UNIDAD 16: Distribuciones bidimensionales. Correlación y regresión

ACTIVIDADES-PÁG. 374

1. La media y la desviación típica son: $\bar{x} = 1{,}866 \;\; y \;\; \sigma = 0{,}065.$

Los jugadores que se encuentran por encima de $\bar{x} + \sigma$ = 1,866 + 0,065 = 1,931 son 5 del intervalo [1,90; 1,95) y 2 del intervalo [1,95; 2,00); en total 7.

2. La media y la desviación típica son $\bar{x} = 105 \;\; y \;\; \sigma = 23{,}95$.

El intervalo buscado es:

$$\left(\bar{x} - \sigma, \bar{x} + \sigma\right) = (105 - 23{,}95;\ 105 + 23{,}95) = (81{,}05;\ 128{,}95).$$

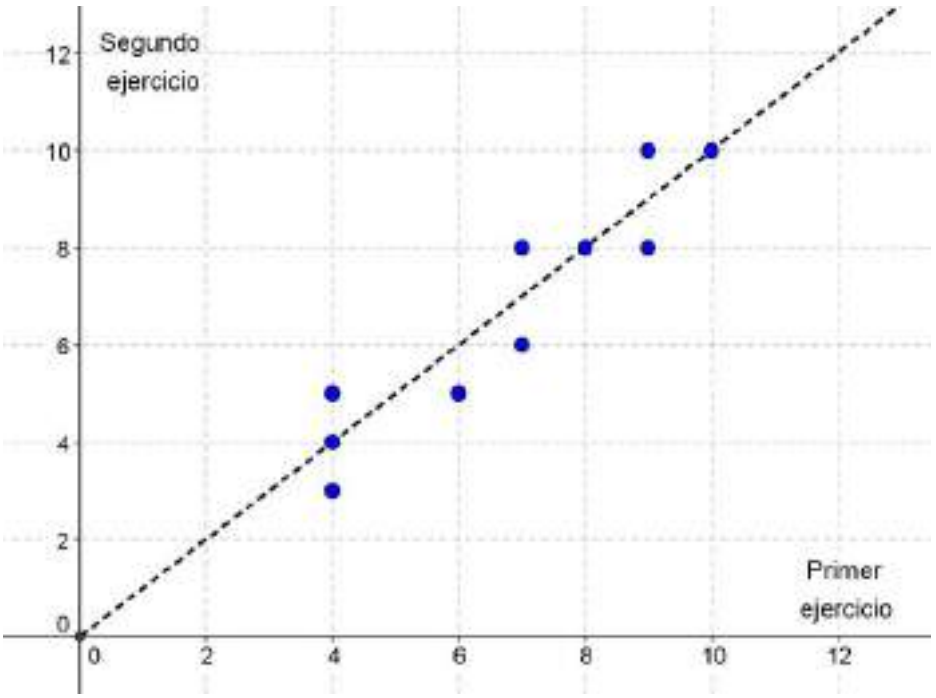
En el intervalo anterior se encuentran 9 + 18 + 19 + 8 = 54 valores del total, que representan el $\frac{54}{80} \cdot 100 = 67{,}5\%$ del total.

3. La nube de puntos parece en el gráfico.

La recta ajustada a ojo puede ser al bisectriz del primer cuadrante, y = x.

La correlación será positiva y fuerte, próxima a 1.

Si calculamos el coeficiente de correlación lineal obtenemos r = 0,927.



ACTIVIDADES-PÁG. 393

1. Veamos los dos casos límite:

1º: Si r = 0, entonces, $V = \pi \cdot h \cdot R^2$, que coincide con el volumen del cilindro.

2º: Si r = R, entonces, $V = \pi \cdot h \cdot (R^2 + R^2) = 2 \cdot \pi \cdot R^2 \cdot h$, pero si r = R el volumen es cero.

Luego la fórmula es falsa.

2. Los números felices de una cifra son 1 y 7.

Los números felices de dos cifras son: 10, 13, 19, 23, 28, 31, 32, 44, 49, 68, 70, 79, 82, 86, 91, 94 y 97.

Los primeros números felices de tres cifras son: 100, 103, 109, 129, 130, 133, 139, 167…

3. Después de varios intentos vemos que la situación final, para lograr el objetivo buscado, que debe quedar en la vía muerta superior es: W_1 W_2 L.

Llamamos A al lugar en que inicialmente está el vagón W_1 y B al lugar donde está inicuamente el vagón W_2.

Los pasos a seguir son:

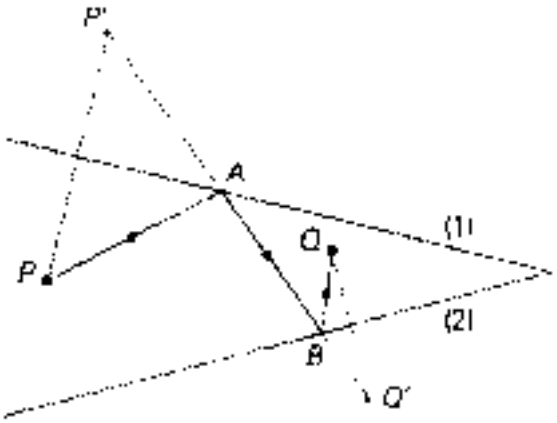
1º L coge W_1 y lo lleva a la vía muerta de abajo.
2º L da la vuelta al circuito pasando por el túnel y empuja a W_2 hasta el punto A.
3º L coge W_1 y lo lleva junto a W_2.
4º L da la vuelta al circuito y empuja a ambos vagones a la vía muerta de arriba, quedando la situación que buscábamos, W_1 W_2 L.
5º L remolca a W_2 hasta el punto A.
6º L da la vuelta al circuito y engancha a W_1 llevándolo a la posición B.
7º L vuelve a la vía muerta de arriba y los vagones han cambiado de posición.

4. Este problema es una doble simetría.

Construimos P ´, simétrico de P respecto a la banda (1), y Q ´ simétrico de Q respecto a la banda (2).

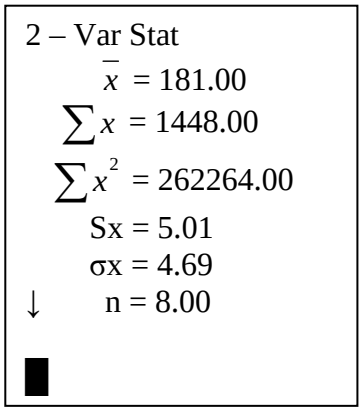
Unimos P ´ y Q ´ y llamamos A y B a los puntos en que la recta P ´Q ´ corta a las bandas.

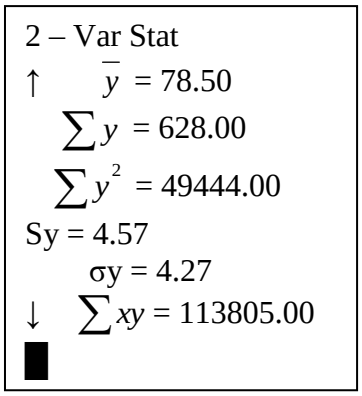
La trayectoria pedida es PABQ.

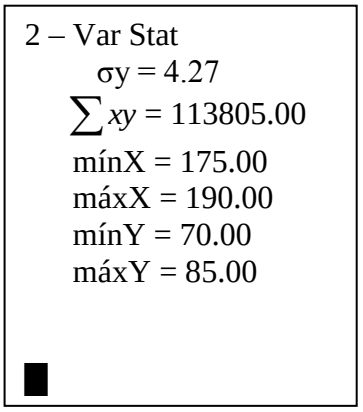


ACTIVIDADES-PÁG. 395

1. a) Utilizando la tecla  y procediendo como se explica en el texto, obtenemos los siguientes parámetros:







Para calcular el coeficiente de correlación de Pearson $r = \dfrac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$, calculamos previamente la covarianza:

$$\sigma_{xy} = \frac{\sum f_{ij}\, x_i\, y_j}{N} - \bar{x} \cdot \bar{y} = \frac{113805}{8} - 181 \cdot 78{,}50 = 17{,}125$$

Con este valor obtenemos: $r = \dfrac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \dfrac{17{,}125}{4{,}69 \cdot 4{,}27} = 0{,}86$

b) La recta de regresión del peso (Y) sobre la estatura (X) es:

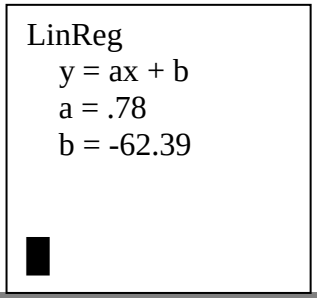
$$y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x}) \quad \Rightarrow \quad y - 78{,}50 = \frac{17{,}125}{4{,}69^2}(x - 181) \quad \Rightarrow \quad y = 0{,}78x - 62{,}39$$

La recta de regresión de la estatura (X) sobre el peso (Y) es:

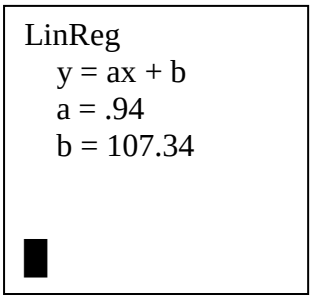
$$x - \bar{x} = \frac{\sigma_{xy}}{\sigma_y^2}(y - \bar{y}) \quad \Rightarrow \quad x - 181 = \frac{17{,}125}{4{,}27^2}(y - 78{,}50) \quad \Rightarrow \quad x = 0{,}94y + 107{,}34$$

Con la calculadora se determinan así:

● Para la recta de regresión del peso (Y) sobre la estatura (X), en el menú de tecla **STAT**, elegimos **CALC** seguido de la opción **4:LinReg(ax+b),** tecleando posteriormente **L1, L2** (teclas 2nd 1; tecla , teclas 2nd 2) y obtenemos, como vemos en la imagen, la recta de ecuación y = 0,78 x – 62,39

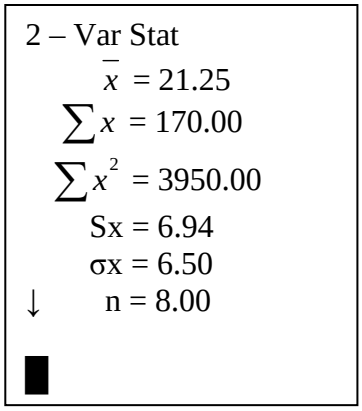


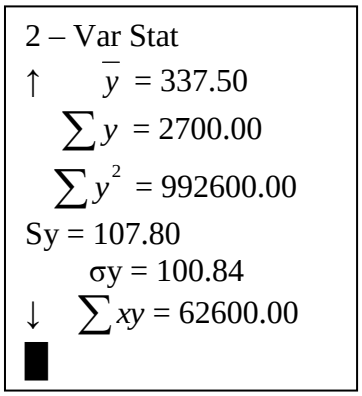
● Para la recta de regresión de la estatura (X) sobre el peso (Y), en el menú de tecla **STAT**, elegimos **CALC** seguido de la opción **4:LinReg(ax+b),** tecleando posteriormente **L2, L1** (teclas 2nd 2; tecla , teclas 2nd 1) y obtenemos, como vemos en la imagen, la recta de ecuación x = 0,94 y + 107,34

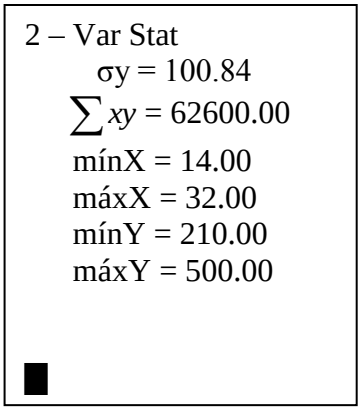


2. a) Utilizando la tecla 

y procediendo como se explica en el texto, obtenemos los siguientes parámetros para los datos de la tabla del enunciado:







Para calcular el coeficiente de correlación de Pearson $r = \dfrac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$, calculamos previamente la covarianza:

$$\sigma_{xy} = \frac{\sum f_{ij}\, x_i\, y_j}{N} - \bar{x} \cdot \bar{y} = \frac{62600}{8} - 21{,}25 \cdot 337{,}50 = 653{,}125$$

17

Con este valor obtenemos: $r = \dfrac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \dfrac{653{,}125}{6{,}50 \cdot 100{,}84} = 0{,}996$

La recta de regresión del número de conejos (Y) sobre el número de zorros (X) es:

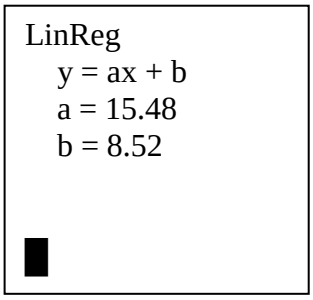
$$y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x}) \quad \Rightarrow \quad y - 337{,}50 = \frac{635{,}125}{6{,}50^2}(x - 21{,}25) \quad \Rightarrow \quad y = 15{,}48x + 8{,}52$$

La recta de regresión del número de zorros (X) sobre el número de conejos (Y) es:

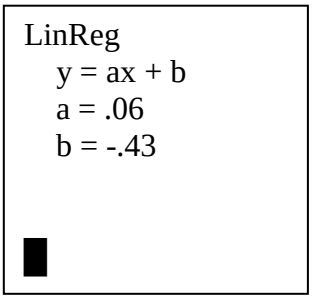
$$x - \bar{x} = \frac{\sigma_{xy}}{\sigma_y^2}(y - \bar{y}) \quad \Rightarrow \quad x - 21{,}25 = \frac{635{,}125}{100{,}84^2}(y - 337{,}50) \quad \Rightarrow \quad x = 0{,}06y - 0{,}43$$

Con la calculadora se determinan así:

● Para la recta de regresión del número de conejos (Y) sobre el número de zorros (X), en el menú de tecla **STAT**, elegimos **CALC** seguido de la opción **4:LinReg(ax+b),** tecleando posteriormente **L1, L2** (teclas 2nd 1; tecla , teclas 2nd 2) y obtenemos, como vemos en la imagen, la recta de ecuación $y = 15{,}48\,x + 8{,}52$



● Para la recta de regresión del número de zorros (X) sobre el número de conejos (Y), en el menú de tecla **STAT**, elegimos **CALC** seguido de la opción **4:LinReg(ax+b),** tecleando posteriormente **L2, L1** (teclas 2nd 2; tecla , teclas 2nd 1) y obtenemos, como vemos en la imagen, la recta de ecuación $x = 0{,}06\,y - 0{,}43$



b) Estimamos la cantidad de conejos que habría si hubiera 10 zorros, calculando en la recta de regresión de Y sobre X, de ecuación $y = 15{,}48x + 8{,}52$, el valor que se obtiene al hacer $x = 10$.

Operando, obtenemos:

Si $x = 10 \Rightarrow y = 15{,}48 \cdot 10 + 8{,}52 \Rightarrow y = 163{,}32$.

Por tanto, si hubiera 10 zorros, la cantidad de conejos estimada sería 163.

c) Estimamos la cantidad de zorros que habría si hubiéramos contado 350 conejos, calculando en la recta de regresión de X sobre Y, de ecuación $x = 0{,}06y - 0{,}43$, el valor que se obtiene al hacer $y = 350$.

Operando, obtenemos:

Si $y = 350 \Rightarrow x = 0{,}06 \cdot 350 - 0{,}43 \Rightarrow x = 20{,}57$.

Por tanto, si hubiera 350 conejos, la cantidad de zorros estimada sería 21.

ACTIVIDADES-PÁG. 396

1. Las soluciones son:

La media: $\overline{x} = 172{,}5$.

La desviación típica: $\sigma = 12{,}91$

El número de países en:

$(\overline{x} - \sigma, \overline{x} + \sigma) = (159{,}59; 185{,}41)$ es 161.

$(\overline{x} - 2\sigma, \overline{x} + 2\sigma) = (146{,}68; 198{,}32)$ es 200.

$(\overline{x} - 3\sigma, \overline{x} + 3\sigma) = (133{,}77; 211{,}23)$ es 200.

2. Los valores pedidos son:

Las medias aritméticas son: $\overline{x_A} = \frac{276}{10} = 26{,}7$ y $\overline{x_B} = \frac{285}{10} = 28{,}5$.

Las desviaciones típicas son: $\sigma_A = \sqrt{\frac{7243}{10} - (26{,}7)^2} = 3{,}38$ y $\sigma_B = \sqrt{\frac{8209}{10} - (28{,}5)^2} = 2{,}94$

Será aconsejable optar por la marca B, ya que tiene mayor media y, a su vez, menos desviación típica.

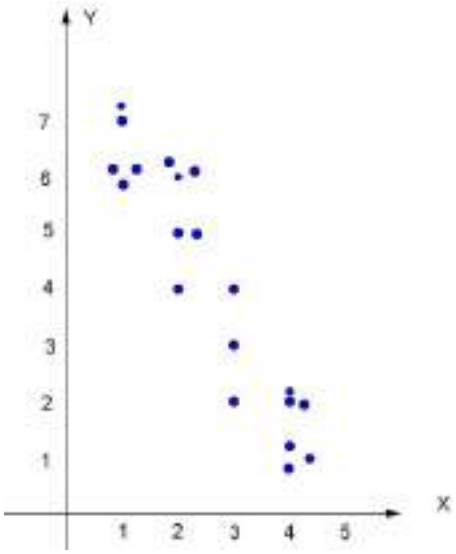
3. En cada caso queda:

a) No existe correlación.
b) Existe correlación negativa y fuerte.
c) Existe correlación positiva y fuerte.
d) No existe correlación.

4. a) La tabla de doble entrada es:

Y Viajes hijos / X Viajes padres	**1**	**2**	**3**	**4**	**TOTALES**
1				3	3
2			1	3	4
3			1		1
4		1	1		2
5		2			2
6	3	3			6
7	2				1
TOTALES	5	6	3	6	20

b) El diagrama de dispersión es:



Se observa una correlación negativa fuerte (puede calcularse el coeficiente de correlación r = - 0,944).

ACTIVIDADES-PÁG. 397

5. a) La tabla bidimensional de doble entrada es:

X / Y	3	4	5	6	7	8	9	10	Totales
3		2							2
4	1	2	2						5
5		3	3	2	5				13
6			3						3
7									0
8				5		6			11
9									0
10						5	7	4	16
Totales	1	7	8	7	5	11	7	4	50

b) La tabla bidimensional simple es:

x_i	3	4	4	4	5	5	5	6	6	7	8	8	9	10
y_i	4	3	4	5	4	5	6	5	8	5	8	10	10	10
f_i	1	2	2	3	2	3	3	2	5	5	6	5	7	4

c) Las tablas de las distribuciones marginales son:

x_i	3	4	5	6	7	8	9	10	Total
f_i	1	7	8	7	5	11	7	4	50

y_i	3	4	5	6	7	8	9	10	Total
f_i	2	5	13	3	0	11	0	16	50

d) La distribución correspondiente a la variable X condicionada a que Y tome el valor 5 es:

$x_i / _{Y=5}$	3	4	5	6	7	8	9	10	Total
f_i	0	3	3	2	5	0	0	0	13

e) La distribución correspondiente a la variable Y condicionada a que X tome el valor 5 es:

$y_i / _{X=5}$	3	4	5	6	7	8	9	10	Total
f_i	0	2	3	3	0	0	0	0	8

6. a) La tabla bidimensional simple es:

x_i	3	3	4	4	5	5	6	6	7	7
y_i	1	2	2	3	3	4	4	5	4	5
f_i	1	2	4	6	10	12	15	5	1	4

b) Los parámetros buscados son:

x_i	f_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$
3	3	9	27
4	10	40	160
5	22	110	550
6	20	120	720
7	5	35	245
Sumas	60	314	1702

$$\bar{x} = \frac{314}{60} = 5{,}23 \qquad \sigma_x = \sqrt{\frac{1702}{60} - (5{,}23)^2} = 1{,}01$$

y_i	f_i	$f_i \cdot y_i$	$f_i \cdot yx_i^2$
1	1	1	1
2	6	12	24
3	16	48	144
4	28	112	448
5	9	345	225
Sumas	60	218	8402

$$\bar{y} = \frac{218}{60} = 3{,}63 \qquad \sigma_y = \sqrt{\frac{842}{60} - (3{,}63)^2} = 0{,}93$$

c) La media aritmética y la desviación típica de la distribución de la variable X condicionada a que Y valga 4 es:

$x_i / _{y=4}$	f_i	$f_i \cdot x_i / _{y=4}$	$f_i \cdot (x_i / _{y=4})^2$
3	0	0	0
4	0	0	0
5	12	60	300
6	15	90	540
7	1	7	49
Sumas	28	157	889

$$\bar{x}/_{Y=4} = \frac{157}{28} = 5{,}607 \qquad \sigma_{x/y=4} = \sqrt{\frac{889}{28} - (5{,}607)^2} = 0{,}56$$

d) Calcula los parámetros anteriores para la distribución de la variable Y condicionada a que X valga 5.

$y_i / _{x=54}$	f_i	$f_i \cdot y_i / _{x=5}$	$f_i \cdot (y_i / _{x=5})^2$
1	0	0	0
2	0	0	0
3	10	30	90
4	12	48	192
5	0	0	0
Sumas	22	78	182

$$\bar{y}/_{x=5} = \frac{282}{22} = 3{,}55 \qquad \sigma_{y/x=5} = \sqrt{\frac{282}{22} - (3{,}55)^2} = 0{,}50$$

7. Las respuestas a los apartados son:

a)

X/Y	**1**	**2**	**3**	**Total**
1	9	0	0	9
2	14	7	0	21
3	16	9	5	30
4	20	12	8	40
Total	59	28	13	**100**

x_i	f_i
1	9
2	21
3	30
4	40

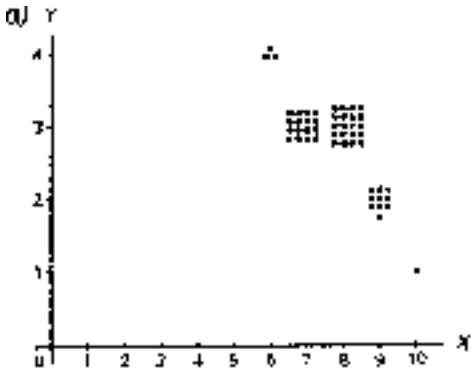
y_i	f_i
1	59
2	28
3	13

b) $\bar{x} = 3{,}01 \quad \sigma_X = 0{,}98$

$\bar{y} = 1{,}54 \quad \sigma_Y = 0{,}71$

8. Las soluciones son:

a)



b) Para amabas variables queda:

$$\bar{x} = \frac{390}{50} = 7{,}8 \text{ horas dormidas y } \sigma_X = 0{,}89$$

$$\bar{y} = \frac{141}{50} = 2{,}82 \text{ horas televisión y } \sigma_Y = 0{,}55$$

c) El porcentaje de individuos por encima de al media es $\frac{20 + 10 + 1}{50} = 0{,}62$, es decir, el 62%.

d) Para el cálculo de $r = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$, calculamos la covarianza: $\sigma_{XY} = \frac{1078}{50} - 7{,}8 \cdot 2{,}82 = -0{,}436.$

El coeficiente de correlación es: $r = \frac{-0{,}436}{0{,}89 \cdot 0{,}55} = -0{,}89.$

La correlación es muy fuerte y negativa.

ACTIVIDADES-PÁG. 398

9. La covarianza es $\sigma_{AB} = \frac{1619}{10} - 11{,}5 \cdot 14{,}3 = -2{,}55.$

El coeficiente de correlación es: $r = \frac{-2{,}55}{3{,}67 \cdot 2{,}72} = -0{,}255.$

La correlación es negativa y débil.

10. La correspondencia de cada gráfico con su coeficiente de correlación es:

a) r = 0,05 b) r = 0,71 c) r = - 0,98 d) r = 0,93 e) r = - 0,62

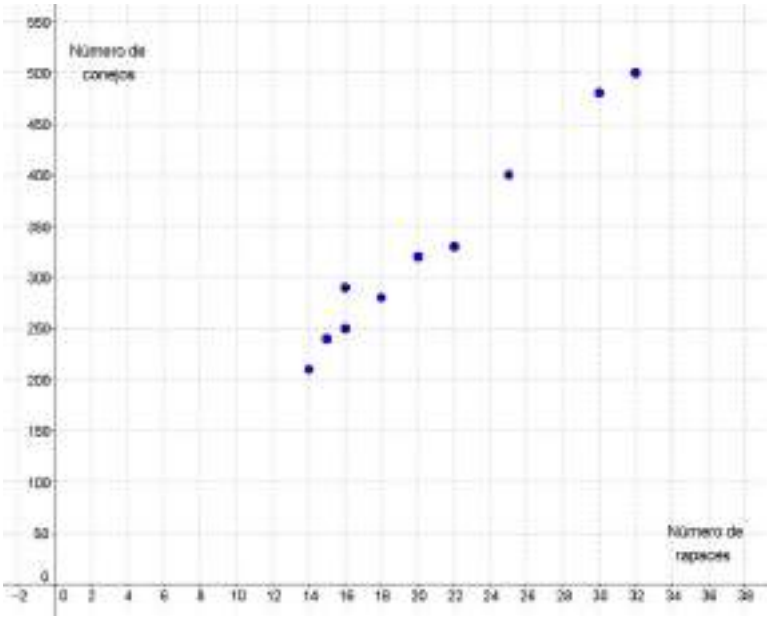
11. Los parámetros estadísticos son:

$$\bar{x} = 2{,}68;\ \bar{y} = 15{,}4;\ \sigma_X = 1{,}82;\ \sigma_Y = 7{,}97;\ \sigma_{XY} = 8{,}47$$

a) El coeficiente de correlación es: $r = \dfrac{8{,}47}{1{,}82 \cdot 7{,}96} = 0{,}58$.

b) La recta de regresión es: $y - 15{,}4 = \dfrac{8{,}47}{3{,}31}(x - 2{,}68)$, es decir, y = 2,56x + 8,54.

12. a) El diagrama de dispersión puede verse en el dibujo.



Los parámetros que se obtienen en el cálculo del coeficiente de correlación lineal son:

$$\bar{x} = 20{,}8 \qquad \sigma_X = 6{,}03 \qquad \bar{y} = 330 \qquad \sigma_Y = 94{,}55 \qquad \sigma_{XY} = 564$$

El valor del coeficiente es:

$$r = \frac{564}{6{,}03 \cdot 94{,}55} = 0{,}9892$$

Observamos que el valor obtenido nos permite afirmar que existe un excelente grado de dependencia positiva, es decir, que a mayor número de conejos, existe mayor número de rapaces.

b) Las rectas de regresión son:

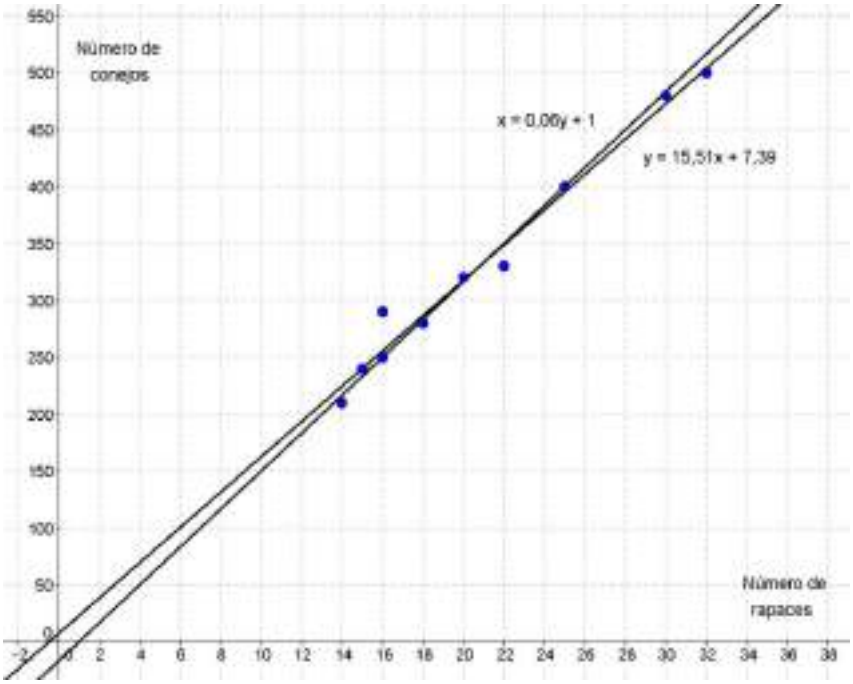
De Y sobre X es:

$$y - 330 = \frac{564}{6{,}03^2}(x - 20{,}8) \qquad \Rightarrow \qquad y = 15{,}51x + 7{,}39$$

De X sobre Y:

$$x - 20{,}8 = \frac{564}{94{,}55^2}(y - 330) \quad \Rightarrow \quad x = 0{,}06y + 1$$

Sus gráficas pueden verse en el dibujo.



c) Estimamos la cantidad de conejos que habría si hubiera 10 rapaces:

En la recta de regresión de Y sobre X: si x = 10, entonces y = 15,51 · 10 + 7,39 = 162,49 ≈ 162 conejos.

En la recta de regresión de X sobre Y: si x = 10, entonces 10 = 0,06y + 1 ⇒ y = 150 conejos.

d) Estimamos la cantidad de rapaces que habría si hubiera 350 conejos:

En la recta de regresión de Y sobre X: si y = 350, entonces 350 = 15,51 · y + 7,39 ⇒ 22,09 ≈ 22 rapaces.

En la recta de regresión de X sobre Y: si y = 350, entonces x = 350y + 1 = 22 rapaces.

Es más fiable la segunda estimación, ya que el valor inicial de la primera se aleja bastante de la media de rapaces.

13. Al ser el coeficiente de correlación r = 0,7; obtenemos:

$$r = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} \quad \Rightarrow \quad 0{,}7 = \frac{\sigma_{XY}}{5 \cdot 7{,}5} = \quad \Rightarrow \quad \sigma_{XY} = 26{,}25.$$

La recta de regresión de Y (estatura de los hijos) sobre X (estatura de los padres) es:

$$y - 170 = \frac{26{,}25}{5^2}(x - 168) \quad \Rightarrow \quad y = 1{,}05x - 6{,}4$$

Si un padre mide 180 cm, se estima que su hijo tendrá y = 1,05 · 180 – 6,4 = 182, 6 cm.

Nota: Todos los datos se han convertido en centímetros.

ACTIVIDADES-PÁG. 399

14. Calculamos previamente los parámetros correspondientes a las distribuciones marginales y la covarianza, obteniendo:

$$\bar{x} = \frac{108}{9} = 12 \qquad \sigma_x = \sqrt{\frac{1836}{9} - 12^2} = 7{,}75$$

$$\bar{y} = \frac{84{,}40}{9} = 9{,}38 \quad \sigma_y = \sqrt{\frac{863{,}52}{9} - (9{,}38)^2} = 2{,}83$$

$$\sigma_{xy} = \frac{1201{,}20}{9} - 12 \cdot 9{,}38 = 20{,}93$$

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
0	3,50	0	12,25	0,00
3	6,25	9	39,06	18,75
6	8,00	36	64,00	48,00
9	9,20	81	84,64	82,80
12	10,20	144	104,04	122,40
15	11,00	225	121,00	165,00
18	11,60	324	134,56	208,80
21	12,05	441	145,20	253,05
24	12,60	576	158,76	302,40
108	**84,40**	**1836**	**836,52**	**1201,20**

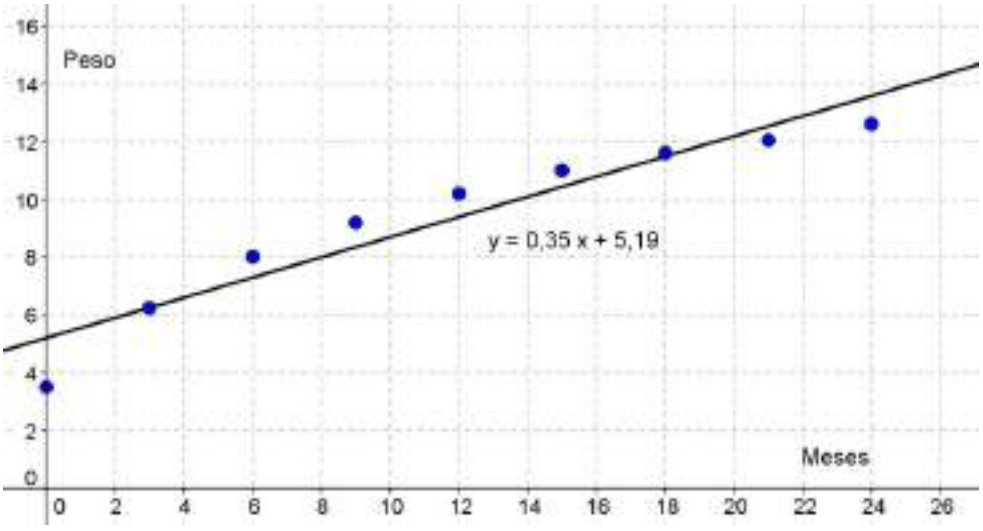
a) El coeficiente de correlación lineal vale:

$$r = \frac{20{,}93}{7{,}75 \cdot 2{,}83} = 0{,}96$$

La recta de regresión del peso (Y) en función de la edad (X) es:

$$y - 9{,}38 = \frac{20{,}93}{7{,}75^2}(x - 12) \qquad \Rightarrow \qquad y = 0{,}35x + 5{,}19$$

En el dibujo puede verse la nube de puntos y la gráfica de la recta de regresión.



b) Los valores de la varianza residual y el coeficiente de determinación son:

La varianza residual vale:

$$\sigma_e^2 = \frac{6{,}30}{9} = 0{,}70$$

El coeficiente de determinación es:

$$R^2 = 1 - \frac{0{,}70}{8{,}00} = 0{,}91$$

x_i	y_i	$\hat{y}_i = 0{,}35x_i + 5{,}19$	$e_i = \hat{y}_i - y_i$	e_i^2
0	3,50	5,19	- 1,69	2,86
3	6,25	6,24	0,01	0,00
6	8,00	7,29	0,71	0,50
9	9,20	8,34	0,86	0,74
12	10,20	9,39	0,81	0,66
15	11,00	10,44	0,56	0,31
18	11,60	11,49	0,11	0,01
21	12,05	12,54	- 0,49	0,24
24	12,60	13,59	- 0,99	0,98
				6,30

c) El incremento del peso esperado en un mes, podemos calcularlo como la diferencia de los pesos esperados para dos meses consecutivos, por ejemplo para x = 1 y x = 2:

Si x = 1, entonces $\hat{y}\,(1) = 0{,}35 \cdot 1 + 5{,}19 = 5{,}54\ kg.$

Si x = 2, entonces $\hat{y}\,(2) = 0{,}35 \cdot 2 + 5{,}19 = 5{,}89\ kg.$

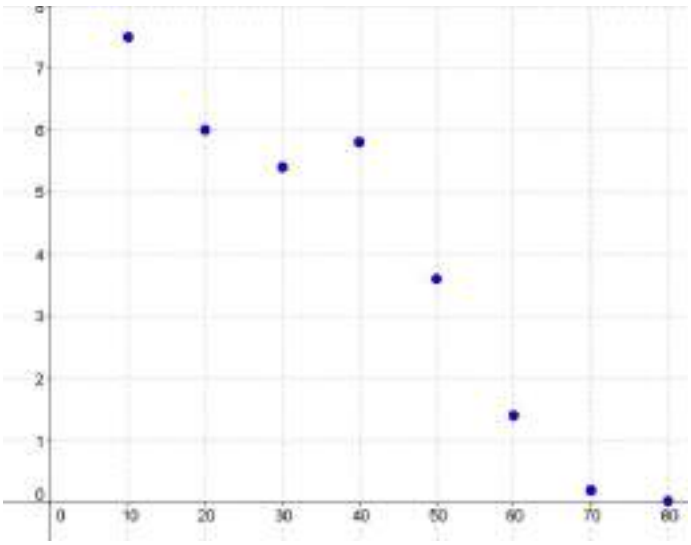
La diferencia es $\hat{y}\,(2) - \hat{y}\,(1) = 5{,}89 - 5{,}54 = 0{,}35\ kg.$

Puede observarse que el peso esperado en un mes coincide con el coeficiente de regresión $m = \frac{20{,}93}{7{,}75^2} = 0{,}35$.

d) El peso esperado para un niño de 14 meses es: $\hat{y}\,(14) = 0{,}35 \cdot 14 + 5{,}19 = 10{,}08\ kg.$

El peso esperado para un niño de dos años y medio (30 meses) es: $\hat{y}\,(30) = 0{,}35 \cdot 30 + 5{,}19 = 15{,}66\ kg.$

15. a) El diagrama de dispersión puede verse en el dibujo.



b) Los parámetros que se obtienen en el cálculo del

coeficiente de correlación lineal son:

$$\bar{x} = 45 \qquad \sigma_X = 22{,}91 \qquad \bar{y} = 3{,}75 \qquad \sigma_Y = 2{,}68 \qquad \sigma_{XY} = -59{,}41$$

El valor del coeficiente es:

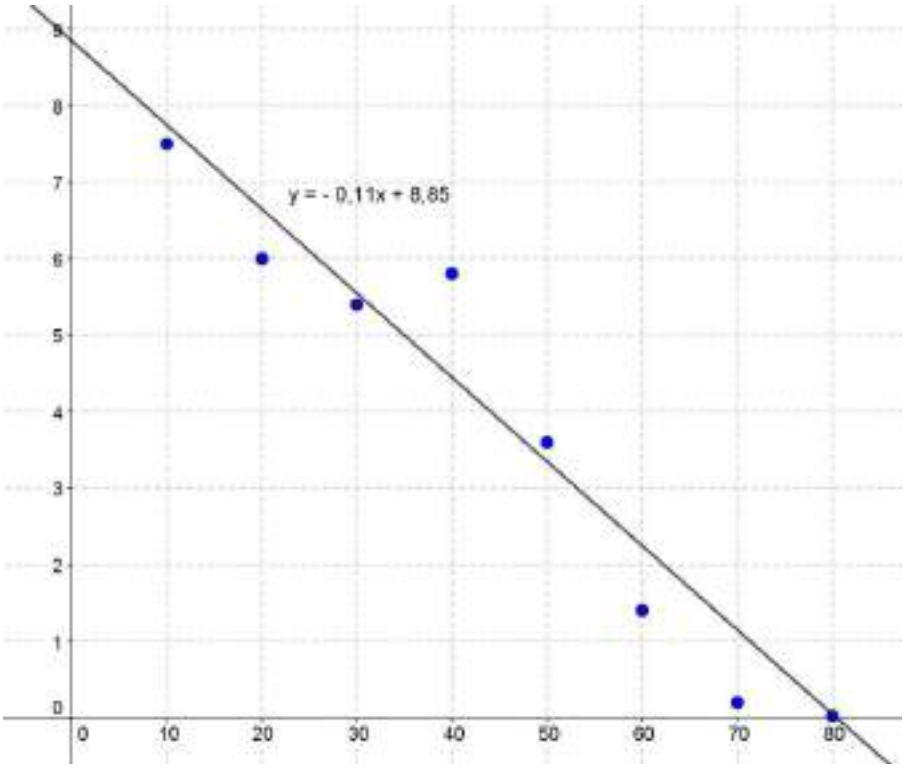
$$r = \frac{-59{,}41}{22{,}91 \cdot 2{,}68} = -0{,}968$$

Observamos que el valor obtenido nos permite afirmar que existe un excelente grado de dependencia negativa, es decir, que a mayor profundidad, existe menos oxígeno en el agua del embalse.

c) La recta de regresión de Y sobre X es:

$$y - 3{,}75 = \frac{-59{,}41}{22{,}91^2}(x - 45) \qquad \Rightarrow \qquad y = -0{,}11x + 8{,}85$$

Su gráfica puede verse en el dibujo.



d) Calculamos las estimaciones de la cantidad de oxígeno en el agua a las distintas profundidades que se piden:

Para x = 25 m, tenemos que y = - 0,11 · 25 + 8,85 = 6,1 mg/L.

Para x = 55 m, tenemos que y = - 0,11 · 55 + 8,85 = 2,8 mg/L.

Para x = 85 m, tenemos que y = - 0,11 · 85 + 8,85 = - 0,5 mg/L.

Puede observarse que los dos primeros valores son razonables, pero el último carece de sentido.

e) Nos ayudamos de los cálculos que aparecen en la tabla.

x_i	y_i	$\hat{y}_i = -0{,}11x_i + 8{,}85$	$e_i = \hat{y}_i - y_i$	e_i^2
10	7,50	7,75	- 0,25	0,0625
20	6,00	6,65	- 0,65	0,4225
30	5,40	5,55	- 0,15	0,0225
40	5,80	4,45	1,35	1,8225
50	3,60	3,35	0,25	0,0625
60	1,40	2,25	- 0,85	0,7225
70	0,30	1,15	- 0,85	0,7225
80	0,02	0,05	- 0,03	0,0009
				3,8384

La varianza residual vale: $\sigma_e^2 = \dfrac{3{,}84}{8} = 0{,}48$

El coeficiente de determinación es: $R^2 = 1 - \dfrac{0{,}48}{7{,}18} = 0{,}93$

16. a) Como la recta de regresión de Y sobre X es 4x – 3y = 0, su pendiente es el coeficiente de regresión y vale:

$$m = \frac{4}{3} = \frac{\sigma_{xy}}{\sigma_x^2}$$

La pendiente de la recta de regresión de X sobre Y, 3x – 2y = 1, es:

$$m' = \frac{3}{2} = \frac{\sigma_{xy}}{\sigma_y^2}$$

La relación entre el coeficiente de correlación lineal y los coeficientes de regresión nos permite calcular:

$$r = \sqrt{m \cdot m'} = \sqrt{\frac{4}{3} \cdot \frac{3}{2}} = \sqrt{2} = 1{,}41$$

El coeficiente de correlación es muy alto y nos permite afirmar que las variables están muy relacionadas.

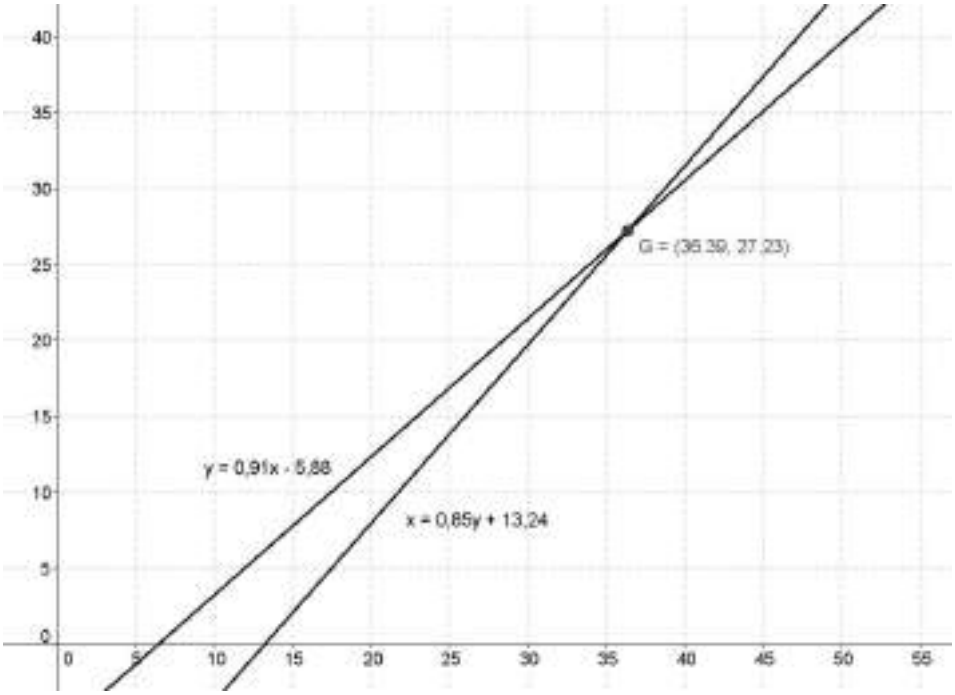
b) Sabemos que las dos rectas de regresión pasan por el punto $(\bar{x}, \bar{y})$, centro de gravedad de la nube de puntos.

Para calcular las medias de las variables, calculamos el punto de corte de las dos rectas. Resolviendo el sistema, obtenemos:

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

La nota media en teoría es $\bar{x} = 3$ y la nota media en práctica es $\bar{y} = 4$.

17. La representación gráfica puede verse en el dibujo.



El centro de gravedad de la distribución es el punto de corte de las rectas de regresión. Por tanto:

$$\begin{cases} y = 0{,}91x - 5{,}88 \\ x = 0{,}85\,y + 13{,}24 \end{cases} \Rightarrow \begin{cases} x = 36{,}39 \\ y = 27{,}23 \end{cases}$$

El centro de gravedad es el punto $G\left(\bar{x} = 36{,}39;\ \bar{y} = 27{,}23\right)$.

El cuadrado del coeficiente de correlación lineal es igual al producto de los coeficientes de regresión. Sustituyendo, obtenemos:

$$r^2 = m \cdot m' \quad \Rightarrow \quad r^2 = 0{,}91 \cdot 0{,}85 \quad \Rightarrow \quad r = \sqrt{0{,}7735} = 0{,}8795.$$

18. Observando los gráficos vemos que el ángulo formado por las rectas es más pequeño en las distribuciones II) y IV). Por tanto, en estos casos es más significativo.

Analizando las ecuaciones de las rectas obtenemos los resultados que siguen.

I) El coeficiente de regresión de la recta y = x + 2 vale m = 1, lo que significa que la covarianza σ_{xy} es no nula. Por lo tanto, no puede ser el coeficiente de regresión de la otra recta m´ = 0, como ocurre con la recta x = 4. Es decir, esta situación carece de sentido, ya que no es posible que haya una distribución con estas dos rectas de regresión.

II) En este caso, $m = \frac{4}{5}$, $m' = \frac{5}{6}$ y $r = \sqrt{\frac{4}{5} \cdot \frac{5}{6}} = \sqrt{\frac{2}{3}} = 0{,}82$.

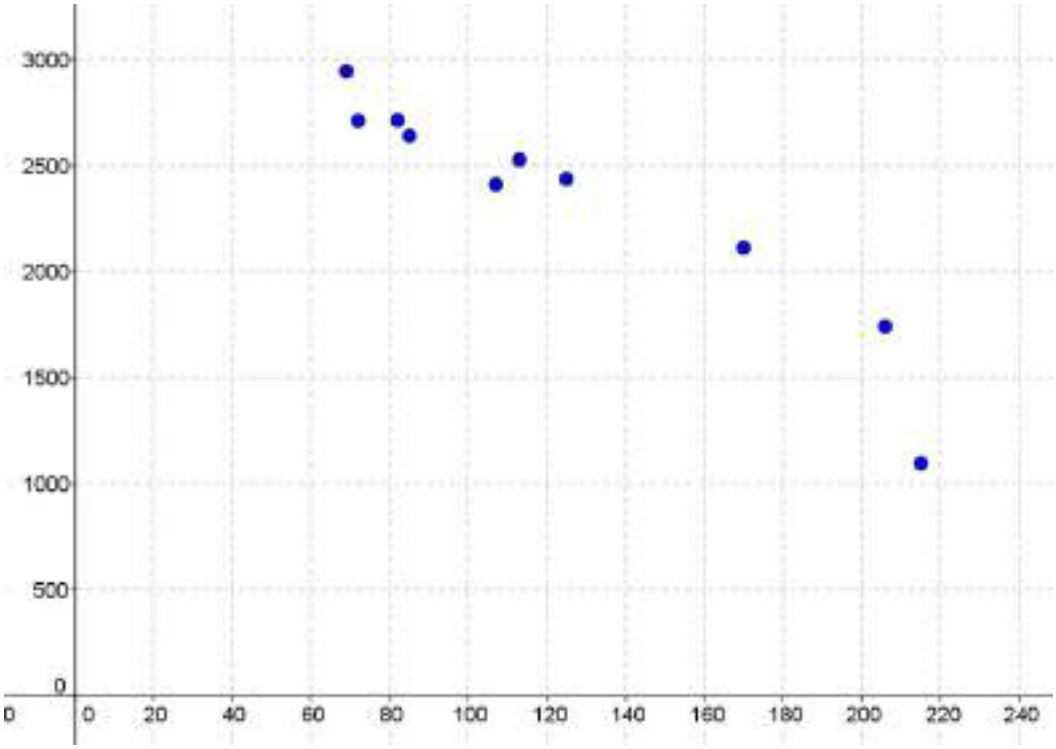
III) Para esta distribución $m = 0$, $m' = 0$ y r = 0.

IV) En esta distribución, $m = 1$, $m' = \frac{4}{5}$ y $r = \sqrt{\frac{4}{5}} = 0{,}89$.

De nueve podemos ver que la correlación es significativa en los apartados II) y IV).

ACTIVIDADES-PÁG. 400

19. El diagrama de dispersión puede verse en el dibujo.



Los parámetros que se obtienen en el cálculo del coeficiente de correlación lineal son:

$$\bar{x} = 124{,}4 \qquad \sigma_X = 51{,}52 \qquad \bar{y} = 2333{,}7 \qquad \sigma_Y = 523{,}61 \qquad \sigma_{XY} = -25593{,}18$$

El valor del coeficiente es:

$$r = \frac{-25593{,}18}{51{,}52 \cdot 523{,}61} = -0{,}9487$$

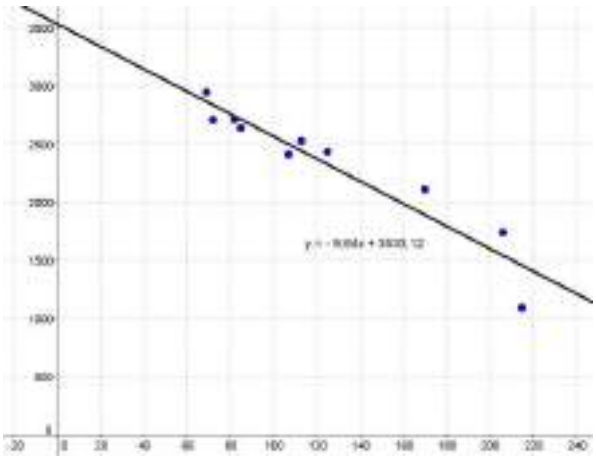
Se trata de una correlación negativa, en los lugares con más días de lluvia hay menos horas de sol y recíprocamente.

La recta de regresión de Y sobre X es:

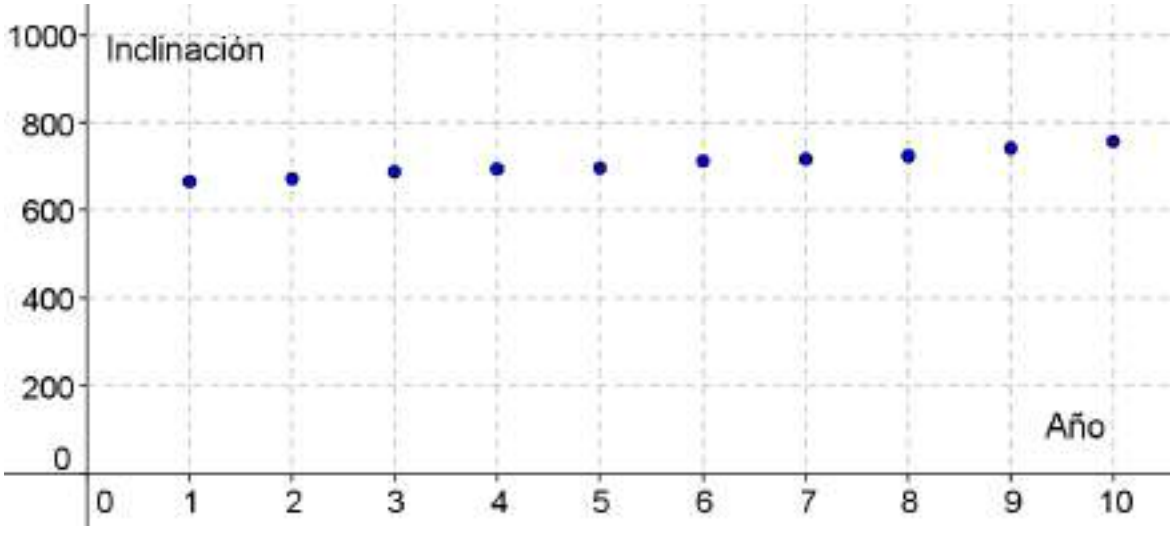
$$y - 2333{,}7 = \frac{-25593{,}18}{51{,}52^2}(x - 124{,}4) \quad \Rightarrow \quad y = -9{,}64x + 3533{,}12$$

Si se han registrado x = 100 días de lluvia se predicen:

$$y = -9{,}64 \cdot 100 + 3533{,}12 \approx 2568 \text{ horas de sol.}$$



20. a) Tomando el año 1978 como año 1, la representación gráfica puede verse en el dibujo.



A la vista de la nube de puntos parece que tiene una tendencia lineal que crece con el tiempo.

Para poder confirmarlo hallamos el coeficiente de correlación lineal.

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
1	667	1	444889	667
2	673	4	452929	1346
3	688	9	473344	2064
4	696	16	484416	2784
5	698	25	487204	3490
6	713	36	508369	4278
7	717	49	514089	5019
8	725	64	525625	5800
9	742	81	550564	6678
10	757	100	573049	7570
55	**7076**	**385**	**5014478**	**39696**

$$\bar{x} = \frac{55}{10} = 5{,}5$$

$$\sigma_x = \sqrt{\frac{385}{10} - 5{,}5^2} = 2{,}87$$

$$\bar{y} = \frac{7076}{10} = 707{,}6$$

$$\sigma_y = \sqrt{\frac{5014478}{10} - (707{,}6)^2} = 27{,}39$$

$$\sigma_{xy} = \frac{39696}{10} - 5{,}5 \cdot 707{,}6 = 77{,}8$$

El coeficiente de correlación lineal vale $r = \dfrac{77{,}8}{2{,}87 \cdot 27{,}39} = 0{,}99$.

b) La ecuación de la recta de regresión de la inclinación (Y) en función del tiempo (X) es:

$$y - 707{,}6 = \frac{77{,}8}{8{,}24}(x - 5{,}5) \quad \Rightarrow \quad y = 9{,}44x + 655{,}68$$

c) Calculamos el coeficiente de determinación.

x_i	y_i	$\hat{y}_i = 9{,}44x_i + 655{,}68$	$e_i = \hat{y}_i - y_i$	e_i^2
1	667	665,12	-1,88	3,5344
2	673	674,56	1,56	2,4336
3	688	684	-4	16
4	696	693,44	-2,56	6,5536
5	698	702,88	4,88	23,8144
6	713	712,32	-0,68	0,4624
7	717	721,76	4,76	22,6576
8	725	731,2	6,20	38,44
9	742	740,64	1,36	1,8496
10	757	750,08	-6,92	47,8864
				163,6322

La varianza residual vale: $\sigma_e^2 = \dfrac{163{,}6322}{10} = 16{,}37$

El coeficiente de determinación es: $R^2 = 1 - \dfrac{16{,}37}{750{,}21} = 0{,}98$

d) El valor ajustado para 1918 en la recta de regresión es:

$$\hat{y}\,(-59) = 9{,}47 \cdot (-59) + 655{,}68 = 96{,}95$$

El valor obtenido es muy diferente de 71, esto es debido a que el año 1918 está muy alejado del intervalo de años que estamos considerando.

21. Calculamos los parámetros de la distribución bidimensional considerando el número de horas como variable X y el número de gérmenes como la variable Y.

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
0	20	0	400	0
1	26	1	676	26
2	33	4	1089	66
3	41	9	1681	123
4	47	16	2209	188
5	53	25	2809	265
15	**220**	**55**	**8864**	**668**

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

$$\sigma_x = \sqrt{\frac{55}{6} - 2{,}5^2} = 1{,}71$$

$$\bar{y} = \frac{220}{5} = 36{,}67$$

$$\sigma_y = \sqrt{\frac{8864}{6} - (36{,}67)^2} = 11{,}53$$

$$\sigma_{xy} = \frac{668}{6} - 2{,}5 \cdot 36{,}67 = 19{,}67$$

a) La ecuación de la recta de regresión del número de gérmenes (Y), por centímetro cúbico, en función del tiempo (X) es:

$$y - 36{,}67 = \frac{19{,}67}{1{,}71^2}(x - 2{,}5) \qquad \Rightarrow \qquad y = 6{,}73x + 19{,}85$$

b) Calculamos el coeficiente de determinación.

x_i	y_i	$\hat{y}_i = 6{,}73x_i + 19{,}85$	$e_i = \hat{y}_i - y_i$	e_i^2
0	20	19,85	0,15	0,0225
1	267	26,58	- 0,58	0,3364
2	33	33,31	- 0,31	0,0961
3	41	40,04	0,96	0,9216
4	47	46,77	0,23	0,0529
5	53	53,50	- 0,50	0,2500
				1,6795

La varianza residual vale: $\sigma_e^2 = \frac{1{,}6795}{6} = 0{,}2799$

El coeficiente de determinación vale $R^2 = 1 - \frac{0{,}2799}{11{,}53^2} = 0{,}9979$

El coeficiente de correlación es $r = \sqrt{0{,}9979} = 0{,}9989$.

c) Estimamos el número de gérmenes a las 6 horas:

$$\hat{y}\,(6) = 6{,}73 \cdot 6 + 19{,}85 = 60{,}26$$

Al cabo de 6 horas habrá uno 60 miles de gérmenes por centímetro cúbico. Esta estimación tiene una gran probabilidad de ser válida ya que el coeficiente de determinación es muy alto.

22. De las rectas de regresión no podemos asegurar cuál es la de regresión de Y sobre X y cuál la de X sobre Y.

Supongamos que la primera de ellas es la de regresión de Y sobre X, se tiene:

$$y = -2x - 1$$

y su coeficiente de regresión es m = - 2.

La segunda corresponderá a la de regresión de X sobre Y, se tiene:

$$x = -\frac{3}{5}y - \frac{4}{5}$$

y su coeficiente de regresión es $m' = -\frac{3}{5}$.

Con los datos anteriores se obtiene el coeficiente de determinación es:

$$R^2 = m \cdot m' = (-2) \cdot \left(-\frac{3}{5}\right) = \frac{6}{5} > 1$$

lo cual carece de sentido.

En consecuencia, es necesario elegir las rectas de la otra forma posible.

La recta de regresión de Y sobre X es 5x + 3y + 4 = 0, se tiene:

$$y = -\frac{5}{3}y - \frac{4}{3}$$

y su coeficiente de regresión es $m = -\frac{5}{3}$.

La recta de regresión de X sobre Y es 2x + y + 1 = 0, se tiene:

$$x = -\frac{1}{2}y - \frac{1}{2}$$

y su coeficiente de regresión es $m' = -\frac{1}{2}$.

El sigo negativo de m y m ´ nos indica que la dependencia lineal entre las variables es de tipo inverso, y el coeficiente de determinación es:

$$R^2 = m \cdot m' = \left(-\frac{3}{5}\right) \cdot = \frac{5}{6} = 0{,}83$$

Como el coeficiente de correlación es $r = \pm\sqrt{R^2}$ y estamos ante una dependencia de tipo inverso, este coeficiente vale:

$$r = -\sqrt{0{,}83} = -0{,}91.$$