1 **Calcula las medias, las varianzas, las desviaciones típicas y la covarianza de las dos distribuciones siguientes:**

a)

x_i	1	2	3	4	5
y_i	3	5	7	9	11

b)

x_i	–2	–1	0	1	3	5
y_i	4	2	1	1	3	6

a) $\bar{x} = 3$; $s_x = 1{,}41$; $s_x^2 = 2$ $\bar{y} = 7$; $s_y = 2{,}83$; $s_y^2 = 8$ $s_{xy} = 4$

b) $\bar{x} = 1$; $s_x = 2{,}38$; $s_x^2 = 5{,}67$ $\bar{y} = 2{,}83$; $s_y = 1{,}77$; $s_y^2 = 3{,}14$ $s_{xy} = 2{,}17$

2 **Calcula e interpreta el coeficiente de correlación de las tablas del ejercicio 1.**

a) $r = 1$. La dependencia es funcional, es la recta $= 2x + 1$.

b) $r = 0{,}52$. La correlación es directa y débil.

3 **Con la tabla b) del ejercicio 1 calcula y dibuja la recta de regresión de *y* sobre . Para un valor de = 5,5, obtén el valor estimado de *y*.**

$y = 2{,}45098 + 0{,}38235$

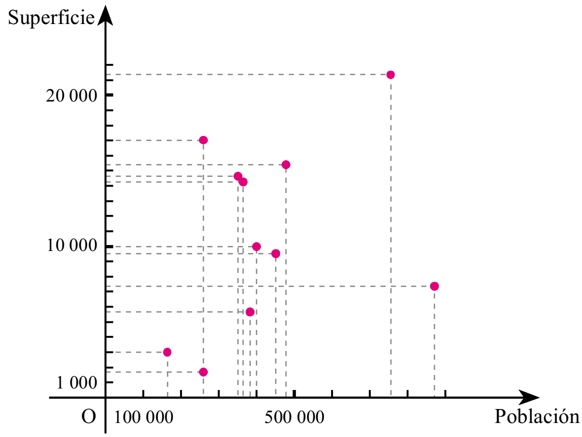
$y(5{,}5) = 4{,}5539$

4 **En el cuadro siguiente tienes la superficie en km² y la población de 10 provincias españolas. Estudia su posible correlación.**

Provincia	Población	Superficie
Alava	179 869	3 047
Albacete	341 812	14 858
Badajoz	754 454	21 657
Burgos	362 787	14 269
Cádiz	878 518	7 385
Cuenca	266 590	17 061
Girona	394 786	5 886
Huelva	401 549	10 085
Lugo	458 197	9 803
Toledo	488 599	15 368

$r = 0{,}228$ la correlación es escasa

Observando la nube de puntos obtendríamos la misma conclusión



5 **Vamos a buscar la relación entre la densidad de algunas sustancias expresada en (g/cm^3) y su calor específico expresado en cal/g. °C. Calcula el coeficiente de correlación lineal.**

Sustancia	Densidad	Calor específico
Agua	1	1
Hielo	0,92	0,5
Aluminio	2,69	0,217
Hierro	7,85	0,113
Plata	10,42	0,057
Oro	19,3	0,032
Plomo	11,005	0,031
Níquel	8,8	0,106
Cinc	7,1	0,093
Cobre	8,9	0,092
Estaño	7,29	0,054
Hidrógeno	0,07	3,4

$r = -0{,}57$ débil y negativa

6 **Considerar la serie estadística bidimensional**

x	–2	–1	0	1	23
y	–7	–4	–1	2	58

Calcular:

a) Las rectas de regresión.

b) El coeficiente de correlación.

c) Interpretar el resultado.

a) recta de regresión de y sobre x

$y = 2{,}58x - 1{,}24$

recta de regresión de x sobre y

$x = 0{,}39y + 0{,}48$

b) $r = 0{,}99$. El coeficiente nos indica que la correlación es muy fuerte y positiva

7 **Idem, para la serie:**

$$(x_i , y_j) = \{(1, 4), (2, 5), (4, 3), (2, 0), (5, 4)\}$$

a) $y = 0{,}11x + 2{,}89$

$x = 0{,}08y + 2{,}54$

b) $r = 0{,}09$ la correlación es escasa o nula

8 **Dada la distribución bidimensional:**

x	1	2	3	4	5
y	4	7	10	13	16

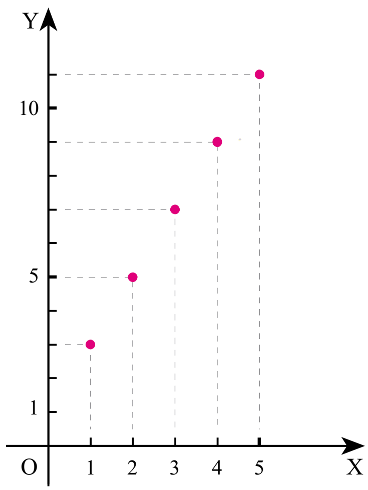
a) Dibujar la nube de puntos.

b) Comprobar que $S_x^2 = 2$; $S_y^2 = 18$; $S_{xy} = 6$.

c) Comprobar que las dos rectas de regresión coinciden.

d) ¿Puede hablarse de una relación funcional?

a)



b) $\bar{x} = 3$; $s_x^2 = 2$; $\bar{y} = 7$; $s_y^2 = 8$; $s_{xy} = 25 - 21 = 4$

c) $y = 2x + 1$; $2x - y + 1 = 0$ recta de regresión de y sobre x

$x = \dfrac{y-1}{2}$; $2x - y + 1 = 0$ recta de regresión de x sobre y

d) Si, la recta es $y = 2x + 1$

9 **Dada la distribución:**

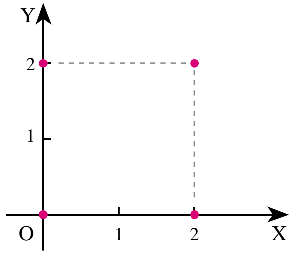
x	0	0	2	2
y	0	2	0	2

a) Dibujar la nube de puntos.

b) Comprobar que $S_x^2 = S_y^2 = 1$ y que $S_{xy} = 0$.

c) Comprobar la perpendicularidad de las rectas de regresión.

a)



b) $s_x^2 = s_y^2 = 1;\ s_{xy} = 0$

c) recta de regresión de y sobre x, $y = 1$

recta de regresión de x sobre y, $x = 1$

10 **La distribución de edades y presión arterial de 10 personas es:**

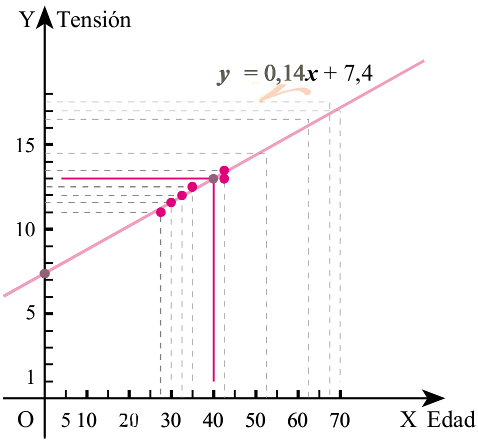
Edad (X)	30	28	35	42	51	42	63	32	70	67
Tensión (Y)	11,5	11,3	12,5	13,5	14,6	13	16,6	12	16,9	17

a) Representar la nube de puntos. ¿Se puede proceder a un ajuste lineal?

b) Calcular el coeficiente de correlación lineal.

c) Prever la tensión de una persona de sesenta años.

a)



Si se puede proceder a un ajuste lineal

b) $r = 0,99$

c) Con 60 años la tensión será 15,8

En las series estadísticas de los ejercicios 11 y 12 se pide:

a) Dibujar la nube de puntos.

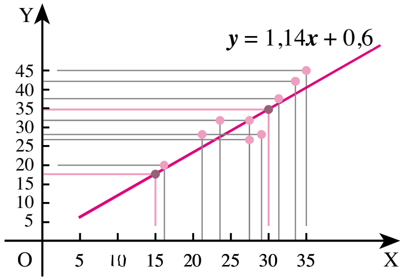
b) Calcular la recta de regresión de *Y* sobre *X* y representarla en la nube de puntos.

c) Calcular el coeficiente de correlación y comentar el resultado.

11

16	28	23	24	28	29	31	34	35
20	26	28	32	32	28	36	41	45

a)



Si se puede proceder a un ajuste lineal

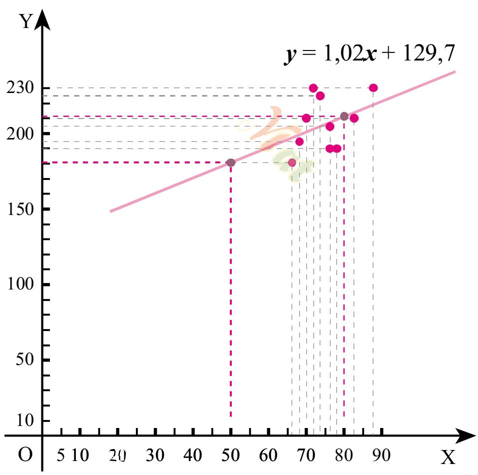
b) $y = 1,14x + 0,6$

c) $r = 0,87$ es fuerte y positiva

12

68	76	78	70	74	72	78	66	86	82
195	205	190	210	225	230	190	180	230	205

a)



b) $y = 1,02x + 129,7$

c) $r = 0,36$ muy débil la correlación existente

13 **La tabla adjunta proporciona las tasas de variación interanual del Índice de Precios al Consumo y los miles de parados registrados en las oficinas del INEM, de los doce meses del año 2006:**

	Tasa IPC	Parados		Tasa IPC	Parados
Enero	4,2	2171,5	Julio	4	1955,0
Febrero	4	2169,3	Agosto	3,7	1983,7
Marzo	3,9	2148,5	Septiembre	2,9	1966,2
Abril	3,9	2075,7	Octubre	2,5	1992,8
Mayo	4	2004,5	Noviembre	2,6	2023,2
Junio	3,9	1959,8	Diciembre	2,7	2022,9

Valora, con el coeficiente de correlación lineal, el grado de dependencia lineal entre la tasa del IPC y el número de parados.

$r = 0{,}399 \Rightarrow$ la relación es débil y en sentido creciente.

14 **Las tasas de variación mensual e interanual del Índice de Precios al Consumo correspondientes a los seis primeros meses del año 2007 se recogen en la tabla siguiente:**

	Variación mensual	Variación anual
Enero	–0,7	2,4
Febrero	0,1	2,4
Marzo	0,8	2,5
Abril	1,4	2,4
Mayo	0,3	2,3
Junio	0,2	2,4

Se pide:

a) Las medidas aritméticas y las desviaciones típicas de ambas variaciones.

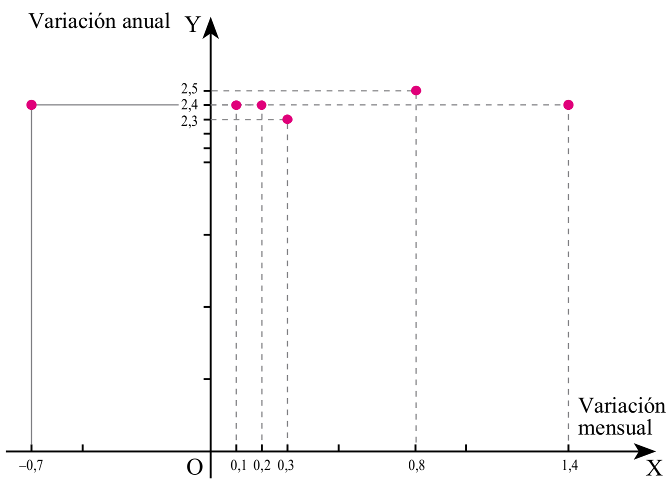
b) El coeficiente de correlación lineal entre las dos variaciones.

c) Relacionar el coeficiente de correlación anterior con la nube de puntos.

a) 0,35 y 0,645; 2,4 y 0,058

b) $r = 0{,}223$ relación débil y positiva

c)



La nube de puntos puede representarse por una recta paralela al eje de abcisas, y se confirma la débil relación lineal de las variaciones mensuales y anuales del IPC en el primer semestre del año 2007.

15 **La nota media del expediente (x) y la nota obtenida de pruebas de acceso a la Universidad (y) de ocho alumnos elegidos al azar fueron:**

x	6,24	7,91	7,04	6,13	6,38	6,48	6,44	5,99
y	4,20	4,65	6,51	6,73	5,20	4,60	5,69	3,42

Se pide:

a) Obtener el coeficiente de correlación entre ambas variables e interpretar el resultado.

b) Calcular la recta de regresión de y sobre x.

c) Según el ajuste que ofrece la recta de regresión, ¿qué nota sería esperable que sacara en las pruebas de acceso un alumno con nota media de expediente 8,31?

a) $r = 0,1$ escasa correlación

b) $y = 0,2x + 3,8$

c) $y = 5,46$

16 **Considera la serie estadística bidimensional**

x	2	4	5	5	8	9
y	3	3	1	3	6	4
n	2	1	3	4	2	3

a) Representa la nube de puntos.

b) Calcula el coeficiente de correlación e indica qué significa el valor obtenido.

c) Calcula las rectas de regresión y represéntalas en la nube de puntos.

$r = 0,557$; $y = 1,13 + 0,36x$; $x = 3 + 0,858$

17 **Una empresa dispone de los datos de la tabla**

número de vendedores	3	4	5	8	10
número de pedidos	90	110	140	190	235

Estimar el número de pedidos que obtendrían 9 vendedores. Indica el método utilizado en el cálculo de la estimación y la fiabilidad de esta estimación.

9 vendedores tendrían un número de pedidos estimado de 214. Se ha utilizado la recta de regresión $y = 20,26x + 31,45$.

Con un coeficiente $r = 0,9$, dependencia muy fuerte.

18 **Una empresa tiene los datos de la tabla**

mill. en publicidad	1	2	3	4	5	6	7	8
ventas	15	16	14	18	21	19	19	21

Estima las ventas esperadas al invertir 10 millones en publicidad.

Explica la fiabilidad de la estimación realizada. Los datos de ventas de la tabla son también en millones.

Ventas esperadas con 10 millones de publicidad serán 22,77 millones

$r = 0{,}83$, es una correlación fuerte

19 **Construye, razonadamente, dos distribuciones bidimensionales que tengan coeficientes de correlación cercanos a 0 y a 1.**

x	2	3	5	6
y	5	8	15	19

$r \Rightarrow 0{,}99$

x	2	3	5	6
y	1	8	1	2

$r \Rightarrow -0{,}27$

20 **Se ha realizado un estudio sobre las preferencias de las ratas con respecto a la temperatura del agua. Un grupo de 10 ratas fue sometido a dos temperaturas del agua diferentes y se midió el tiempo de permanencia en este medio. Estos fueron los resultados.**

26°C	6	7	8	5	9	3	6	4	7	2
30°C	1	8	3	10	4	7	9	2	4	6

Calcular la covarianza y el coeficiente de correlación. Analizar qué tipo de dependencia existe entre las variables.

$s_{xy} = -1{,}28$; $r = -0{,}21$

Existe una dependencia escasa

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

$x = 0{,}4y - 3$; x = peso en kg; y = estatura en cm;

$y = 180$ cm $\rightarrow$ peso = 69 kg

22 **La siguiente tabla ofrece los resultados de 6 pares de observaciones realizadas para analizar el grado de relación existente entre dos variables X e Y.**

X	2	2	3	3	3	4
Y	0	1	1	2	4	3

Obtener:

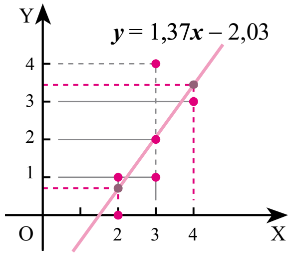
a) Recta de regresión de Y sobre X.

b) Representación gráfica de la misma, así como de los pares de observaciones anteriores.

c) ¿Qué grado de relación lineal existe entre ambas variables?

a) $y = 1{,}37x - 2{,}03$

b)



c) El coeficiente de correlación $r = 0{,}71$ positiva y media.

23 **Calcula la recta de regresión correspondiente a la distribución siguiente**

Altura sobre el nivel del mar	0	184	231	481	730	911	1550
Presión atmosférica	760	745	740	720	700	685	650

¿Qué presión atmosférica habría sobre Peña Vieja (2.600 metros de altitud aproximadamente)?

$y = -0{,}07x + 756{,}40$.

Presión atmosférica sobre Peña Vieja

$x = 2.600$; $y = 574{,}4$

24 **¿Qué significa que en una distribución bidimensional el coeficiente de correlación sea 0? ¿Y que sea –1?**

$r = 0 \rightarrow$ las variables son linealmente independientes y no están correlacionadas

$r = -1 \rightarrow$ la dependencia lineal es funcional, pero inversa

25 **En una muestra fiable de una determinada región se han obtenidos los siguientes datos:**

AÑO	1994	1995	1996	1997	1998	1999
X	13	20	23	25	27	31
Y	80	80	90	100	110	110

AÑO	2000	2001	2002	2003	2004	2005
X	36	46	55	63	70	76
Y	120	160	180	190	200	210

(*X* = nº de telespectadores en miles. *Y* = nº de enfermos mentales).

Calcular el coeficiente de correlación.

¿Se puede inferir de este resultado que la televisión altera la salud mental de los telespectadores? ¿Se puede deducir que en esa región compran televisiones una vez que se han vuelto locos?

$r = 0{,}99$. Son dos variables que aunque su correlación sea alta no tienen ningún tipo de dependencia.

26 **Una compañía desea hacer predicciones del valor anual de sus ventas totales en cierto país a partir de la relación entre éstas y la renta nacional. Para investigar la relación cuenta con los siguientes datos:**

X	189	190	208	227	239	252	257	274	293	308	316
Y	402	404	412	425	429	436	440	447	458	469	469

donde X representa la renta nacional en millones de dólares e Y representa las ventas de la compañía en miles de dólares en el periodo desde 1995 hasta 2005 (ambos inclusive).

Se pide:

a) Obtener la recta de regresión de Y sobre X. Brevemente: ¿qué representa esta recta?

b) Calcular el coeficiente de correlación lineal entre X e Y e interpretarlo.

c) En 2006 se espera que la renta nacional del país sea 325 millones de dólares. ¿Cuál será la predicción para las ventas de la compañía en este año?

a) $y = 0{,}53x + 301{,}67$

b) $r = 0{,}998$ es una correlación muy fuerte y positiva

c) $y = 473{,}92$ millones de dólares, siendo $x = 325$ millones de dólares

27 **Se han medidos los pesos y tallas de 5 personas, obteniéndose los siguientes datos:**

Peso (x) kg	50				
Talla (y) cm	150	160	170	180	190

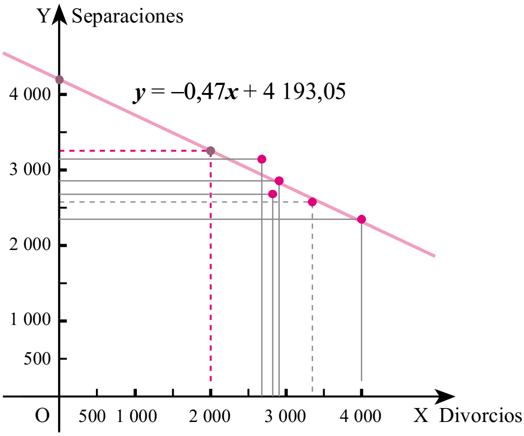
Calcular la recta de regresión de *y* (talla) sobre *x* (peso). ¿Cuál es la talla esperada para una persona que pese 62 kg?

$y = 3{,}57\,x - 30{,}18$ si $x = 62$kg; $y = 191{,}16$ cm

28 **Durante cinco años, los Juzgados de Madrid tramitaron los siguientes casos de separaciones y divorcios:**

Separaciones	2 357	2 586	2 689	3 073	2 821
Divorcios	4 000	3 428	2 903	2 711	2 910

Representar la nube de puntos de esta tabla y calcular el coeficiente de correlación lineal de las dos variables (número de separaciones y número de divorcios). Interpretar el resultado. Ajustando una recta de regresión lineal, ¿cuántas separaciones se prevé que se produzcan en un determinado año, sabiendo que hubo 3.600 divorcios? Dibujar la recta ajustada.



$r = -0,92$ es una correlación fuerte y negativa

$y = -0,47x + 4193,05$ recta de regresión

$x = 3600$ divorcios; $y = 2501$ separaciones

29 **Una asociación dedicada a la protección de la infancia desea estudiar la relación entre la mortalidad infantil en cada país y el número de camas de hospital por cada mil habitantes. Para ello, posee los siguientes datos sobre diez países concretos que pueden considerarse representativos del resto:**

X	50	100	70	60	120	180	200	250	30	90
Y	5	2	2,5	3,75	4	1	1,25	0,75	7	3

donde X representa el número de camas por cada mil habitantes e Y el tanto por ciento de mortalidad infantil en el país correspondiente. Se pide:

a) Calcular razonadamente la media y la desviación típica de X.

b) Calcular razonadamente la media y la desviación típica de Y.

c) ¿Qué distribución está más dispersa? Razona la respuesta.

d) Calcular el coeficiente de correlación lineal e interpretarlo.

a) $\overline{x} = 115$ $s_x = 68,59$

b) $\overline{y} = 3,03$ $s_y = 1,87$

c) Está más dispersa «y», calculando los coeficientes de variación de ambos $CV(x) = 60\%$, $CV(y) = 62\%$

d) $r = -0,82$ es una correlación fuerte y negativa, luego al aumentar el nº de camas, disminuye el índice de mortalidad

30 **La creciente inclinación de la torre de Pisa ha generado numerosos estudios sobre su futura estabilidad. En la tabla siguiente se presentan las medidas de su inclinación durante los años 1978 a 1987. Los años se han codificado a las dos últimas cifras y los de la inclinación como décimas de mm por exceso de 2,9000 m, de forma que la inclinación en el año 1978, que fue de 2,9667 aparece en la tabla como 667.**

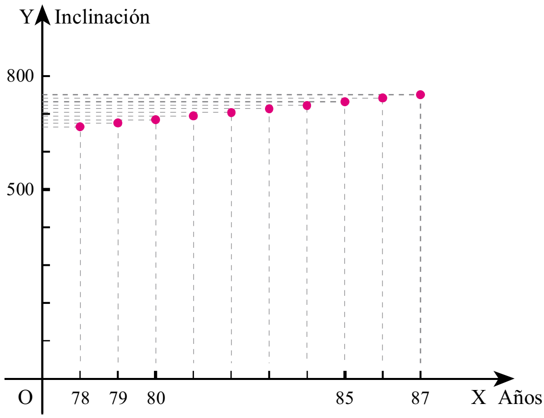
Año	78	79	80	81	82	83	84	85	86	87
Incl.	667	673	688	696	698	713	717	725	742	757

a) Representa gráficamente estos datos. ¿Crees que la inclinación de la torre tiene una tendencia lineal que crece con el tiempo?

b) Calcula la recta de mínimos cuadrados de la inclinación sobre el tiempo.

c) En 1918 la inclinación de la torre era de 2,9071 m. ¿Cuál sería el valor ajustado según la recta que has obtenido en el aparatado b)? ¿Cuál crees que es la causa de la diferencia entre ambos valores?

a)



b) $y = 9{,}45x - 71{,}64$

c) en 1918; $x = 18$, $y = 98{,}46$ luego la inclinación sería 2,9098 según la recta de mínimos cuadrados. La diferencia se debe a que la recta de mínimos cuadrados proporciona un valor esperado.

31 **Se han hecho dos pruebas de Historia a un grupo de 10 alumnos de 2º de ESO para valorar sus conocimientos. Los resultados obtenidos son:**

Alumno	1	2	3	4	5	6	7	8	9	10
A	14	12	15	12	13	12	17	7	9	14
B	14	13	17	15	16	12	22	10	14	20

Calcular la covarianza y el coeficiente de correlación. ¿Existe dependencia entre ambas pruebas?

$s_{xy} = 7{,}65$; $r = 0{,}81$

Existe una dependencia fuerte y positiva

32 **Cinco niñas, de 2, 3, 5, 7 y 8 años, pesan respectivamente, 11, 13, 21, 25y 30 kg.**

a) Hallar la ecuación de la recta de regresión del peso sobre las edades.

b) ¿Cuál sería el peso estimado para una niña de 6 años?

a) $y = 5{,}16x + 4{,}22$

b) $x = 6$ años; $y = 35{,}18$ kg

33 **Dada la distribución bidimensional (X, Y):**

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

a) Calcular el coeficiente de correlación lineal, interpretando el resultado.

b) Determinar la recta de regresión de Y sobre X.

c) Hallar el punto donde se cortan las dos rectas de regresión.

a) $r = 0,94$ es una correlación fuerte y positiva

b) $y = 0,8x + 1,23$

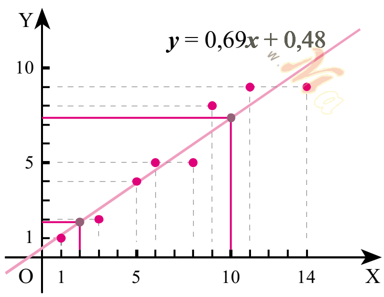
c) (4,82; 5,08)

34 **Calcular la ecuación de la recta de regresión correspondiente a la distribución:**

(1, 1), (3, 2), (5, 4), (6, 5), (8, 5), (9, 8), (11, 9), (14, 9)

Dibujar posteriormente la nube de puntos y la recta de regresión.

$y = 0,69x + 0,48$



35 **La tabla siguiente muestra el número de gérmenes patógenos (en miles por centímetro cúbico) de un determinado cultivo, según el tiempo transcurrido. Calcula una recta de regresión para predecir el número de gérmenes por centímetro cúbico en función del tiempo. ¿Qué cantidad de gérmenes por centímetro cúbico es predecible encontrar cuando hayan transcurrido seis horas? ¿Es buena esa predicción?**

Nº horas	0	1	2	3	4	5
Nº gérmenes	20	26	33	41	47	53

$y = 6,72x + 19,86$ recta de regresión

$x = 6$ horas, $y = 60,18$ gérmenes por cm^3

Es muy buena la predicción porque $r = 0,99$

36 **Se observaron las edades de 5 niños y sus pesos respectivos, obteniéndose los siguientes resultados:**

Edad, en años (x)	2	4,5	6	7,2	8
Peso, en Kg (y)	15	19	25	33	34

a) Hallar el coeficiente de correlación y las rectas de regresión de Y sobre X y de X sobre Y.

b) ¿Qué peso corresponderá a un niño de 5 años? ¿Qué edad corresponderá a un peso de 36 Kg?

a) $r = 0{,}97$

$y = 3{,}4x + 6{,}36$

$x = 0{,}27y - 1{,}38$

b) $x = 5$ años, $y = 23{,}36$ kg

$y = 36$ kg, $x = 8{,}34$ años

37 **Una Compañía Telefónica está interesada en efectuar previsiones sobre sus ingresos en los próximos años, para lo que toma como referencia la información disponible en una determinada zona sobre el número X de líneas en servicio y el total de ingresos en miles Y por operaciones, para los últimos 9 años:**

$\sum_i x_i = 8\,055$ $\qquad S_x = 110$

$\sum_i y_i = 336\,600$ $\qquad S_y = 1\,120$

$\sum_i x_i y_i = 302\,335\,110$

Obtener, a partir de esos datos, el coeficiente de correlación lineal e interpretar el resultado.

Calculando la recta de regresión de Y sobre X, ¿cuáles serán los ingresos esperados por la Compañía el próximo año, si el número de líneas en servicio va a ser de 1 200?

$r = 0{,}97$ la correlación es fuerte

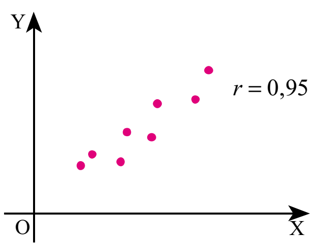
recta de regresión $y = 9{,}9x + 28539{,}5$

$x = 1200 \Rightarrow y = 40419{,}5$ ingresos

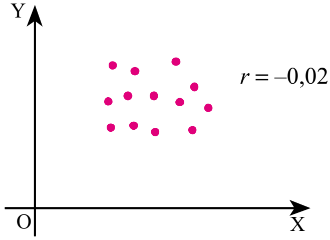
38 **Representa de forma aproximada una nube de puntos que corresponde a una distribución bidimensional según los siguientes valores del coeficiente de correlación lineal:**

a) r = 0,95; b) r = –0,02; c) r = –0,93

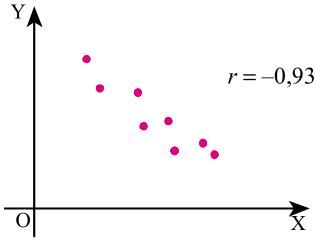
a)



b)



c)



39 **Justifica si es posible que las rectas de regresión de *y sobre x* y de *x sobre y* en una distribución bidimensional sean $2x + 3y = 7$; $2y - 3x = 22$.**

No son posibles, pues una de las rectas es creciente y la otra decreciente, mientras que las pendientes de las dos rectas de regresión deben tener el mismo signo.

40 **Calcula las pendientes de las dos rectas de regresión en una distribución bidimensional, sabiendo que $S_{xy} = -20$; $S_x^2 = 10$; $S_y^2 = 40$.**

Regresión Y sobre X: pendiente $= -2$

Regresión X sobre Y: pendiente $= -0{,}5$