Problema 1

La temperatura media anual y la latitud de las capitales de los países de la Comunidad Económica Europea es aproximadamente la siguiente:

Capitales	Temperatura (º C)	Latitud (º)
Amsterdam	13	54
Atenas	24	37
Bonn	13	52
Bruselas	14	52
Copenhague	11	54
Dublín	13	53
Lisboa	19	39
Londres	14	53
Luxemburgo	14	50
Madrid	19	40
París	15	49
Roma	22	42

a) Calcular las medias, las varianzas y la covarianza.

b) Estudiar si existe algún tipo de correlación entre ambas variables.

c) Analizar el tipo de dependencia entre ambas variables

d) Calcular la ecuación de la recta de regresión de la latitud respecto de la temperatura media.

e) Calcular la ecuación de la recta de regresión de la temperatura media respecto de la latitud.

f) ¿Cuál será la temperatura media esperada, o prevista, para una ciudad que se encuentra situada a 47º de latitud norte?

g) ¿Cuál será la latitud esperada para una ciudad cuya temperatura media es de 20 grados centígrados?

h) Hacer una representación gráfica de los datos mediante una nube de puntos y mediante un diagrama de columnas tridimensional.

i) Representar gráficamente las rectas de regresión de Y sobre X y de X sobre Y en el diagrama de dispersión o nube de puntos.

a)

x_i	y_i	f_i	x_if_i	y_if_i	$x_i^2f_i$	$y_i^2f_i$	$x_iy_if_i$
13	54	1	13	54	169	2916	702
24	37	1	24	37	576	1369	888
13	52	1	13	52	169	2704	676
14	52	1	14	52	196	2704	728
11	54	1	11	54	121	2916	594
13	53	1	13	53	169	2809	689
19	39	1	19	39	361	1521	741
14	53	1	14	53	196	2809	742
14	50	1	14	50	196	2500	700
19	40	1	19	40	361	1600	760
15	49	1	15	49	225	2401	735
22	42	1	22	42	484	1764	924
		12	191	575	3223	28013	8879

$$\bar{x} = \frac{191}{12} = 15'9167$$

$$\bar{y} = \frac{575}{12} = 47'9167$$

$$s_x^2 = \frac{3223}{12} - 15'9167^2 = 15'241994$$

$$s_x = 3'90424$$

$$s_y^2 = \frac{28013}{12} - 47'9167^2 = 38'406527$$

$$s_y = 6'19756$$

$$s_{xy} = \frac{8879}{12} - 15'9167 \cdot 47'9167 = -22'7569$$

b)

$$r = \frac{-22'7569}{3'90424 \cdot 6'19756} = -0'94049686$$

Al ser el coeficiente de correlación negativo, quiere decir que la correlación es inversa, lo que significa que a mayor temperatura menos latitud y recíprocamente a mayor latitud menos temperatura.

Por ser próximo a -1 la correlación es bastante fuerte.

c) La dependencia que existe entre las variables es de tipo aleatorio fuerte; esto quiere decir que, si bien cuando se conoce una latitud no es posible dar con total exactitud su temperatura media, sí es posible aproximarse con un error relativamente pequeño.

d) La ecuación de la recta de regresión de la latitud respecto de la temperatura media es:

$$y - \bar{y} = \frac{s_{xy}}{s_x^2}(x - \bar{x}) \quad \rightarrow \quad y - 47'9167 = \frac{-22'7569}{15'241994} \cdot (x - 15'9167) \quad \rightarrow \quad y = -1'49x + 71'67$$

e) La ecuación de la recta de regresión de la temperatura media respecto de la latitud es:

$$x - \bar{x} = \frac{s_{xy}}{s_y^2}(y - \bar{y}) \quad \rightarrow \quad x - 15'9167 = \frac{-22'7569}{38'406527} \cdot (y - 47'9167) \quad \rightarrow \quad x = -0'59\,y + 44'30$$

f) La temperatura media esperada para una ciudad que se encuentra situada a 47º de latitud norte se obtiene sustituyendo la y por 47 en la ecuación de la recta de regresión de x sobre y, es decir:

$$\hat{x} = -0'59 \cdot 47 + 44'30 = 16'57$$

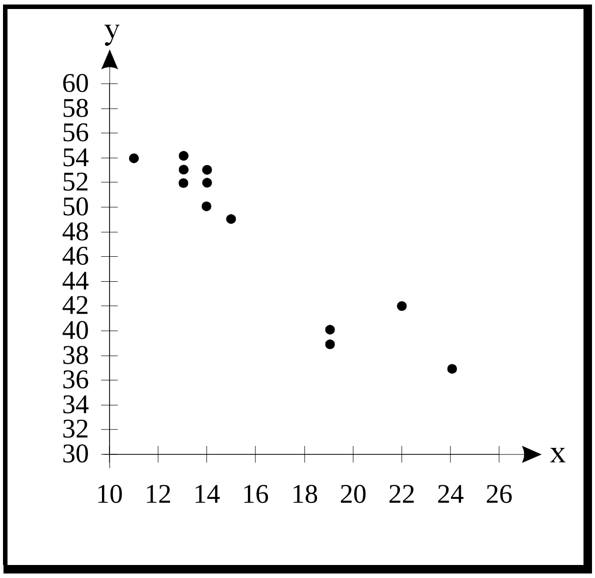
Es decir, se prevé que la temperatura media anual sea de 16'5 grados centígrados aproximadamente.

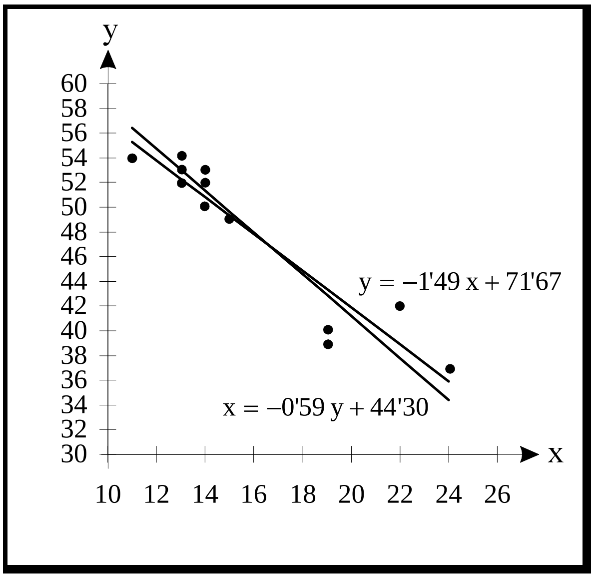
g) Sustituyendo la x por 20 en la recta de regresión de y sobre x tenemos:

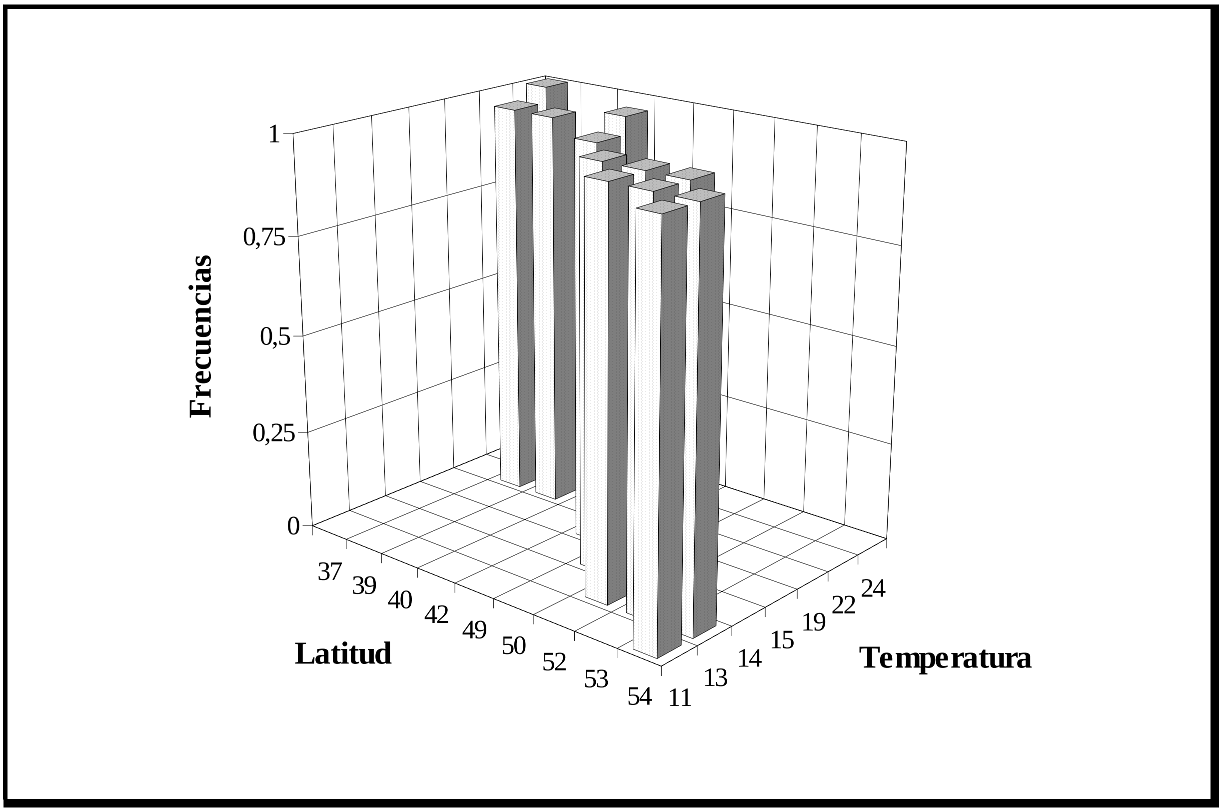
$$\hat{y} = -1'49 \cdot 20 + 71'67 = 41'87$$

Es decir, se prevé que la latitud sea de 42º aproximadamente.

h) , i)







Problema 2

Se han clasificado 50 familias con arreglo al número de hijos e hijas, obteniéndose los siguientes resultados:

Número de hijos x_i	Número de hijas y_i	Número de familias f_i
0	0	2
0	1	3
0	3	1
1	2	6
1	3	4
2	0	4
2	1	9
2	4	2
3	0	3
3	2	6
3	5	1
4	0	1
4	3	2
4	4	1
5	1	3
5	3	1
6	2	1

a) **Calcular las medias, las varianzas y la covarianza.**

b) **Estudiar si existe algún tipo de correlación entre ambas variables y analizar el tipo de dependencia entre ellas.**

c) **Calcular la ecuación de la recta de regresión de Y/X y de X/Y**

d) **¿Cuál será el número esperado de hijas para una familia que tenga 8 hijos? ¿Y el número esperado de hijos para una familia que tenga 6 hijas?**

e) **Hacer una representación gráfica de los datos mediante una nube de puntos y mediante un diagrama de columnas tridimensional.**

f) **Representar gráficamente las rectas de regresión de Y sobre X y de X sobre Y en el diagrama de dispersión o nube de puntos.**

a)

x_i	y_i	f_i	$x_i f_i$	$y_i f_i$	$x_i^2 f_i$	$y_i^2 f_i$	$x_i y_i f_i$
0	0	2	0	0	0	0	0
0	1	3	0	3	0	3	0
0	3	1	0	3	0	9	0
1	2	6	6	12	6	24	12
1	3	4	4	12	4	36	12
2	0	4	8	0	16	0	0
2	1	9	18	9	36	9	18
2	4	2	4	8	8	32	16
3	0	3	9	0	27	0	0
3	2	6	18	12	54	24	36
3	5	1	3	5	9	25	15
4	0	1	4	0	16	0	0
4	3	2	8	6	32	18	24
4	4	1	4	4	16	16	16
5	1	3	15	3	75	3	15
5	3	1	5	3	25	9	15
6	2	1	6	2	36	4	12
		50	112	82	360	212	191

$$\bar{x} = \frac{112}{50} = 2'24$$

$$\bar{y} = \frac{82}{50} = 1'64$$

$$s_x^2 = \frac{360}{50} - 2'24^2 = 2'1824$$

$$s_x = 1'4772$$

$$s_y^2 = \frac{212}{50} - 1'64^2 = 1'5504$$

$$s_y = 1'2451$$

$$s_{xy} = \frac{191}{50} - 2'24 \cdot 1'64 = 0'1464$$

b) El coeficiente de correlación lineal es:

$$r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{0'1464}{1'4772 \cdot 1'2451} = 0'0795$$

lo que significa que la correlación existente entre dichas variables es positiva, pero irrelevante, por ser el valor de r muy próximo a 0. En este caso se dice que las variables están incorrelacionadas, y en consecuencia, se puede afirmar que ambas variables son aleatoriamente independientes.

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

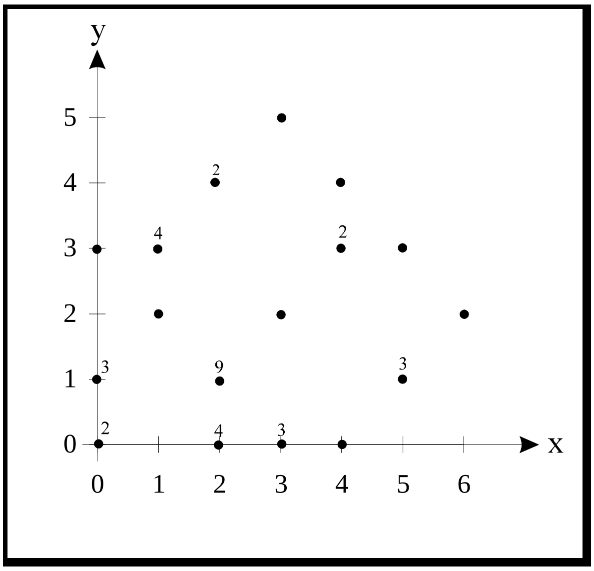
$$y - \bar{y} = \frac{s_{xy}}{s_x^2}(x - \bar{x}) \quad \rightarrow \quad y - 1'64 = \frac{0'1464}{2'1824} \cdot (x - 2'24) \quad \rightarrow \quad y = 0'06x + 1'48$$

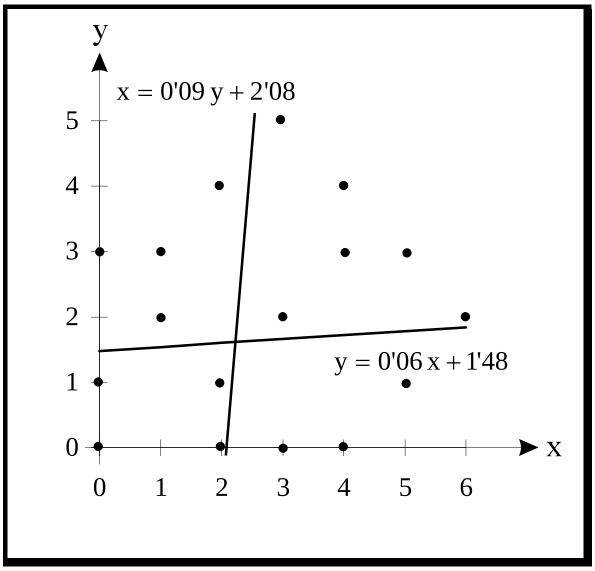
La ecuación de la recta de regresión de X sobre Y:

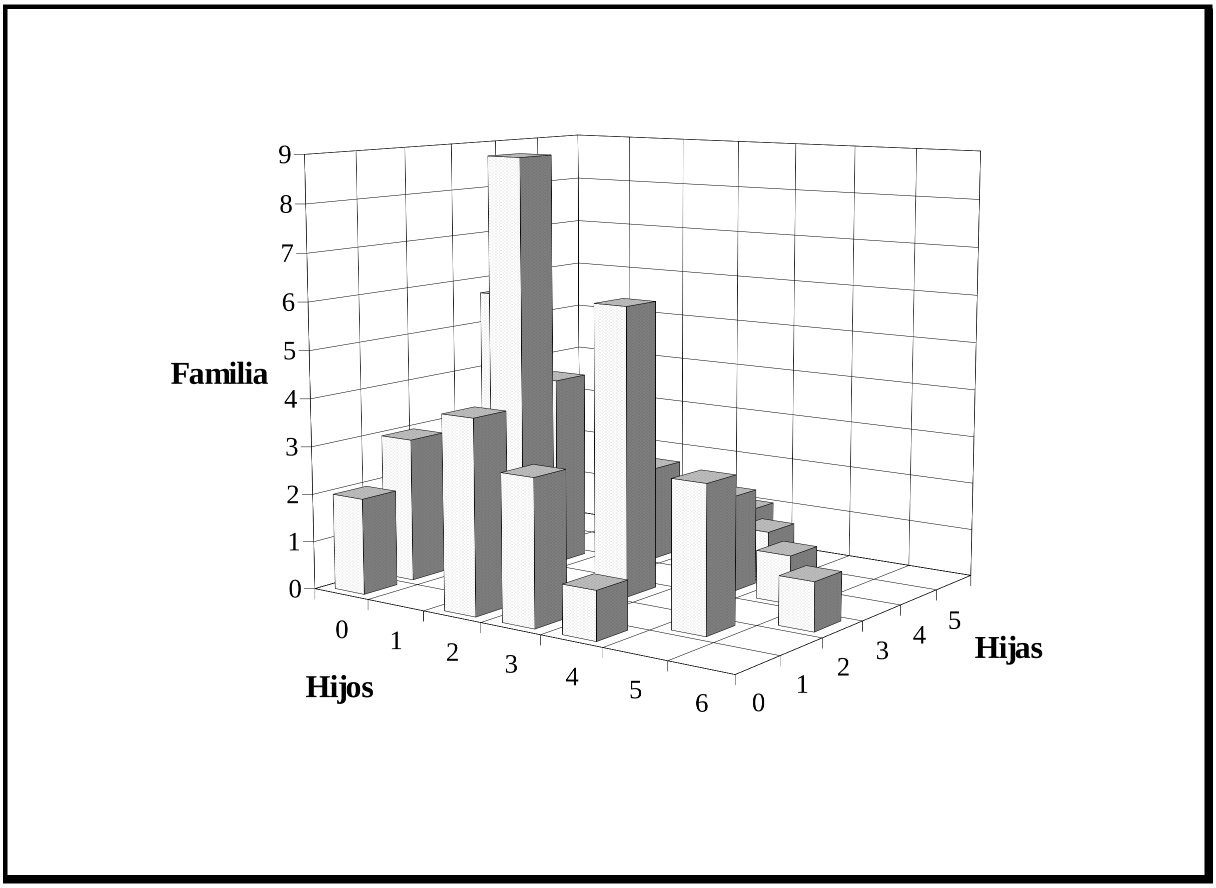
$$x - \bar{x} = \frac{s_{xy}}{s_y^2}(y - \bar{y}) \quad \rightarrow \quad x - 2'24 = \frac{0'1464}{1'5504} \cdot (y - 1'64) \quad \rightarrow \quad x = 0'09y + 2'08$$

d) Al ser la correlación lineal casi nula no tiene sentido calcular este apartado.

e) , f)







Problema 3

Consideremos la distribución dada por la tabla adjunta, donde x_i representa la edad e y_i el grado de psicomotricidad de un grupo de 44 minusválidos mentales:

y_i ↓ x_i →	$[5-7)$	$[7-9)$	$[9-11)$	$[11-13)$	
$[25-30)$	**4**	**3**	**–**	**1**	**8**
$[30-35)$	**2**	**7**	**2**	**–**	**11**
$[35-40)$	**1**	**1**	**11**	**1**	**14**
$[40-45)$	**–**	**2**	**–**	**6**	**8**
$[45-50)$	**–**	**–**	**–**	**3**	**3**
	7	**13**	**13**	**11**	**44**

a) Calcular las medias, las varianzas y la covarianza.

b) Estudiar si existe algún tipo de correlación entre ambas variables y analizar el tipo de dependencia entre ellas.

c) Calcular la ecuación de la recta de regresión de Y/X y de X/Y

d) ¿Cuál será el grado esperado de psicomotricidad para un minusválido cuya edad esté comprendida entre 13 y 15 años? ¿Cuál será la edad esperada de un minusválido cuyo grado de psicomotricidad sea de 62?

e) Hacer una representación gráfica de los datos mediante una nube de puntos y mediante un diagrama de columnas tridimensional o estereograma.

f) Representar gráficamente las rectas de regresión de Y sobre X y de X sobre Y en el diagrama de dispersión o nube de puntos.

a)

x_i	y_i	f_i	$x_i f_i$	$y_i f_i$	$x_i^2 f_i$	$y_i^2 f_i$	$x_i y_i f_i$
6	27'5	4	24	110	144	3025	660
6	32'5	2	12	65	72	2112'5	390
6	37'5	1	6	37'5	36	1406'25	225
8	27'5	3	24	82'5	192	2268'75	660
8	32'5	7	56	227'5	448	7393'75	1820
8	37'5	1	8	37'5	64	1406'25	300
8	42'5	2	16	85	128	3612'5	680
10	32'5	2	20	65	200	2112'5	650
10	37'5	11	110	412'5	1100	15468'75	4125
12	27'5	1	12	27'5	144	756'25	330
12	37'5	1	12	37'5	144	1406'25	450
12	42'5	6	72	255	864	10837'5	3060
12	47'5	3	36	142'5	432	6768'75	1710
		44	408	1585	3968	58575	15060

$$\bar{x} = \frac{408}{44} = 9'27$$

$$\bar{y} = \frac{1585}{44} = 36'02$$

$$s_x^2 = \frac{3968}{44} - 9'27^2 = 4'1983$$

$$s_x = 2'0489$$

$$s_y^2 = \frac{58575}{44} - 36^2 = 33'6131$$

$$s_y = 5'7976$$

$$s_{xy} = \frac{15060}{44} - 9'27 \cdot 36'02 = 8'2438$$

b) El coeficiente de correlación lineal es

$$r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{8'2438}{2'0489 \cdot 5'7976} = 0'6939$$

lo que significa que la correlación existente entre dichas variables es positiva y aproximadamente del 70%.

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

$$y - \overline{y} = \frac{s_{xy}}{s_x^2}(x - \overline{x}) \quad \rightarrow \quad y - 36'02 = \frac{8'2438}{4'1983} \cdot (x - 9'27) \quad \rightarrow \quad y = 1'96x + 17'81$$

La ecuación de la recta de regresión de X sobre Y:

$$x - \overline{x} = \frac{s_{xy}}{s_y^2}(y - \overline{y}) \quad \rightarrow \quad x - 9'27 = \frac{8'2438}{33'6131} \cdot (y - 36'02) \quad \rightarrow \quad x = 0'24y + 0'43$$

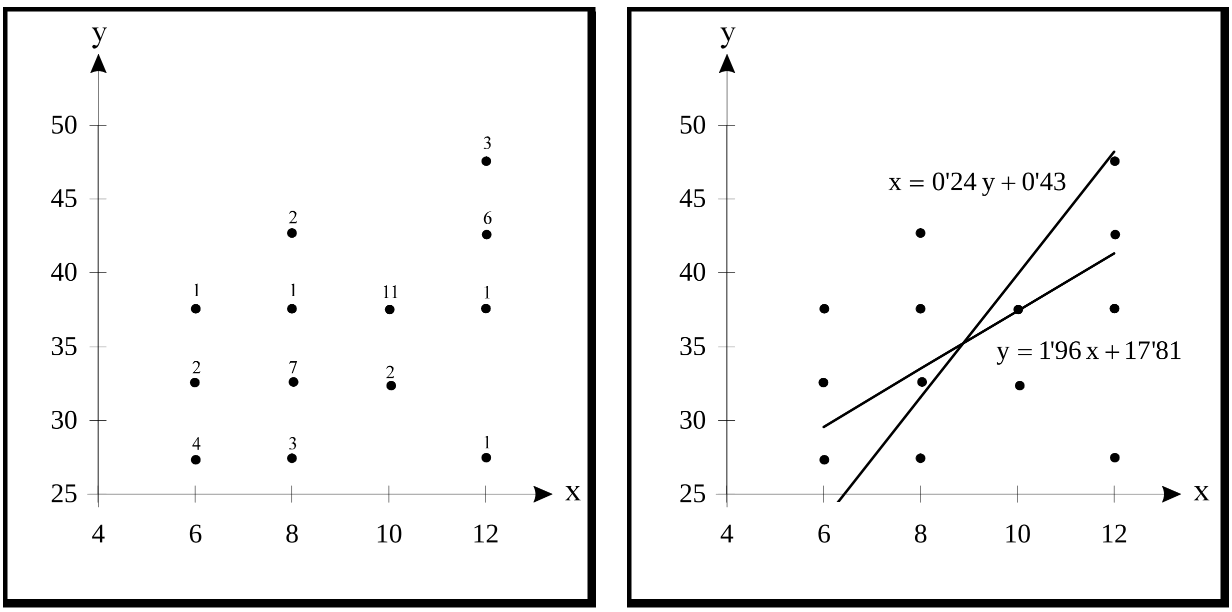
d) El grado de psicomotricidad esperado para un minusválido cuya edad esté comprendida entre 13 y 15 años será:

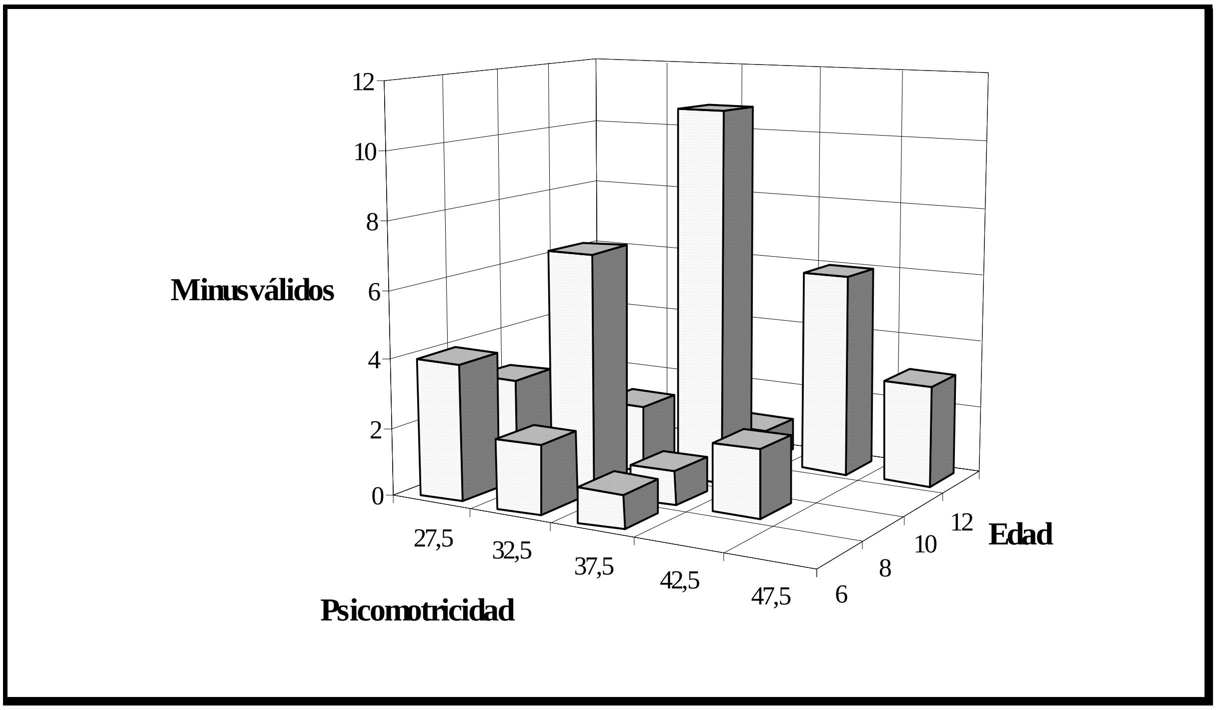
$$\hat{y} = 1'96 \bullet 14 + 17'81 = 45'25 \approx 45$$

La edad esperada para un minusválido cuya psicomotricidad sea de 62 será:

$$\hat{x} = 0'24 \cdot 62 + 0'43 = 15'31 \approx 15$$

e) , f)





Problema 4

Explíquese si cabe esperar correlación lineal, positiva, negativa o nula, en cada uno de los siguientes pares de valores:

a) La edad de un determinado modelo de coche y su precio en el mercado de ocasión.

b) El peso de una persona y su coeficiente de inteligencia.

Vaya por delante, que, sin una constatación "seria", basada en la toma de datos y posteriores cálculos, toda afirmación apriorística es cuando menos, muy discutible. Hecha esta observación, diremos que:

a) El precio de un coche usado decrece con la edad. Su depreciación no tiene porqué ser la misma cantidad cada año, es decir, el descenso del valor del coche no es lineal con respecto al tiempo. No obstante, parece que debe existir una fuerte correlación lineal y, por supuesto, negativa, pues el precio decrece al aumentar la edad del coche.

b) Entre personas adultas, no es de esperar correlación lineal entre las dos variables, o, dicho de otro modo, es de esperar que exista correlación nula. Si consideramos niños de distintas edades, es posible que encontremos correlación lineal positiva entre ambas variables, pero ello podría ser debido al siguiente "factor externo": En época de crecimiento, aumenta el peso y también el coeficiente de inteligencia.

Problema 5

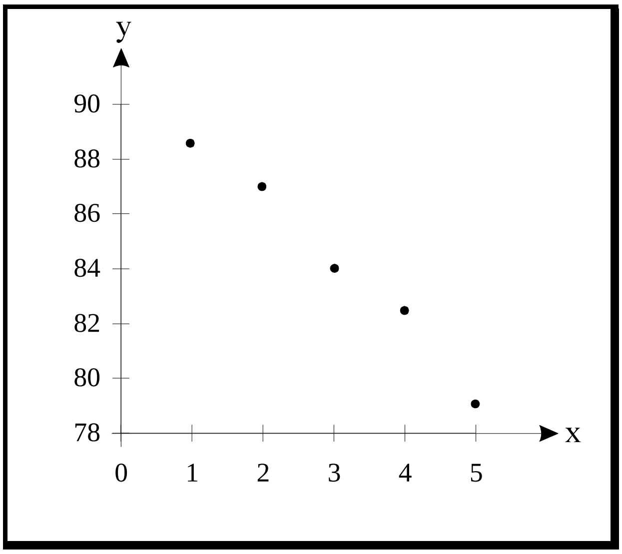
Una persona se somete a una dieta de adelgazamiento durante 5 semanas. A continuación se detalla su peso al término de cada una de las semanas:

Semanas	1	2	3	4	5
Peso en kg	88'5	87	84	82'5	79

a) ¿Es razonable suponer que existe correlación lineal entre el peso y la dieta? ¿Cómo puede expresarse esta correlación?

b) ¿Qué peso es de esperar que alcance esa persona si sigue la dieta durante dos semanas más? ¿Y si hace la dieta durante veinticinco semanas?

a)



Si representamos el diagrama de dispersión de estos datos tenemos la figura adjunta.

Parece indicar la existencia de correlación lineal negativa y fuerte. Comprobémoslo calculando el coeficiente de correlación lineal.

Llamaremos x_i a la variable "número de semanas de dieta" e y_i a la variable "peso". Al haber pocos datos, se facilitan los cálculos disponiéndolos como sigue:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	88'5	1	7832'25	88'5
2	87	4	7569	174
3	84	9	7056	252
4	82'5	16	6806'25	330
5	79	25	6241	395
15	421	55	35504'5	1239'5

$$\overline{x} = \frac{15}{5} = 3 \qquad \overline{y} = \frac{421}{5} = 84'2$$

$$s_x = \sqrt{\frac{55}{5} - 3^2} = \sqrt{2}$$

$$s_y = \sqrt{\frac{35504'5}{5} - 84'2^2} = \sqrt{11'26}$$

$$s_{xy} = \frac{1239'5}{5} - 3 \cdot 84'2 = -4'7$$

Consecuentemente:

$$r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{-4'7}{\sqrt{2} \cdot \sqrt{11'26}} = -0'990$$

lo que demuestra la existencia de correlación lineal, fuerte y negativa, que puede expresarse mediante la recta de regresión de Y sobre X, que es:

$$y - \bar{y} = \frac{s_{xy}}{s_x^2}(x - \bar{x}) \qquad \rightarrow \qquad y - 84'2 = -2'35 \cdot (x - 3) \qquad \rightarrow \qquad y = -2'35x + 91'25$$

b) Tomando $x = 7$ en la ecuación de la recta de regresión, se obtiene una buena estimación del peso que alcanzará esa persona con dos semanas más de dieta: $y = 74'8$ kg.

Sin embargo, no es de esperar que la correlación entre semanas de dieta y peso se mantenga por mucho tiempo. Por eso **no tiene sentido utilizar la fórmula con $x = 25$** para obtener el peso presumible tras 25 semanas de dieta. Tras ese tiempo, o bien la persona ha estabilizado su peso, o bien ha enfermado gravemente, si no es que muere.

Problema 6

En una nube de puntos, (x_i, y_i) con $i = 1,2,\ldots\ldots\ldots,20$ se sabe que el valor de y_i es siempre no superior a 2 y que las x_i son positivas. Si la recta de regresión tiene por ecuación $y = 2 + mx$ ¿Cuáles son los posible valores de m? Dar, además, para el valor máximo de m los valores de los y_i.

Al ser 2 la ordenada en el origen de la recta de regresión, la única posibilidad es que la pendiente m sea un número real negativo para que la recta se ajuste a la nube de puntos. Como el valor máximo de "m" es 0, los valores de los y_i serán todos iguales a 2.

Problema 7

En una distribución bidimensional (X,Y), si el coeficiente de correlación lineal r es próximo a -1 ¿puede decir el efecto sobre la variable Y de un incremento positivo en la variable X? Además, ¿puedes dar los valores de r para los cuales Y es exactamente una función lineal de X? Justifica tus respuestas.

Si el coeficiente de correlación lineal es próximo a -1 significa que al aumentar la variable X disminuye la variable Y y recíprocamente, es decir, al disminuir la variable X aumenta la variable Y, lo que significa que la correlación es inversa.

Si Y es una función lineal de X, el coeficiente de correlación lineal es o bien 1 o bien -1, es decir, todos los puntos de la distribución se encuentran situados sobre una recta.

Problema 8

El número de matrimonios y nacimientos (en miles) en España ha evolucionado como se indica en las tablas siguientes:

x_i (Años)	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983
y_i (Mat.)	268	271'3	261	262	258'1	246'3	220'7	202	193'3	196'2

x_i (Años)	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983
y_i (Nac.)	687'7	669,4	677,5	656'4	636'9	602	571	533	515'7	485'4

Tomando como variable x_i el tiempo y como variable y_i el número de matrimonios o el número de nacimientos, obtener las ecuaciones de las rectas de regresión de Y sobre X, y los coeficientes de correlación; mediante las rectas de regresión obtenidas, ¿qué estimación se podría realizar para los años 1984 y 1988? Comentar los resultados.

Si calculamos las rectas de regresión tanto para los matrimonios como para los nacimientos obtenemos:

Matrimonios

$$\bar{x} = 1978'5 \qquad \bar{y} = 237'89 \qquad s_x = 2'872281 \qquad s_y = 29'8785 \qquad s_{xy} = -81'145$$

$$r = -0'94 \qquad\qquad y = -9'83576x + 19697'94$$

Para $x = 1984$ tendremos $\hat{y} = -9'83576 \cdot 1984 + 19697'94 = 183'79 \rightarrow \hat{y} = 183.790$ matrimonios

Para $x = 1988$ tendremos $\hat{y} = -9'83576 \cdot 1988 + 19697'94 = 144'45 \rightarrow \hat{y} = 144.450$ matrimonios

Nacimientos

$$\bar{x} = 1978'5 \qquad \bar{y} = 603,5 \qquad s_x = 2'872281 \qquad s_y = 69'6474 \qquad s_{xy} = -195'51$$

$$r = -0'97 \qquad\qquad y = -23'6982x + 47490'35$$

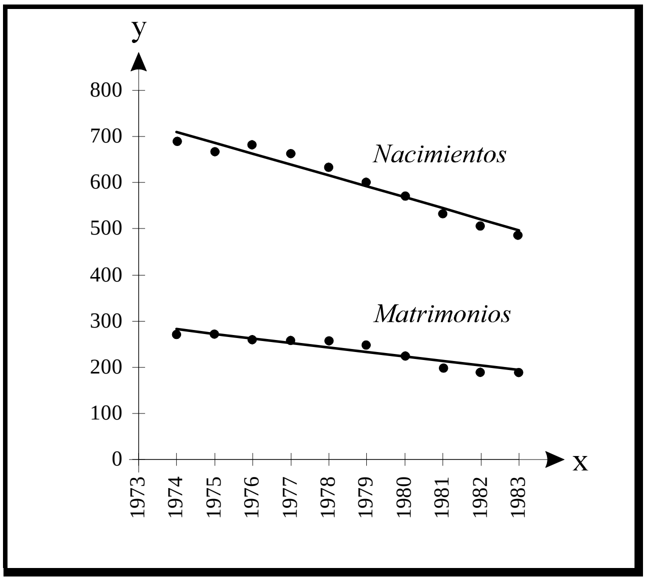
Para $x = 1984$ tendremos $\hat{y} = -23'6982 \cdot 1984 + 47490'35 = 473'12 \rightarrow \hat{y} = 473.120$ nacimientos

Para $x = 1988$ tendremos $\hat{y} = -23'6982 \cdot 1988 + 47490'35 = 378'32 \rightarrow \hat{y} = 378.320$ nacimientos

Conclusión

Los resultados obtenidos indican un fuerte descenso en el número de matrimonios celebrados, aproximadamente 9835 menos cada año ya que entre 1984 y 1988 el número de matrimonios ha descendido de 83.790 a 144.450, es decir, 39.340.

Por otra parte se aprecia que el descenso en el número de nacimientos es más fuerte que el de matrimonios, ya que descienden de 473.120 a 378.320 es decir 94.800 lo que significa una media de 23.700 menos cada año.



Los resultados de la extrapolación deben ser tomados con suma cautela, pues es evidente que esta tendencia a la baja no puede prolongarse indefinidamente.

Problema 9

En una muestra fiable de una determinada región se han obtenido los siguientes datos:

Año	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979
x	13	20	23	25	27	31	36	46	55	63	70	76
y	80	80	90	100	110	110	120	160	180	190	200	210

(x = nº de telespectadores en miles. y = nº de enfermos mentales)

a) Calcular el coeficiente de correlación.

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

a)

x	y	xy	x^2	y^2
13	80	1040	169	6400
20	80	1600	400	6400
23	90	2070	529	8100
25	100	2500	625	10000
27	110	2970	729	12100
31	110	3410	961	12100
36	120	4320	1296	14400
46	160	7360	2116	25600
55	180	9900	3025	32400
63	190	11970	3969	36100
70	200	14000	4900	40000
76	210	15960	5776	44100
485	1630	77100	24495	247700

$$\bar{x} = \frac{485}{12} = 40'41 \qquad \bar{y} = \frac{1630}{12} = 135'83$$

$$s_x = \sqrt{\frac{24495}{12} - 40'41^2} = 20'20$$

$$s_y = \sqrt{\frac{247700}{12} - 135'83^2} = 46'81$$

$$s_{xy} = \frac{77100}{12} - 40'41 \cdot 135'83 = 936'10$$

$$r = \frac{936'10}{20'20 \cdot 46'81} = 0'989$$

b) Vemos que el coeficiente de correlación es muy alto, lo que significa que hay una relación fuerte entre esas dos variables. Pero de ello no debemos deducir relaciones de *causa-efecto* entre ellas, como en tono de broma se nos dice en el enunciado. La razón hay que buscarla en variables intermedias de las cuales dependen ambas variables estudiadas. Por ejemplo, pueden ser:

- Aumento del número de habitantes de la región: *a más habitantes, más telespectadores y más enfermos mentales.*
- Subida del nivel de vida: *más televisores y más enfermos mentales detectados y atendidos.*

Problema 10

Una compañía discográfica ha recopilado la siguiente información sobre el nº de conciertos dados, durante el verano, por 15 grupos musicales y las ventas de discos de estos grupos (expresados en miles de LPs), obteniendo los siguientes datos:

LPs \ Conciertos	10 – 30	30 – 40	40 – 80
1 – 5	3	0	0
5 – 10	1	4	1
10 – 20	0	1	5

a) Calcular el nº medio de LPs vendidos por estos grupos.

b) ¿Cómo es el grado de dependencia lineal del número de conciertos dado por un grupo con respecto al nº de discos que ha vendido.

c) Obtener la recta de regresión que explica la dependencia anterior.

d) Si un grupo musical ha vendido 18.000 LPs, ¿qué número de conciertos es previsible que dé?

a) En cada uno de los 15 grupos musicales se da la relación entre las dos variables:

$x \rightarrow$ nº de discos vendidos $\qquad$ $y \rightarrow$ nº de conciertos dados

Puesto que las variables vienen dadas en intervalos, convendría reinterpretar la tabla poniendo las marcas de clase de cada intervalo, para propiciar los cálculos necesarios y obtener los parámetros que se nos piden.

x \ y	20	35	60
3	3	0	0
7'5	1	4	1
15	0	1	5

x_i	y_i	f_i	$x_i f_i$	$y_i f_i$	$x_i^2 f_i$	$y_i^2 f_i$	$x_i y_i f_i$
3	20	3	9	60	27	1200	180
7'5	20	1	7'5	20	56'25	400	150
7'5	35	4	30	140	225	4900	1050
7'5	60	1	7'5	60	56'25	3600	450
15	35	1	15	35	225	1225	525
15	60	5	75	300	1125	18000	4500
		15	144	615	1714'5	29325	6855

$$\bar{x} = \frac{144}{15} = 9'6 \quad \Rightarrow \quad \text{venden por término medio 9.600 LPs}$$

b) Hemos de obtener el coeficiente de correlación.

$$s_x = \sqrt{\frac{1714'5}{15} - 9'6^2} = 4'71 \qquad \bar{y} = \frac{615}{15} = 41 \qquad s_y = \sqrt{\frac{29325}{15} - 41^2} = 16'6$$

$$s_{xy} = \frac{6855}{15} - 9'6 \cdot 41 = 63'4 \qquad r = \frac{s_{xy}}{s_x s_y} = \frac{63'4}{4'71 \cdot 16'6} = 0'81$$

El valor del coeficiente de correlación nos indica que la dependencia entre las dos variables es alta.

c) $\quad m_{yx} = \dfrac{s_{xy}}{s_x^2} = \dfrac{63'4}{22'14} = 2'86 \quad \rightarrow \quad y - 41 = 2'86(x - 9'6) \quad \rightarrow \quad y = 2'86x + 13'5$

d) $\quad \hat{y}(18) = 2'86 \cdot 18 + 13'5 = 64'98 \cong 65$ $\quad$ Se prevé un número de conciertos próximo a 65.