individuos se midió la concentración de la sustancia A en LCR y en suero, obteniéndose:

Concentración de A en suero : x	11	12	15	8	7	0
Concentración de A en LCR : y	15	21	24	11	7	3

(a) ¿Puede calcularse la concentración de A en LCR a partir de la obtenida en suero?

(b) ¿Qué ecuación lineal permite hacerlo?

(c) Representa el diagrama de dispersión y la recta de regresión obtenida.

◊◊◊▣▣⊙▣▣◊◊◊

(a) Veamos qué correlación existe entre las variables:

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
11	15	121	225	165
12	21	144	441	252
15	24	225	576	360
8	11	64	121	88
7	7	49	49	49
0	3	0	9	0
$\sum x_i = 53$	$\sum y_i = 81$	$\sum x_i^2 = 603$	$\sum y_i^2 = 1421$	$\sum x_i \cdot y_i = 914$

❂ **Medias :**

$$\bar{x} = \frac{\sum_{i=1}^{6} x_i}{n} = \frac{53}{6} = 8'8\widehat{3}\ ;\ \bar{y} = \frac{\sum_{i=1}^{6} y_i}{n} = \frac{81}{6} = 13'5$$

❂ **Desviaciones típicas:**

$$s_x = \sqrt{\frac{\sum_{i=1}^{6} x_i^2}{n} - \bar{x}^2} = \sqrt{\frac{603}{6} - 8{,}8\widehat{3}^2} = \sqrt{22{,}47} = 4{,}74\ ;\ s_y = \sqrt{\frac{\sum_{i=1}^{6} y_i^2}{n} - \bar{y}^2} = \sqrt{\frac{1421}{6} - 13{,}5^2} = \sqrt{54{,}58} = 7{,}39$$

❂ **Covarianza y coeficiente de correlación :**

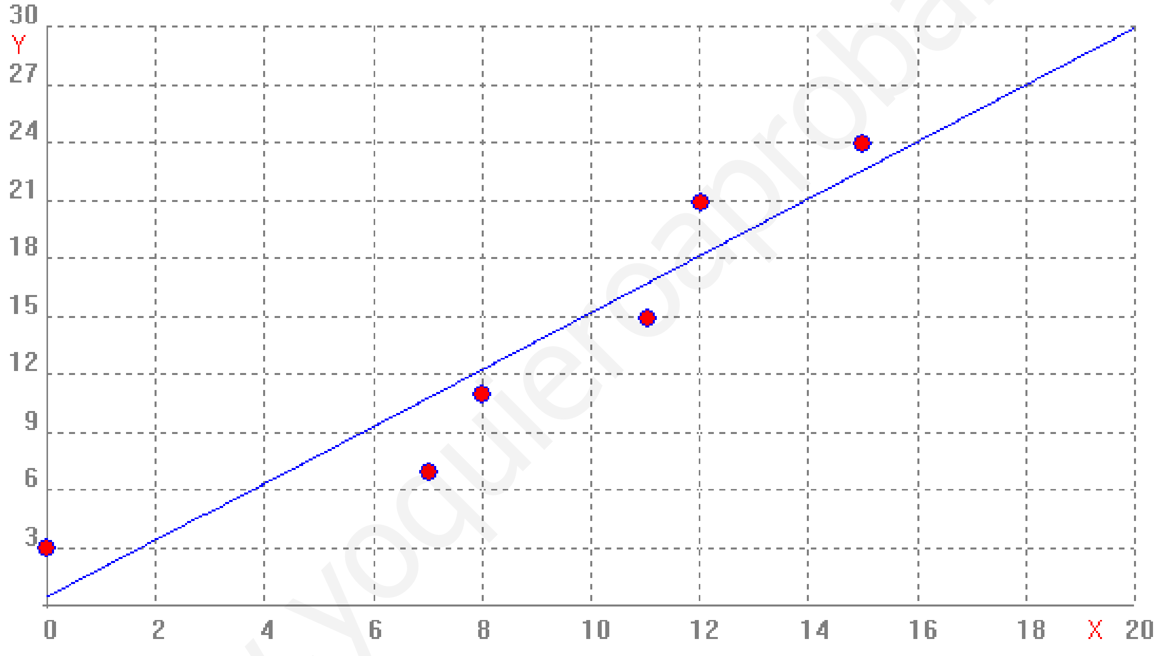
$$s_{xy} = \frac{\sum_{i=1}^{6} x_i y_i}{n} - \bar{x}\cdot\bar{y} = \frac{914}{6} - 8{,}8\bar{3}\cdot 13'5 = 33{,}08 \qquad r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{33{,}08}{4'74 \cdot 7'39} = 0'94$$

Como la correlación es alta, sí podemos calcular la concentración en LCR a partir de la suero.

(b) Para calcularla utilizamos la ecuación de la recta de regresión de y sobre x:

Recta de regresión de Y sobre X :

$$y-\bar{y}=\frac{s_{xy}}{s_x^2}(x-\bar{x}) \to y-13'53=\frac{33'08}{22,47}(x-8'8\widehat{3}) \Leftrightarrow y-13'53=1{,}47(x-8'8\widehat{3}) \Leftrightarrow y=1{,}47x+0{,}55$$



❄❄❄❁❁✪❁❁❄❄❄

16 En una muestra de 7(8) individuos se midió la concentración de dos sustancias X e Y en plasma sanguíneo:

x	11	13	17	19	14	12	16	18
y	12	14	20	21	17	12	17	12

Halla la recta que permita estimar la concentración de X conociendo la de Y. ¿Qué valor de x estimamos para un individuo con concentración 15 de la sustancia y.

◊◊◊▣▣⊙▣▣◊◊◊

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
11	12	121	144	132
13	14	169	196	182
17	20	289	400	340
19	21	361	441	399
14	17	196	289	238
12	12	144	144	144
16	17	256	289	272
18	12	324	144	216
$\sum x_i = 120$	$\sum y_i = 125$	$\sum x_i^2 = 1860$	$\sum y_i^2 = 2047$	$\sum x_i \cdot y_i = 1923$

Medias :

$$\bar{x} = \frac{\sum_{i=1}^{6} x_i}{n} = \frac{120}{8} = 15\ ;\ \bar{y} = \frac{\sum_{i=1}^{6} y_i}{n} = \frac{125}{8} = 15{,}625$$

Desviaciones típicas:

$$s_x = \sqrt{\frac{\sum_{i=1}^{6} x_i^2}{n} - \bar{x}^2} = \sqrt{\frac{1860}{8} - 15^2} = \sqrt{7{,}5} = 2{,}74\ ;\ s_y = \sqrt{\frac{\sum_{i=1}^{6} y_i^2}{n} - \bar{y}^2} = \sqrt{\frac{2047}{8} - 15{,}625^2} = \sqrt{11{,}73} = 3{,}43$$

Covarianza :

$$s_{xy} = \frac{\sum_{i=1}^{6} x_i y_i}{n} - \bar{x} \cdot \bar{y} = \frac{1923}{8} - 15 \cdot 15{,}625 = 6$$

Recta de regresión de X sobre Y :

$$x - \bar{x} = \frac{s_{xy}}{s_y^2}(y - \bar{y}) \rightarrow x - 15 = \frac{6}{11{,}73}(y - 15'625) \Leftrightarrow x - 15 = 0{,}51(y - 15'625) \Leftrightarrow x = 0{,}51x + 7{,}03$$

Si y = 15, sustituyendo en la ecuación de regresión anterior tenemos:

$$x = 0{,}51 \cdot 15 + 7{,}03 = 14{,}68$$

17 Para una variable bidimensional se conoce r = - 0,5; s_x = 2; s_y = 3. Razona si alguna de las siguientes rectas de regresión de Y sobre X corresponde a estos datos:

(a) y= -x+2

(b) y = 0,5x - 1

(c) 3x+4y-4=0

◊◊◊▣▣☉▣▣◊◊◊

Como:

$r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{s_{xy}}{2 \cdot 3} = -0{,}5 \Rightarrow s_{xy} = -3$, la pendiente es $a = \frac{s_{xy}}{s_x^2} = \frac{-3}{4}$ que se corresponde con la pendiente de la recta **(c)** y = -(3/4) x + 1

❄❄❄❁❁✪❁❁❄❄❄

18 Considera los datos del Ejercicio de aplicación 6 y la recta de regresión hallada. ¿Qué cosecha cabría esperar si al año siguiente repetimos el experimento en idénticas condiciones; esto es, si las mismas parcelas son tratadas con los mismos kilogramos de fertilizante? ¿Cómo interpretas los cambios respecto al año anterior?

◊◊◊▣▣☉▣▣◊◊◊

La cosecha sería similar pero distinta debido a que habrá factores que cambian como la temperatura, pluviosidad, agotamiento del terreno, etc.

❄❄❄❁❁✪❁❁❄❄❄

19 El número de parejas de cigüeñas y de pollos habidos en Alcalá de Henares en los años que se indican, vienen dados en la siguiente tabla:

Año	1988	1989	1990	1991	1992	1993
Parejas	12	22	25	27	27	43
Pollos	32	35	54	42	49	73

Fuente: Juan Prieto Martín.

(a) Calcula los coeficientes de correlación entre: año-parejas; año-pollos; parejas-pollos.

(b) Si en 1994 hubo 45 parejas de cigüeñas, ¿cuántos pollos deben esperarse?

◊◊◊▣▣☉▣▣◊◊◊

(a)

Año - parejas

Año (x)	1988	1989	1990	1991	1992	1993
Parejas(y)	12	22	25	27	27	43

Media Aritmética X = 1990,5
Media Aritmética Y = 26
Desviación Típica X = 1,707825
Desviación Típica Y = 9,165151
Covariancia = 14,333333

$$r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{14,3}{1,71 \cdot 9,17} = 0,91$$

Año - pollos

Año (x)	1988	1989	1990	1991	1992	1993
Pollos (y)	32	35	54	42	49	73

Media Aritmética X = 1990,5
Media Aritmética Y = 47,5
Desviación Típica X = 1,707825
Desviación Típica Y = 13,671747
Covariancia = 19,583333

$$r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{19,58}{1,71 \cdot 13,67} = 0,84$$

Parejas – pollos

Parejas	12	22	25	27	27	43
Pollos	32	35	54	42	49	73

Media Aritmética X = 26
Media Aritmética Y = 47,5
Desviación Típica X = 9,165151
Desviación Típica Y = 13,671747
Covariancia = 115

$$r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{115}{9,165 \cdot 13,67} = 0,918$$

(b) Hallamos la recta de regresión de y sobre x en el último caso:

❂ Recta de regresión de Y sobre X :

$$y - \bar{y} = \frac{s_{xy}}{s_x^2}(x - \bar{x}) \to y - 47'5 = \frac{115}{84}(x - 26) \Leftrightarrow y - 47'5 = 1,37(x - 26) \Leftrightarrow y = 1,37x + 11,88$$

Luego si x = 45 parejas y = 1,37 · 45 + 11,88 = 73,53 = 74 pollos.

❉❉❉❁❁✪❁❁❉❉❉

20 La temperatura media de varias ciudades y el gasto medio anual en calefacción por habitante (en miles de pesetas) fue:

T. (ºC) (X)	6	10	14	18	20	25
Gasto (Y)	50	45	25	15	10	2

Utilizando la calculadora halla la recta de regresión de Y (gasto en calefacción) sobre X (temperatura media de la ciudad). ¿Qué gasto cabe esperar en ciudades con temperaturas media de 5, 15 y 30 °C? Comenta los resultados.

◊◊◊▣▣⊙▣▣◊◊◊

Media Aritmética X = 15,5
Media Aritmética Y = 24,5
Desviación Típica X = 6,317964
Desviación Típica Y = 17,689451
Covariancia = -109,75

✪ **Recta de regresión de Y sobre X :**

$$y-\bar{y}=\frac{s_{xy}}{s_x^2}(x-\bar{x}) \rightarrow y-24{,}5=\frac{-109{,}75}{40}(x-15{,}5) \Leftrightarrow y-24'5=-2{,}74(x-15{,}5) \Leftrightarrow y=-2{,}74x+67$$

Luego si:

x = 5 ºC, y = -2,74 · 5 + 67 = 53 miles de pesetas.

x = 15 ºC, y = -2,74 · 15 + 67 = 26 miles de pesetas.

x = 30 ºC, y = -2,74 · 30 + 67 = -15,2 miles de pesetas, lo que significaría que la compañía eléctrica nos daría ese dinero, IMPOSIBLE, ¿verdad?, no se puede extrapolar a esa temperatura, no pondríamos la calefacción y el gasto sería nulo.

❉❉❉❁❁✪❁❁❉❉❉

21 Para España, el número de trabajadores agrícolas (en millones), para los años que se indican, fue:

Año	1955	1960	1965	1970	1975	1980	1985	1990
Trabajadores	2,56	2,45	2,20	2	1,85	1,60	1,25	1,10

(a) Deduce el tipo de dependencia entre las variables.
(b) Estima el número de trabajadores agrícolas para el año 2000.

(c) ¿Qué ocurre si intentamos estimar el número de trabajadores agrícolas para el año 2050?; ¿cómo explicas el error de dicha predicción?

◈◈◈▣▣⊙▣▣◈◈◈

(a)

Hallamos la correlación:
Media Aritmética X = 1972,5
Media Aritmética Y = 1,87625
Desviación Típica X = 11,456439
Desviación Típica Y = 0,498571
Covariancia = -5,678125

$$r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{-5{,}678125}{11{,}456439 \cdot 0{,}498571} = -0{,}994$$

Vemos que la correlación es muy fuerte pero negativa, con el paso de los años se ha ido reduciendo el número de trabajadores agrícolas por emigración a las grandes ciudades.

(b) Hallamos la recta de regresión de y sobre x:

Recta de regresión de Y sobre X :

$$y - \bar{y} = \frac{s_{xy}}{s_x^2}(x - \bar{x}) \to y - 1'88 = \frac{-5{,}68}{131{,}25}(x - 1972{,}5) \Leftrightarrow y - 1'88 = -0{,}043(x - 1972{,}5) \Leftrightarrow y = -0{,}043x + 86{,}7$$

Para x = 2000, y = - 0,043 · 2000 + 86,7 = 0,7 millones de trabajadores habrá si se mantiene la tendencia.

(c) Que da un valor negativo y = - 0,043 · 2050 + 86,7 = - 1, 45, ya que la función se anula en x = 86,7/0,043 = 2016, 27, es decir en el 2 017, lo cual no es cierto, pues habrá campesinos, pocos si sigue la tendencia pero algunos habría.

❄❄❄❁❁✪❁❁❄❄❄

22 En la siguiente tabla se muestran las temperaturas máximas y mínimas (en °C) y la precipitación (en mm) tomadas en el Observatorio Universitario de Sierra Nevada, para el período 1986-1987.

	Temperatura		Precipitación
	Máxima	Mínima	
Diciembre	8,2	-15,5	152,4
Enero	7,9	-12,8	117,7
Febrero	8,5	-13,2	110,6
Marzo	9,3	-12,7	115,5
Abril	10,6	-11,6	78,4
Mayo	15,3	-3,0	74,7
Junio	20,4	-1,2	31,8
Julio	23,1	4,0	3,4
Agosto	23,2	2,9	5,3
Septiembre	18,7	-3,5	61,7
Octubre	15,2	-7,5	105,4
Noviembre	9,9	-10,3	124,4

(a) Representa las nubes de puntos correspondientes a:

1) las temperaturas máximas y mínimas;

2) La temperatura máxima y la precipitación.

(b) A la vista de esas nubes determinar si existe correlación entre las variables.

(c) ¿Qué signo tienen?

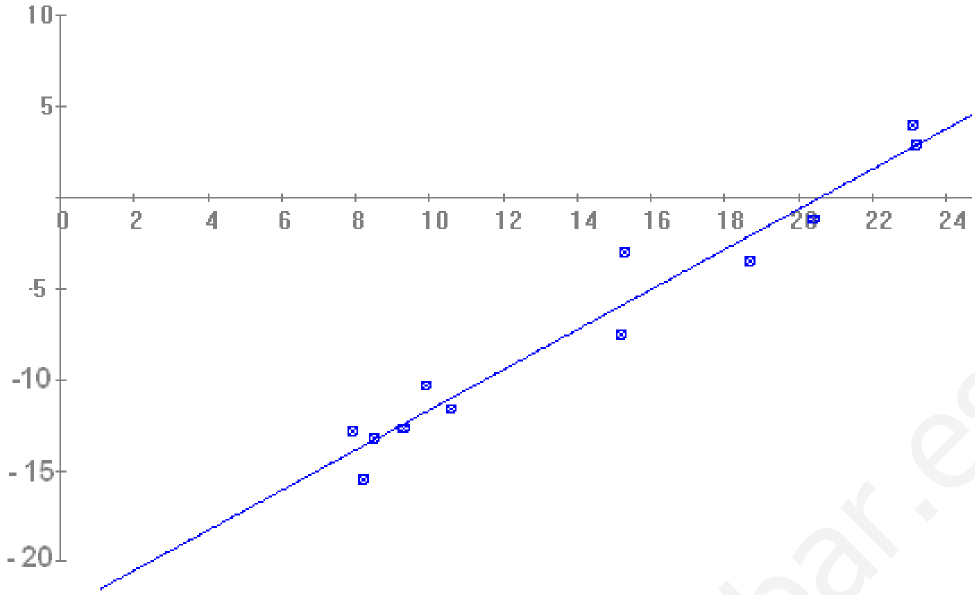
(d) Estima, para cada caso, su valor aproximado.

◊◊◊▣▣⊙▣▣◊◊◊

(a)

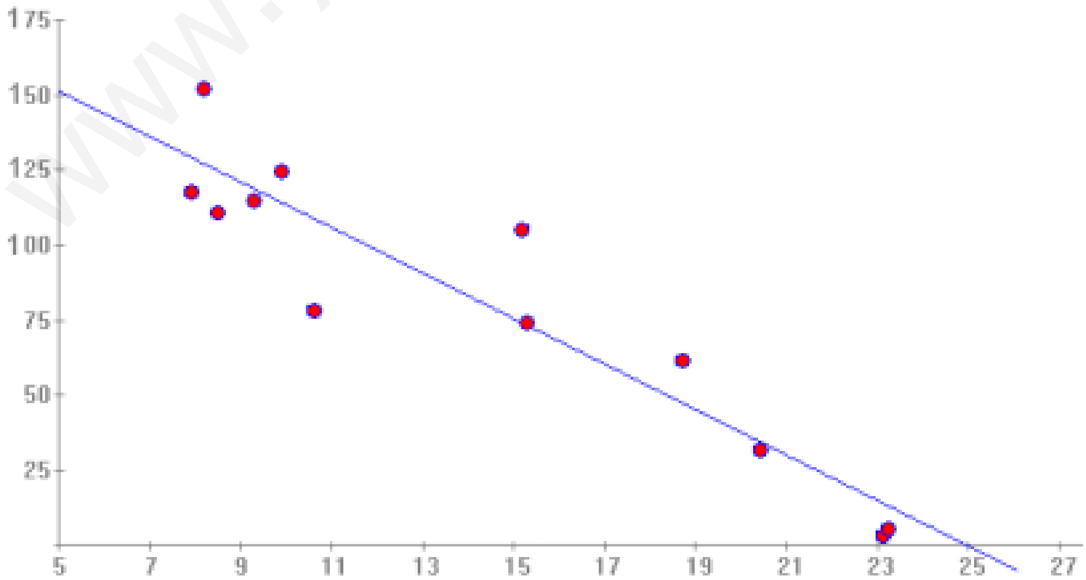
1)

	Temperatura	
	Máxima	Mínima
Diciembre	8,2	-15,5
Enero	7,9	-12,8
Febrero	8,5	-13,2
Marzo	9,3	-12,7
Abril	10,6	-11,6
Mayo	15,3	-3,0
Junio	20,4	-1,2
Julio	23,1	4,0
Agosto	23,2	2,9
Septiembre	18,7	-3,5
Octubre	15,2	-7,5
Noviembre	9,9	-10,3



2)

	Máxima	Precipitación
Diciembre	8,2	152,4
Enero	7,9	117,7
Febrero	8,5	110,6
Marzo	9,3	115,5
Abril	10,6	78,4
Mayo	15,3	74,7
Junio	20,4	31,8
Julio	23,1	3,4
Agosto	23,2	5,3
Septiembre	18,7	61,7
Octubre	15,2	105,4
Noviembre	9,9	124,4



(b) Sí existe una correlación negativa en ambos casos, pero más fuerte entre las temperaturas.

(c) Correlación de signo negativo ya que al aumentar las temperaturas máximas, disminuyen las mínimas y la pluviosidad.

(d) El coeficiente de correlación puede ser r = - 0,95 en el primer caso y r = - 0,90 en el segundo.

AUTOEVALUACIÓN

1 ¿Qué significa que dos variables estén correlacionadas? Existe correlación entre los kilovatios gastados y la factura eléctrica que paga una familia? ¿Y entre su consumo eléctrico y la temperatura media diaria?

- Si están correlacionadas, existe alguna relación entre ambas.
- Es una relación funcional no hay correlación estadística sino funcional, es una función lineal del tipo: Facturación = término fijo(potencia contratada, etc) + precio/kw · Nº de Kw.
- Sí existe una correlación estadística inversa pues cuando la temperatura media aumenta el consumo de calefacción eléctrica y agua caliente disminuye, ya que se necesita menos calor.

2 Da un ejemplo de dos variables que estén correlacionadas directamente. Haz una representación aproximada de la nube de puntos; traza la recta de regresión.

Véanse los ejercicios Nº 12 y Nº 15, por ejemplo.

3 ¿Qué es el centro medio de una distribución bidimensional? ¿Tiene alguna aplicación? ¿Recuerdas cómo se calcula?

- Es el punto formado por la medias de las dos variables $(\bar{x}, \bar{y})$.
- Sí siempre que se requiera saber el centro de una distribución, por ejemplo en distribuciones logísticas entre distintos puntos de la geografía.
- Para su cálculo usamos las fórmulas : $\bar{x} = \frac{\sum_{i=1}^{6} x_i}{n}$; $\bar{y} = \frac{\sum_{i=1}^{6} y_i}{n}$

4 ¿Qué es el coeficiente de correlación lineal r? ¿Qué valores puede tomar? ¿Qué deducimos si r es positivo?

❑ Es un número que mide la mayor o menor correlación entre dos variables estadísticas.

❑ Puede tomar valores entre – 1y 1, $-1 \leq r \leq 1$.

❑ Si $r > 0$, la correlación entre las variables es positiva de manera que si una aumenta también lo hace la otra y al contrario.

5 ¿Para qué sirve la recta de regresión? La bondad de una estimación mediante la recta de regresión, ¿depende de la pendiente de la recta? Si no es así, ¿de qué depende?

❑ La recta de regresión sirve para medir la regresión de las variables y poder hacer estimaciones de una variable a partir de valores de la otra.

❑ La bondad del ajuste, no depende de la pendiente sino del coeficiente de correlación (r) y del número de puntos (directamente) que se han tomado para su cálculo.

6 Una de las siguientes afirmaciones sobre el coeficiente de correlación lineal **r** es falsa. Indícala.

(a) Si r = **-** 0,9, la correlación lineal es débil.
(b) Si r = 0, no hay correlación lineal.
(c) Si r = 0,85, la correlación es directa y fuerte.
(d) Si r = 0,2, la correlación es débil.

Es falsa la **(a)** ya que si r = - 0,9 la correlación es fuerte (negativa pero fuerte).

7 ¿Cuál de las siguientes relaciones estadísticas presentan correlación negativa?

(a) La temperatura media diaria y el consumo de refrescos.

(b) El número de médicos de un país y la tasa de mortalidad infantil de ese mismo país.

(c) La velocidad de un automóvil y su consumo de gasolina.

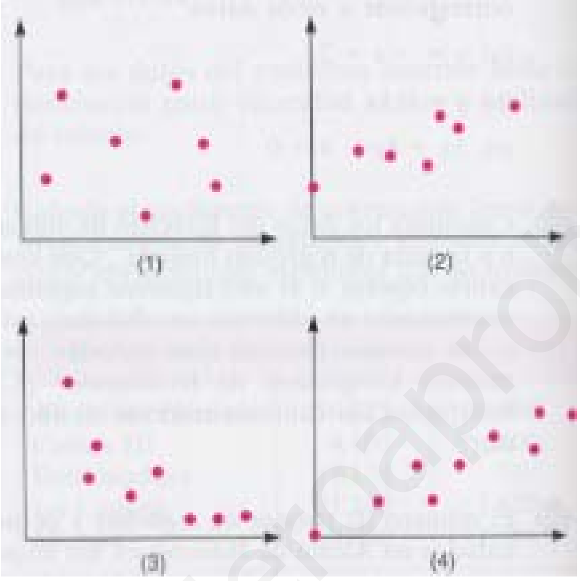
(d) El rendimiento académico y la estatura.

◊◊◊▣▣⊙▣▣◊◊◊

Si la temperatura aumenta suele aumentar el consumo de refrescos (positiva), si se pisa el acelerador y la velocidad aumenta el consumo aumenta (positiva), el rendimiento académico y la estatura no tiene correlación y si el número de médicos aumenta debe disminuir l mortalidad infantil que es la negativa.

❈❈❈❁❁✪❁❁❈❈❈

8 Asocia a las variables de la pregunta anterior, los diagramas:



◊◊◊▣▣⊙▣▣◊◊◊

(a) ⇆ (2) **(b)** ⇆ (3) **(c)** ⇆ (4) (no hay duda, pues si v = 0 el consumo puede ser nulo, si el automóvil permanece sin arrancar, mientras que el consumo de refrescos aunque baje no llega a ser nulo en ninguna época) **(d)** ⇆ (1)

❈❈❈❁❁✪❁❁❈❈❈

9 ¿Qué coeficiente de correlación asignarías a cada una de las nubles de puntos de la pregunta 7?

(a) r = 0,8
(b) r = -0,1
(c) r = - 0,65
(d) r = 0,8

◊◊◊▣▣⊙▣▣◊◊◊

(a) r = 0,8, (b) r = - 0,65, (c) r = 0,8 y (d) r = - 0,1

❈❈❈❁❁✪❁❁❈❈❈

10 Una de las siguientes afirmaciones sobre la recta de regresión, $y = ax + b$, es falsa. Indícala.

(a) Si $a = 0$, entonces no hay correlación entre x e y.
(b) El signo de a indica el sentido de la correlación.
(c) El valor de a, como el de r, no puede ser mayor que 1.
(d) La recta de regresión siempre pasa por el centro medio.

(a) Para que $a = 0$, ha de ser 0 la covarianza, y es cierto que no hay correlación la y toma siempre el mismo valor $y = b$ independiente de los valores de x.

(b) Si es cierto pues si $a < 0$ $r < 0$ y la covarianza (que da el signo a la pendiente y a r, ya que las varianzas y desviaciones típicas han de ser siempre positivas por definición) también es negativa.

(c) Esta es la falsa ya que aunque $-1 \le r \le 1$, la pendiente $-\infty < a < +\infty$.

(d) ya hemos dicho que el centro de la distribución es el punto formado por los puntos medios, sustituyéndolo en la ecuación de la recta de regresión se cumple pues 0 = 0.