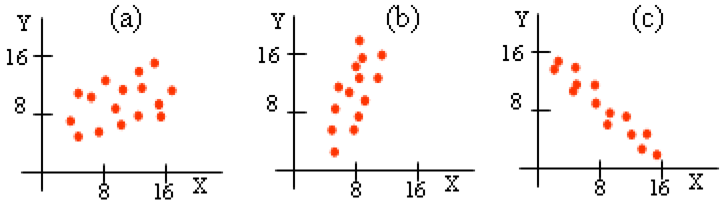
DISTRIBUCIONES BIDIMENSIONALES

PROBLEMAS RESUELTOS

1. a) Asocia las rectas de regresión: $y = -x + 16$, $y = 2x - 12$ e $y = 0{,}5x + 5$ a las nubes de puntos siguientes:



b) Asigna los coeficientes de correlación lineal $r = 0{,}4$, $r = -0{,}85$ y $r = 0{,}7$, a las mismas nubes de puntos.

<u>Solución</u>:

a) Respectivamente: (c), (b), (a).

b) Respectivamente: (a), (b), (c).

2. Se han tomado ocho medidas de la temperatura de una batería y de su voltaje, y se obtuvieron los siguientes datos:

X: temperatura	10,0	10,0	23,1	23,5	34,0	34,5	45,0	45,6
Y: voltaje	430	425	450	460	470	480	495	510

a) Sin efectuar cálculos, razona cuál de las siguientes rectas es la recta de regresión de *Y* sobre *X* para los datos anteriores:

$y = 350 - 2{,}1x$ $\quad$ $y = 460 - 2{,}1x$ $\quad$ $y = 406 + 2{,}1x$

b) Para 25 grados, ¿qué voltaje sería razonable suponer?

<u>Solución</u>:

a) Puede observarse que al aumentar la temperatura también lo hace el voltaje; por tanto, la correlación es positiva. Como el signo de la correlación es el mismo que el de la pendiente de la recta de regresión, la única recta posible es $y = 406 + 2{,}1x$.

b) Para esa ecuación, si $x = 25$ se tiene $y = 406 + 2{,}1 \cdot 25 = 458{,}5$.

3. El número de horas de estudio de una asignatura y la calificación obtenida en el examen correspondiente, fue, para 7 personas, la siguiente:

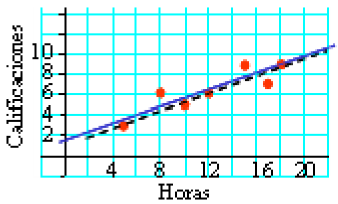
Horas	5	8	10	12	15	17	18
Calificación	3	6	5	6	9	7	9

a) Dibuja la nube de puntos y traza, aproximadamente, la recta de regresión asociada.

b) Indica el carácter y estima la fuerza de la correlación.

<u>Solución</u>:

a) En la figura adjunta se representan los puntos de la tabla anterior. La línea de trazos puede ser una recta de regresión aceptable.



b) Obviamente, la correlación es directa y fuerte.

4. Calcula, paso a paso (sin utilizar la calculadora en modo estadístico), el coeficiente de correlación y la recta de regresión asociada a los datos del problema anterior.
Solución:
El coeficiente de correlación vale $r = \dfrac{s_{xy}}{s_x \cdot s_y}$. Para hallarlo hay que calcular las desviaciones típicas marginales y la covarianza.
Utilizaremos las fórmulas:

$$s_x = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2} \ ; \ s_y = \sqrt{\frac{\sum y_i^2}{n} - \bar{y}^2} \ ; \ s_{xy} = \frac{\sum x_i y_i}{n} - \overline{xy}$$

Observa que debemos hacer sumas, sumas de cuadrados y sumas de productos; para ello resulta eficaz la siguiente tabla:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
5	3	25	9	15
8	6	64	36	48
10	5	100	25	50
12	6	144	36	72
15	9	225	81	135
17	7	289	49	119
18	9	324	81	162
$\sum x_i = 85$	$\sum y_i = 45$	$\sum x_i^2 = 1171$	$\sum y_i^2 = 317$	$\sum x_i y_i = 601$

Con esto:

$$\bar{x} = \frac{85}{7} = 12{,}14 \qquad \bar{y} = \frac{45}{7} = 6{,}43$$

$$s_x = \sqrt{\frac{1171}{7} - 12{,}14^2} = 4{,}46 ; \ s_y = \sqrt{\frac{317}{7} - 6{,}43^2} = 1{,}99 ; \ s_{xy} = \frac{601}{7} - 12{,}14 \cdot 6{,}43 = 7{,}80$$

Por último, $r = \dfrac{7{,}80}{4{,}46 \cdot 1{,}99} = 0{,}88$.
Al ser *r* próximo a +1, la correlación entre las horas de estudio y las notas de examen es directa y fuerte: a más horas de estudio, mejor nota de examen.
La recta de regresión es

$$y - 6{,}43 = \frac{7{,}80}{19{,}90}(x - 12{,}14) \Leftrightarrow \ y = 0{,}39x + 1{,}67$$

(Es la recta continua de la figura anterior).
Nota: Los resultados anteriores están redondeados a las centésimas.
Con más precisión (con la calculadora): $r = 0{,}879689$; $y = 0{,}393004x + 1{,}656379$.

5. a) Calcula la recta de regresión de Y sobre X en la distribución siguiente realizando todos los cálculos intermedios.

X	10	7	5	3	0
Y	2	4	6	8	10

b) ¿Cuál es el valor que correspondería según dicha recta a X = 7?
Solución:
Se forma la tabla:

X	Y	X^2	Y^2	X·Y
10	2	100	4	20
7	4	49	16	28
5	6	25	36	30
3	8	9	64	24
0	10	0	100	0
$\Sigma X_i = 25$	$\Sigma Y_i = 30$	$\Sigma X_i^2 = 183$	$\Sigma Y_i^2 = 220$	$\Sigma X_i Y_i = 102$

Se obtiene.

$$\bar{x} = 5;\quad s_x^{\,2} = \frac{183}{5} - 5^2 = 11{,}6;\quad \bar{y} = 6;\quad s_{xy} = \frac{102}{5} - 5{\cdot}6 = -9{,}6$$

La ecuación de la recta de regresión es

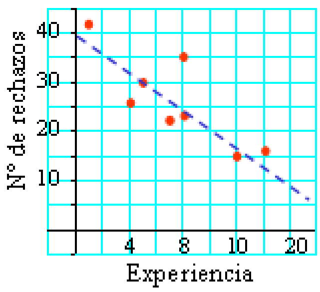
$$y - \bar{y} = \frac{s_{xy}}{s_x^{\,2}}(x - \bar{x}) \quad \Rightarrow \quad y = -0{,}8276x + 10{,}138$$

b) Si X = 7 $\Rightarrow$ Y = 4,3448.

6. El departamento de control de calidad de una empresa de instalación de componentes electrónicos desea determinar la relación entre las semanas de experiencia de sus trabajadores y el número de componentes rechazados a esos trabajadores la semana anterior.

Trabajador examinado	A	B	C	D	E	F	G	H	I	J
Semanas de experiencia (*X*)	7	8	10	1	4	5	15	18	4	8
Número de rechazos (*Y*)	22	35	15	42	26	30	16	20	31	23

a) Representa el diagrama de dispersión asociado a esos datos. ¿Sugiere la gráfica alguna asociación lineal?
b) ¿Cómo calificarías la correlación?
Solución:
a)



Puede ajustarse una recta como la que se ha trazado.

b) Salvo para la persona B, la correlación parece fuerte e inversa.

7. Para los datos del problema anterior, halla con ayuda de la calculadora:
a) Las medias y desviaciones típicas marginales.
b) La covarianza.
c) El coeficiente de correlación lineal.
d) La recta de regresión de Y sobre X.
e) El número de rechazos que hay que esperar para una persona con 20 semanas de experiencia.
Solución:

Sumas:

$\sum x_i = 80;\ \sum y_i = 260;\ \sum x_i^{\ 2} = 884;\ \sum y_i^{\ 2} = 7420;\ \sum x_i y_i = 1788$

a) $\bar{x} = 8$; $s_x = 4{,}93963$; $\bar{y} = 26$; $s_y = 8{,}12403$

b) $s_{xy} = 178{,}8 - 8 \cdot 26 = -29{,}2$

c) $r = -29{,}2 /(4{,}93963 \cdot 8{,}12403) = -0{,}72763$

d) $y = -1{,}19672x + 35{,}5737$

e) 11,6, que se aproxima a 12, el entero más próximo.

8. Se midieron los valores de concentración de una sustancia A en suero fetal y los valores de su concentración en suero materno. Se obtuvieron los siguientes datos en una muestra de seis embarazadas a término:

Madre (*X*)	8	4	12	2	7	9
Feto (*Y*)	6	4	8	1	4	5

a) Calcula el coeficiente de correlación lineal.

b) Halla la expresión de la recta que permita estimar los valores fetales a partir de los maternos.

Solución:

a) Con la calculadora se obtiene: $r = 0{,}93$

Otros parámetros de interés:

$\bar{x} = 7$; $s_x = 3{,}2659$; $\bar{y} = 4{,}667$; $s_y = 2{,}1344$

También se obtiene: A = 0,40; B = 0,609

b) La recta de regresión es: Y = A + BX $\Rightarrow y = 0{,}40 + 0{,}609x$

9. En seis alumnos de bachillerato se observaron dos variables: *X* = puntuación obtenida en un determinado test e *Y* = nota en un examen de matemáticas. Los resultados se indican en la siguiente tabla:

Test: *X*	110	90	140	120	120	100
Examen: *Y*	6	5	9	7	8	6

a) Halla la recta de regresión.

b) Sabiendo que un alumno obtuvo 130 puntos en el test, pero no realizó el examen de matemáticas, predice, si es posible, la nota que hubiese obtenido.

Solución**:**

a) Utilizando la calculadora se obtiene que:

La ecuación de la recta de regresión es: $y = -2{,}28 + 0{,}08x$

El coeficiente de correlación vale $r = 0{,}9569$.

b) Si un alumno obtuvo en el test 130, se puede estimar que su nota en matemáticas sería

$y = 2{,}28 + 0{,}08 \cdot 130 = 8{,}12$

Como *r* es muy alto y 130 está dentro del rango de los datos considerados, la estimación es fiable.

10. La altura, en cm, de 8 padres y del mayor de sus hijos varones, son:

Padre (X)	170	173	178	167	171	169	184	175
Hijo (Y)	172	177	175	170	178	169	180	187

a) Calcula la recta de regresión que permita estimar la altura de los hijos dependiendo de la del padre; y la del padre conociendo la del hijo.

b) ¿Qué altura cabría esperar para un hijo si su padre mide 174? ¿Y para un padre, si su hijo mide 190 cm?

Solución:
a) Se utilizará la calculadora en el modo estadístico.
Si X indica la atura del padre e Y la del hijo, se tendrá:
$Y = 68{,}1853 + 0{,}621859 \cdot X$; $X = 77{,}4406 + 0{,}545082 \cdot Y$.

b) Si el padre (X) mide 174 cm (en la primera ecuación) $\Rightarrow$ para el hijo caber esperar una estatura de Y = 176,4 cm.
Si el hijo (Y) mide 190 cm (en la segunda ecuación) $\Rightarrow$ para el padre puede suponerse una estatura de X = 181 cm.

11. Los años de siete árboles y el diámetro de su tronco, en cm, se dan en la siguiente tabla:

Años	2	4	5	8	10	14	20
Diámetro	10	15	17	20	23	25	27

a) Calcula, utilizando la recta de regresión, el diámetro que se puede predecir para árboles de 10 y 20 años.
b) Compara el resultado anterior con los valores observados en la tabla. Razona el porqué de las diferencias.
Solución:
a) X = años; Y = diámetro.
$\bar{x} = 9$; $s_x = 5{,}83$; $\bar{y} = 19{,}57$; $s_y = 5{,}55$; $r = 0{,}93563$.
$y = 11{,}55 + 0{,}89x$.

b) Y(10) = 20,45; Y(20) = 29,35.
Las diferencias son debidas a que la recta de regresión da la media del valor esperado.

12. El número de bacterias por unidad de volumen, presentes en un cultivo después de un cierto número de horas, viene expresado en la siguiente tabla:

X: Nº de horas	0	1	2	3	4	5
Y: Nº de bacterias	12	19	23	34	56	62

Calcula:
a) Las medias y desviaciones típicas de las variables, número de horas y número de bacterias.
b) La covarianza de la variable bidimensional.
c) El coeficiente de correlación e interpretación.
d) La recta de regresión de Y sobre X.
Solución:
Sumas:
$\sum x_i = 15$; $\sum y_i = 206$; $\sum x_i^2 = 55$; $\sum y_i^2 = 9170$; $\sum x_i y_i = 701$
a) $\bar{x} = 2{,}5$; $s_x = 1{,}70782$; $\bar{y} = 34{,}3333$; $s_y = 18{,}6964$
b) $s_{xy} = 31$
c) $r = 0{,}97086$
d) $y = 10{,}6285x + 7{,}7619$

OTROS PROBLEMAS

13. Un conjunto de datos bidimensionales (x, y) tiene un coeficiente de correlación $r = 0{,}8$. Las medias marginales valen: $\bar{x} = 2$; $\bar{y} = 4$. Indica si alguna de las siguientes ecuaciones puede corresponder a la recta de regresión de Y sobre X:
$y = -2x + 8$; $y = 0{,}8x + 2$, $y = 1{,}5x + 1$
Solución:

Se sabe que el signo de la pendiente de la recta de regresión es el mismo que el de la correlación. Por tanto, como *r* es positivo, hay que descartar la primera recta.
Por otra parte, la recta de regresión pasa por el centro medio de la distribución, el punto $(\bar{x}, \bar{y}) = (2, 4)$. La única recta que lo contiene es la tercera.
En consecuencia, la recta de regresión pedida es $y = 1,5x + 1$.

14. Halla el centro medio de una distribución sabiendo que sus rectas de regresión valen:
De Y sobre X: $y = 2x + 2$.
De X sobre Y: $x = 0,45y – 0,2$.
Solución:
Como las dos rectas pasan por el centro medio, este punto será la solución del sistema

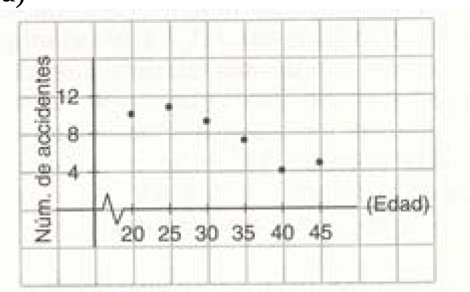
$$\begin{cases} y = 2x + 2 \\ x = 0,45y - 02 \end{cases}$$

Cuya solución es: $x = 7$, $y = 16$.
El centro medio será $(\bar{x}, \bar{y}) = (7, 16)$.

15. Una compañía de seguros sospecha que el número de accidentes está en función de la edad del conductor. Para ello elige 100 personas de cada grupo de edad y contabiliza los accidentes totales del último año. Los datos fueron:

Edad	20	25	30	35	40	45
N.º accidentes	10	11	9	7	4	5

a) Representa gráficamente la nube de puntos asociada a estos datos. ¿Qué correlación se observa?
b) Halla, sin calculadora, el coeficiente de correlación lineal entre las variables medidas. Comenta su valor.
Solución:
a)



Correlación inversa.

b) Observa que debemos hacer sumas, sumas de cuadrados y sumas de productos; para ello resulta eficaz la siguiente tabla:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
20	10	400	100	200
25	11	625	121	275
30	9	900	81	270
35	7	1225	49	245
40	4	1600	16	160
45	5	2025	25	225
$\sum x_i = 195$	$\sum y_i = 46$	$\sum x_i^2 = 6675$	$\sum y_i^2 = 392$	$\sum x_i y_i = 1375$

Con esto:

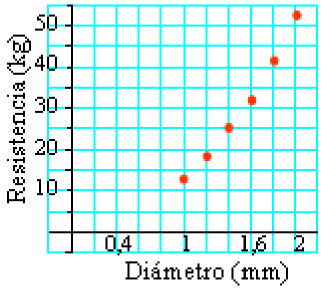
$$\bar{x} = 32{,}5;\ \bar{y} = 7{,}667;\ s_x = 8{,}539;\ s_y = 2{,}56;\ s_{xy} = -20;\ r = -0{,}915.$$

La correlación entre el número de accidentes y la edad es negativa y muy fuerte: los conductores más novatos tiene más accidentes.

16. Se está experimentado la resistencia a la rotura de una determinad fibra textil. Para ello se ha medido el diámetro de la fibra y el peso que soporta hasta la rotura, obteniéndose los siguientes datos:

Diámetro en mm (X)	1	1,2	1,4	1,6	1,8	2
Peso a la rotura en kg (Y)	12,5	18	25	32	41	52

a) Representa el diagrama de dispersión asociado a esos datos. ¿Sugiere la gráfica alguna asociación lineal?
b) ¿Cómo calificarías la correlación?
Solución:
a)



Claramente se adivina una correlación lineal
b) Positiva y muy fuerte.

17. Con los datos del problema anterior, halla:
a) La recta de regresión de Y sobre X y determina la resistencia a la rotura de una fibra de 2,5 mm de diámetro.
b) La recta de regresión de X sobre Y y determina el diámetro mínimo de una fibra para que soporte más de 60 kg.
Solución:
Utilizando la calculadora:
a) $y = 39{,}0714x - 28{,}5238$.
Para x = 2,5 mm, y = 69,1547 kg.

b) $x = 0{,}02522y + 0{,}74112$.
Para y = 60 kg, x = 2,25 mm