

# 13 Distribuciones bidimensionales

## ACTIVIDAD INICIAL

13.I. Considera la población formada por los alumnos de tu clase. Para esta población da tres ejemplos de:

- Caracteres estadísticos cualitativos, indicando algunas de sus modalidades.
- Variables estadísticas discretas, indicando varios valores posibles.
- Variables estadísticas continuas, indicando varios posibles valores.

a) Color de los ojos: azul, marrón, negro.

Género de películas favorito: romántico, bélico, de acción.

Optativa que cursa: Francés, Informática.

b) Número de hermanos: 0, 1, 2, 3.

Número de materias aprobadas en la primera evaluación. 0, 1, 2, 3...

Día del cumpleaños: 1, 2, 3... 31.

c) Peso. A partir de 48 kilos.

Distancia del domicilio al instituto.

Perímetro craneal.

## EJERCICIOS PROPUESTOS

13.1. (PAU) Las edades de un grupo de 19 jóvenes aparecen en la siguiente tabla.

Edad	14	15	17	18	19	20	21
N.º de jóvenes	3	1	2	3	5	3	2

- Calcula la media, la moda y la mediana.
- Halla el rango, la varianza y la desviación típica.

Formamos la siguiente tabla:

$x_i$	$f_i$	$F_i$	$x_i f_i$	$x_i^2 f_i$
14	3	3	42	588
15	1	4	15	225
17	2	6	34	578
18	3	9	54	972
19	5	14	95	1805
20	3	17	60	1200
21	2	19	42	882
	19		342	6250

a) Media:  $\bar{x} = \frac{342}{19} = 18$  años

Moda:  $M_o = 19$  años

Mediana: La mitad de los datos es  $\frac{19}{2} = 9,5$ ; por tanto,  $M = 19$  años.

b) Rango:  $r = 21 - 14 = 7$  años

Varianza:  $s^2 = \frac{6250}{19} - 18^2 = 4,95$  años<sup>2</sup>

Desviación típica:  $s = 2,22$  años

13.2. (PAU) Una variable estadística viene dada por la siguiente tabla.

$x_i$	[10-15)	[15-20)	[20-25)	[25-30)	[30-35)
$p_i$	12	20	28	16	8

a) Halla la media, la moda y la mediana.

b) Calcula el rango, la varianza y la desviación típica.

Formamos la siguiente tabla:

$[L_i, L_{i+1})$	$x_i$	$f_i$	$F_i$	$x_i f_i$	$x_i^2 f_i$
[10, 15)	12,5	12	12	150	1875
[15, 20)	17,5	20	32	350	6125
[20, 25)	22,5	28	60	630	14 175
[25,30)	27,5	16	76	440	12 100
[30, 35)	32,5	8	84	260	8450
		84		1830	42 725

a) Media:  $\bar{x} = \frac{1830}{84} = 21,79$

La clase modal es [20, 25); por tanto,  $Mo = 22,5$ .

La mitad de los datos es  $\frac{84}{2} = 42$ ; por tanto, la clase mediana es [20, 25) y la mediana es  $M = 22,5$ .

b) Rango:  $r = 35 - 10 = 25$

Varianza:  $s^2 = \frac{42\,725}{84} - 21,79^2 = 33,83$

Desviación típica:  $s = \sqrt{33,83} = 5,82$

13.3. La siguiente tabla muestra las calificaciones obtenidas por cinco alumnos en Bachillerato (X) y en la PAU (Y).

Bachillerato	5,4	6,8	5,3	7,4	4,3
PAU	5,8	4,8	5,9	7,4	4,2

A partir de ella, calcula:

a) Las medias y varianzas de X y de Y.

b) La covarianza de (X, Y).

Formamos la tabla:

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$
5,4	5,8	29,16	33,64	31,32
6,8	4,8	46,24	23,04	32,64
5,3	5,9	28,09	34,81	31,27
7,4	7,4	54,76	54,76	54,76
5,3	4,2	28,09	17,64	18,06
30,2	28,1	186,34	163,89	172,25

a)  $\bar{x} = \frac{30,2}{5} = 6,04$        $s_x^2 = \frac{186,34}{5} - 6,04^2 = 0,79$

$\bar{y} = \frac{28,1}{5} = 5,62$        $s_y^2 = \frac{163,89}{5} - 5,62^2 = 1,19$

b)  $s_{xy} = \frac{172,25}{5} - 6,04 \cdot 5,62 = 0,505$

13.4. En un depósito cilíndrico la altura del agua que contiene varía conforme pasa el tiempo según la siguiente tabla:

Tiempo (h)	8	22	27	33	50	70
Altura (m)	17	14	12	11	6	1

Halla:

a) Las medias de X y de Y.

b) Las varianzas de X y de Y.

c) La covarianza de (X, Y).

Formamos la siguiente tabla:

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$
8	17	64	289	136
22	14	484	196	308
27	12	729	144	324
33	11	1089	121	363
50	6	2500	36	300
70	1	4900	1	70
210	61	9766	787	1501

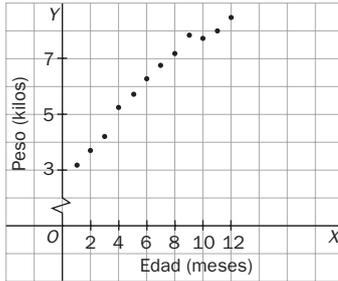
a)  $\bar{x} = \frac{210}{6} = 35$        $\bar{y} = \frac{61}{6} = 10,17$

b)  $s_x^2 = \frac{9766}{6} - 35^2 = 402,67$        $s_y^2 = \frac{787}{6} - 10,17^2 = 27,74$

c)  $s_{xy} = \frac{1501}{6} - 35 \cdot 10,17 = -105,78$

13.5. La tabla adjunta expresa los valores de la variable bidimensional altitud, en metros, y presión, en milímetros, de mercurio de un lugar. Representa la nube de puntos de esta variable e indica la relación existente entre la altitud y la presión.

<b>Altitud (m)</b>	0	184	231	481	730	911	1550
<b>Presión (mm Hg)</b>	760	745	740	720	700	685	650

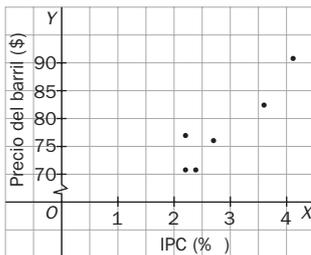


Según se aprecia en el diagrama de dispersión, existe una relación lineal negativa funcional entre la altitud y la presión.

13.6. En la siguiente tabla se recoge la evolución del IPC (índice de precios al consumo) y el precio del barril de petróleo (*brent*) durante el segundo semestre de 2007.

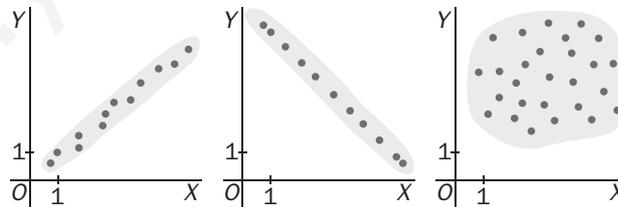
<b>IPC (%)</b>	2,4	2,2	2,2	2,7	3,6	4,1
<b>Precio del barril (\$)</b>	71,54	77,01	70,73	76,87	82,50	90,16

¿Se puede asegurar que la evolución del IPC está directamente relacionada con el precio del petróleo?



Sí, existe una correlación lineal positiva fuerte.

13.7. Los números 0, 0,8 y 1 son los valores absolutos del coeficiente de correlación de las distribuciones bidimensionales cuyas nubes de puntos adjuntamos:



Asigna a cada diagrama su coeficiente de correlación, cambiando el signo cuando sea necesario.

Primero: 0,8      Segundo: -1      Tercero: 0

13.8. (PAU) Las puntuaciones en Matemáticas y Física de siete alumnos han sido las siguientes:

<b>Matemáticas</b>	8	8	6	7	8	6	2
<b>Física</b>	7	7,5	5	7	7,5	5	7

- Halla el coeficiente de correlación de las calificaciones en Matemáticas y Física de los seis primeros alumnos.
- Calcula el coeficiente de correlación de esas dos variables para los siete alumnos.
- Explica la diferencia entre los resultados obtenidos.

a) Formamos la siguiente tabla con los seis primeros alumnos:

$x_i$	$y_i$	$f_i$	$x_i f_i$	$x_i^2 f_i$	$y_i f_i$	$y_i^2 f_i$	$x_i y_i f_i$
8	7	1	8	64	7	49	56
8	7,5	2	16	128	15	112,5	120
6	5	2	12	72	10	50	60
7	7	1	7	49	7	49	49
		6	43	313	39	260,5	285

Medias:  $\bar{x} = 7,167$ ,  $\bar{y} = 6,5$

Varianzas:  $s_x^2 = 0,806$ ,  $s_y^2 = 1,167$

Covarianza:  $s_{xy} = 0,9167$

Coefficiente de correlación:

$$r = \frac{s_{xy}}{s_x s_y} = \frac{0,9167}{0,94} = 0,946$$

b) De manera análoga para los siete alumnos:

$x_i$	$y_i$	$f_i$	$x_i f_i$	$x_i^2 f_i$	$y_i f_i$	$y_i^2 f_i$	$x_i y_i f_i$
8	7	1	8	64	7	49	56
8	7,5	2	16	128	15	112,5	120
6	5	2	12	72	10	50	60
7	7	1	7	49	7	49	49
2	7	1	2	4	7	49	14
		7	45	317	46	309,5	299

Medias:  $\bar{x} = 6,43$ ,  $\bar{y} = 6,57$

Varianzas:  $s_x^2 = 3,959$ ,  $s_y^2 = 1,031$

Covarianza:  $s_{xy} = 0,4694$

Coefficiente de correlación:

$$r = \frac{s_{xy}}{s_x s_y} = \frac{0,4694}{2,02} = 0,232$$

c) Para el conjunto total de los alumnos se observa una correlación directa, pues  $r > 0$  en los dos casos considerados. Sin embargo, en el primer caso, correspondiente a los seis primeros alumnos, la correlación es alta ( $r = 0,97$ ) y las calificaciones en Matemáticas y Física están en dependencia aleatoria directa fuerte, mientras que al considerar el total de los alumnos (segundo caso) la correlación es muy baja ( $r = 0,23$ ), y las calificaciones en Matemáticas y Física son aleatoriamente independientes.

13.9. (PAU) En cierto país, el tipo de interés y el índice de la Bolsa en los últimos seis meses vienen dados por la siguiente tabla.

Tipo de interés (%)	8	7,5	7,2	6	5,5	5
Índice	120	130	134	142	150	165

Halla el índice previsto de la Bolsa en el séptimo mes, suponiendo que el tipo de interés en ese mes fue del 4,1%, y analiza la fiabilidad de la predicción, según el valor del coeficiente de correlación.

Formamos la tabla:

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$
8	120	64	14 400	960
7,5	130	56,25	16 900	975
7,2	134	51,84	17 956	964,8
6	142	36	20 164	852
5,5	150	30,25	22 500	825
5	165	25	27 225	825
39,2	841	263,34	119 145	5401,8

$$\bar{x} = \frac{39,2}{6} = 6,533$$

$$\bar{y} = \frac{841}{6} = 140,167$$

$$s_x^2 = \frac{263,34}{6} - 6,533^2 = 1,21$$

$$s_x = \sqrt{1,21} = 1,10$$

$$s_y^2 = \frac{119 145}{6} - 140,167^2 = 210,71$$

$$s_y = \sqrt{210,71} = 14,52$$

$$s_{xy} = \frac{540,18}{6} - 6,533 \cdot 140,167 = -15,41 = -15,41$$

$$r = \frac{s_{xy}}{s_x s_y} = -0,965$$

A partir de estos datos se obtiene la recta de regresión:  $y = -12,763x + 223,37$ . Sustituyendo  $x$  por 4,1 se obtiene un valor para el índice bursátil de, aproximadamente, 171. Dado el valor de  $r$  que indica una correlación lineal negativa fuerte, y la cercanía del valor 4,1 al rango de datos utilizados en el ajuste, se puede considerar que la predicción es fiable.

13.10. (PAU) Como consecuencia de un estudio estadístico realizado sobre 100 universitarios se ha obtenido una estatura media de 155 cm, con una desviación típica de 15,5 cm. Además se obtuvo la recta de regresión (siendo  $x$  el peso e  $y$  la altura).

Determina el peso medio de estos 100 universitarios.

Las rectas de regresión se cortan en el punto  $(\bar{x}, \bar{y})$ :  $155 = 80 + 1,5\bar{x} \Rightarrow \bar{x} = \frac{155 - 80}{1,5} = 50$  kilos.

13.11. (PAU) Un estudio sociológico proporcionó la siguiente tabla.

Nivel de estudios	1	2	3	4	5
Salario medio (€)	800	1000	1500	2000	3000

a) Calcula el coeficiente de correlación lineal entre el nivel de estudios y el salario medio,  $y$ , en función del valor obtenido, explica si se puede considerar que el salario medio está determinado por el nivel de estudios.

1 = estudios primarios                      2 = estudios secundarios                      3 = formación profesional  
 4 = técnicos de grado medio              5 = técnicos superiores                      6 = doctores

b) Deduce el salario esperado para el nivel de estudios 6.

Formamos la tabla:

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$
1	800	1	640 000	800
2	1000	4	1 000 000	2000
3	1500	9	2 250 000	4500
4	2000	16	4 000 000	8000
5	3000	25	9 000 000	15 000
15	8300	55	16 890 000	30 300

$$\begin{aligned}
 \text{a) } \bar{x} &= \frac{15}{5} = 3 & s_x^2 &= \frac{55}{5} - 3^2 = 2 & s_x &= 1,41 \\
 \bar{y} &= \frac{8300}{5} = 1660 & s_y^2 &= \frac{16\,890\,000}{5} - 1660^2 = 622\,400 \\
 s_y &= 788,92 & s_{xy} &= \frac{30\,300}{5} - 3 \cdot 1660 = 1080 \\
 r &= \frac{s_{xy}}{s_x s_y} = \frac{1080}{1,41 \cdot 788,92} = 0,97
 \end{aligned}$$

Se puede considerar que el salario es en función del nivel de estudios.

b) A partir de los datos obtenidos en a), se obtiene la recta de regresión  $y = 540x + 40$ . Sustituyendo ahora  $x$  por 6 obtendríamos una estimación del salario de 3280 €.

13.12. (PAU) El coeficiente de determinación de una distribución cuya nube de puntos se ajusta a una recta es igual a 0,33.

a) Interpreta este resultado.

b) ¿Tiene sentido encontrar un modelo lineal para esta distribución que permita realizar estimaciones?

a) Como  $r^2 = 0,33$ , en la distribución que estamos tratando, el 33% de la variación de  $Y$  es debido a la variación de  $X$ , mientras que el 67% restante de la variación de  $Y$  se debe al azar o a la influencia sobre  $Y$  de otras variables distintas de  $X$ .

b) Que el coeficiente de determinación sea  $r^2 = 0,33$  implica que el coeficiente de correlación es  $r = \pm 0,57$ , y esto nos indica una dependencia aleatoria media-baja. Así, el modelo lineal solo tendrá sentido a la hora de realizar estimaciones en puntos muy cercanos al par  $(\bar{x}, \bar{y})$ .

13.13. (PAU) Si el coeficiente de correlación vale  $r = 0,7$ :

a) ¿Qué tanto por ciento de la variación de  $Y$  es debido a la variación de  $X$  usando un modelo de regresión lineal?

b) ¿Tiene sentido realizar estimaciones en la recta de regresión obtenida?

a) El coeficiente de determinación para esta distribución es  $r^2 = 0,49$ ; por tanto, un 49% de la variación de  $Y$  es debido a la variación de  $X$ .

b) En este caso, el coeficiente de correlación es  $r = 0,7$ , por lo que esta distribución presenta una dependencia aleatoria intermediafuerte, y las estimaciones realizadas en la recta de regresión solo tendrán sentido si se hacen para puntos cercanos al centro de gravedad de la distribución.

13.14. Sea la variable bidimensional dada por la siguiente tabla.

X	1	2	3	4	5	6	7	8	9
Y	5	6	8	11	1	13	14	14	17

- a) Halla la recta de Tukey.  
 b) Halla la recta de regresión de Y sobre X.  
 c) Representa la nube de puntos y las dos rectas obtenidas.

a) Dividimos el conjunto de datos en los grupos:

$$G_1 = \{(1, 5) (2, 6) (3, 8)\} \quad G_2 = \{(4, 11) (5, 1) (6, 13)\} \quad G_3 = \{(7, 14) (8, 14) (9, 17)\}$$

Mediana de las abscisas de  $G_1$ :  $x_1 = 2 \Rightarrow P_1(2, 6)$

Mediana de las ordenadas de  $G_1$ :  $y_1 = 6$

Mediana de las abscisas de  $G_2$ :  $x_2 = 5 \Rightarrow P_2(5, 11)$

Mediana de las ordenadas de  $G_2$ :  $y_2 = 11$

Mediana de las abscisas de  $G_3$ :  $x_3 = 8 \Rightarrow P_3(8, 14)$

Mediana de las ordenadas de  $G_3$ :  $y_3 = 14$

Baricentro del triángulo  $P_1, P_2, P_3$ :  $G\left(\frac{2+5+8}{3}, \frac{6+11+14}{3}\right) = \left(5, \frac{31}{3}\right)$

Pendiente de la recta que pasa por  $P_1$  y  $P_3$ :  $m = \frac{14-6}{8-2} = \frac{8}{6} = \frac{4}{3}$ .

Recta de Tukey:  $y - \frac{31}{3} = \frac{4}{3}(x - 5) \Rightarrow y = \frac{4}{3}x + \frac{11}{3}$

b) Formamos la tabla:

$x_i$	$y_i$	$x_i^2$	$x_i y_i$
1	5	1	5
2	6	4	12
3	8	9	24
4	11	16	44
5	1	25	5
6	13	36	78
7	14	49	98
8	14	64	112
9	17	81	153
45	89	285	531

$$\bar{x} = \frac{45}{9} = 5$$

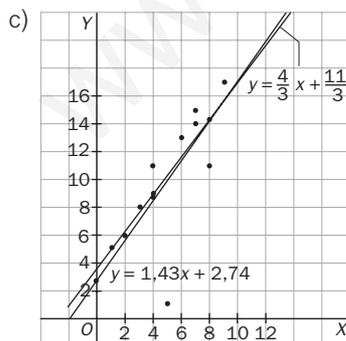
$$\bar{y} = \frac{89}{9} = 9,89$$

$$s_x^2 = \frac{285}{9} - 5^2 = 6,67$$

$$s_{xy} = \frac{531}{9} - 5 \cdot 9,89 = 9,55$$

La recta de regresión de Y sobre X es:

$$y - 9,89 = \frac{9,55}{6,67}(x - 5) \Rightarrow y = 1,43x + 2,74$$



13.15. La siguiente tabla da los datos obtenidos para una variable bidimensional.

X	1	2	3	4	5	6	7	8	9
Y	14	4	18	16	13	18	15	10	11

- a) Halla la recta de regresión de Y sobre X.
- b) Calcula la recta de Tukey.
- c) Representa la nube de puntos y las dos rectas obtenidas.

a) Formamos la tabla:

$x_i$	$y_i$	$x_i^2$	$x_i y_i$
1	14	1	14
2	4	4	8
3	18	9	54
4	16	16	64
5	13	25	65
6	18	36	108
7	15	49	105
8	10	64	80
9	11	81	99
45	119	285	597

$$\bar{x} = \frac{45}{9} = 5$$

$$\bar{y} = \frac{119}{9} = 13,22$$

$$s_x^2 = \frac{285}{9} - 5^2 = 6,67$$

$$s_{xy} = \frac{597}{9} - 5 \cdot 13,22 = 0,23$$

La recta de regresión de Y sobre X es:

$$y - 13,22 = \frac{0,23}{6,67}(x - 5) \Rightarrow y = 0,034x + 13,05$$

b) Formamos con los datos ordenados tres grupos:

$$G_1 = \{(1, 14) (2, 4) (3, 18)\}$$

$$G_2 = \{(4, 16) (5, 13) (6, 18)\}$$

$$G_3 = \{(7, 15) (8, 10) (9, 11)\}$$

Para cada grupo  $G_i$  hallamos el punto  $P_i(x_i, y_i)$ :

$$P_1(2, 14)$$

$$P_2(5, 16)$$

$$P_3(8, 11)$$

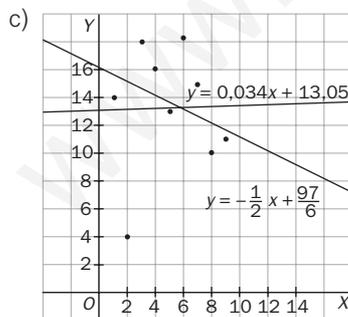
El baricentro del triángulo de vértices  $P_1 P_2 P_3$  tiene por coordenadas:

$$x_G = \frac{2 + 5 + 8}{3} = 5$$

$$y_G = \frac{14 + 16 + 11}{3} = \frac{41}{3}$$

La pendiente  $P_1 P_3$  es:  $m = \frac{11 - 14}{8 - 2} = -\frac{3}{6} = -\frac{1}{2}$ ; por tanto, la recta de Tukey es

$$y - \frac{41}{3} = -\frac{1}{2}(x - 5) \Rightarrow y = -\frac{1}{2}x + \frac{97}{6}$$



## EJERCICIOS

### Distribuciones unidimensionales

13.16. Dadas las series de datos:

A: 3, 5, 2, 7, 6, 4, 9

B: 3, 5, 2, 7, 6, 4, 9, 1

Halla para cada serie:

a) La media, la moda y la mediana.

b) El rango, la varianza y la desviación típica.

Formamos la siguiente tabla:

$x_i$	$f_i^A$	$f_i^B$	$x_i f_i^A$	$x_i f_i^B$	$x_i^2 f_i^A$	$x_i^2 f_i^B$
1	0	1	0	1	0	1
2	1	1	2	2	4	4
3	1	1	3	3	9	9
4	1	1	4	4	16	16
5	1	1	5	5	25	25
6	1	1	6	6	36	36
7	1	1	7	7	49	49
9	1	1	9	9	81	81
	7	8	36	37	220	221

a)  $\bar{A} = \frac{36}{7} = 5,143$ ,  $M_A = 5$ . No hay moda por ser la frecuencia absoluta de cada valor 0 ó 1.

$\bar{B} = \frac{37}{8} = 4,625$ ,  $M_B = 4,5$ . No hay moda por ser la frecuencia absoluta de cada valor 0 ó 1.

b)  $r_A = 9 - 2 = 7$ ,  $s_A^2 = \frac{220}{7} - 5,143^2 = 4,98$ ,  $s_A = 2,23$

$r_B = 9 - 1 = 8$ ,  $s_B^2 = \frac{221}{8} - 4,625^2 = 6,23$ ,  $s_B = 2,49$

13.17. Dada la distribución siguiente:

$x_i$	2	4	6	7	9
$f_i$	3	5	7	4	2

Calcula:

a) La media, la mediana y la moda.

b) El rango, la varianza y la desviación típica.

Formamos la siguiente tabla:

$x_i$	$f_i$	$F_i$	$x_i f_i$	$x_i^2 f_i$
2	3	3	6	12
4	5	8	20	80
6	7	15	42	252
7	4	19	28	196
9	2	21	18	162
	21		114	702

a)  $\bar{x} = \frac{114}{21} = 5,43$ ,  $Mo = 6$ ,  $M = 6$

b)  $r = 9 - 2 = 7$ ,  $s^2 = \frac{702}{21} - 5,43^2 = 3,96$ ,  $s = 1,99$

13.18. (PAU) Dada la distribución estadística:

$x_i$	[0, 5)	[5, 10)	[10, 15)	[15, 20)	[20, 25)	[25, 30)
$f_i$	3	5	7	8	2	5

Calcula la media, la mediana y la moda.

Formamos la siguiente tabla:

$[L_1, L_2)$	$x_i$	$f_i$	$F_i$	$x_i f_i$	$x_i^2 f_i$
[0, 5)	2,5	3	3	7,5	18,75
[5, 10)	7,5	5	8	37,5	281,25
[10, 15)	12,5	7	15	87,5	1093,75
[15, 20)	17,5	8	23	140	2450
[20, 25)	22,5	2	25	45	1012,5
[25, 30)	27,5	5	30	137,5	3781,25
		30		455	8637,5

$\bar{x} = \frac{455}{30} = 15,17$

La mitad de los datos es  $\frac{30}{2} = 15$ ; por tanto, la clase mediana es [10, 15), y la mediana se puede aproximar por la marca de clase correspondiente:  $M = 12,5$ .

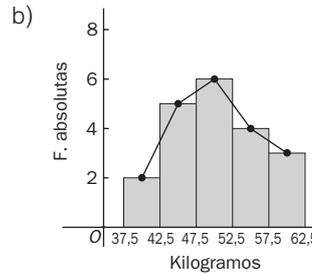
La clase modal es [15, 20); por tanto, la moda se puede aproximar por la marca de clase: será  $M_o = 17,5$ .

13.19. (PAU) Los pesos, en kg, de 20 estudiantes son: 51, 47, 55, 53, 49, 47, 48, 50, 43, 60, 45, 54, 62, 57, 46, 49, 52, 42, 38 y 61.

- a) Agrupa los datos en cinco clases de igual amplitud.
- b) Dibuja el histograma correspondiente.
- c) Halla la media de los datos agrupados.

a) Agrupamos los datos y formamos la tabla.

$[L_i, L_{s})$	$x_i$	$f_i$	$F_i$	$x_i f_i$
[37,5 - 42,5)	40	2	2	80
[42,5 - 47,5)	45	5	7	225
[47,5 - 52,5)	50	6	13	300
[52,5 - 57,5)	55	4	17	220
[57,5 - 62,5)	60	3	20	180
		20		1005



c)  $\bar{x} = \frac{1005}{20} = 50,25$

13.20. (PAU) En una encuesta sobre tráfico se ha preguntado a 1000 conductores sobre el número de multas recibidas, que en todos los casos es mayor o igual a cero y menor o igual a 5. Al efectuar la tabla correspondiente, algún número ha desaparecido, de forma que disponemos de la siguiente información:

N.º de conductores	?	260	150	190	100	90
N.º de multas	0	1	2	3	4	5

Calcula:

- a) La media.
- b) La mediana.
- c) La moda.
- d) La desviación típica.

El dato desconocido es:  $1000 - (260 + 150 + 190 + 100 + 90) = 210$ .

Formamos la tabla:

$x_i$	$f_i$	$F_i$	$x_i f_i$	$x_i^2 f_i$
0	210	210	0	0
1	260	470	260	260
2	150	620	300	600
3	190	810	570	1710
4	100	910	400	1600
5	90	1000	450	2250
	1000		1980	6420

a)  $\bar{x} = \frac{1980}{1000} = 1,98$

b)  $\frac{1000}{2} = 500. M = 2$

c)  $M_0 = 1$

d)  $s^2 = \frac{6420}{1000} - 1,98^2 = 2,4996 \quad s = 1,58$

13.21. La media de  $x, 3, 4x - 3, x + 4, -16, 9$  y  $x - 4$  es 4. ¿Cuál es la mediana de estos 7 números?

$$\bar{x} = \frac{x + 3 + 4x - 3 + x + 4 - 16 + 9 + x - 4}{7} = 4 \Rightarrow x = 5$$

Los números, ordenados de forma creciente, son:  $-16, 1, 3, 5, 9, 9, 17$ . La mediana es  $M = 9$ .

## Distribuciones bidimensionales. Correlación

13.22. ¿Qué significa que en una distribución bidimensional el coeficiente de correlación sea el que se indica en cada uno de los siguientes casos?

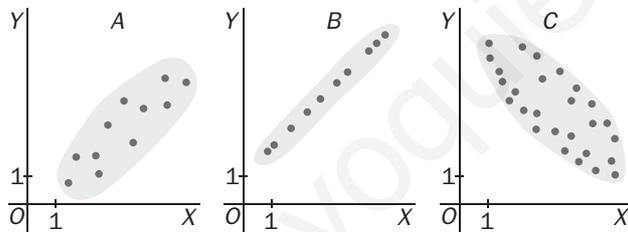
- a)  $r = 1$                                       c)  $r = -1$                                       e)  $r = -0,75$   
 b)  $r = 0$                                       d)  $r = 0,1$                                       f)  $r = 0,9$

- a) En la distribución, las variables  $X$  e  $Y$  están en dependencia funcional lineal directa, y todos los valores  $(X, Y)$  se encuentran sobre una recta de pendiente positiva.  
 b) En la distribución, las variables  $X$  e  $Y$  son aleatoriamente independientes, y todos los valores  $(X, Y)$  forman una nube de puntos sin tendencia alguna (variables incorreladas).  
 c) En la distribución, las variables  $X$  e  $Y$  están en dependencia funcional lineal inversa, y todos los valores  $(X, Y)$  se encuentran sobre una recta de pendiente negativa.  
 d) Las variables  $X$  e  $Y$  están en dependencia aleatoria directa débil, y todos los valores  $(X, Y)$  forman una nube de puntos ligeramente agrupada en torno a una recta de pendiente positiva.  
 e) Las variables  $X$  e  $Y$  están en dependencia aleatoria inversa fuerte, y todos los valores  $(X, Y)$  forman una nube de puntos medianamente agrupada en torno a una recta de pendiente negativa.  
 f) Las variables  $X$  e  $Y$  están en dependencia aleatoria directa fuerte, y todos los valores  $(X, Y)$  forman una nube de puntos notablemente agrupada en torno a una recta de pendiente positiva.

13.23. Los coeficientes de correlación de dos conjuntos de datos estadísticos bidimensionales son  $r_1 = 0,87$  y  $r_2 = 0,37$ . Razona en cuál de los dos conjuntos es mejor el ajuste mediante una recta de una variable en términos de la otra.

El ajuste será mejor en el primer conjunto ( $r_1 = 0,87$ ), ya que el coeficiente de correlación es más cercano a 1, y en este caso la posible dependencia lineal de una de las variables con la otra es más fuerte.

13.24. Considera las siguientes nubes de puntos.



a) ¿En cuál de ellas los datos se ajustarán mejor a una recta?

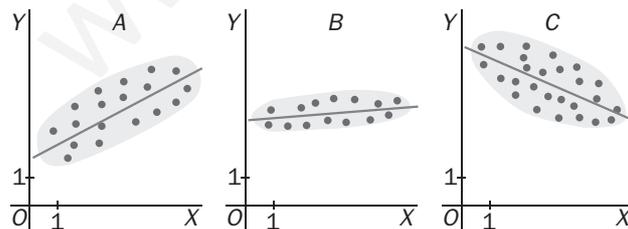
b) Asigna a cada una de las nubes uno de los siguientes coeficientes de correlación, fijando el signo en cada caso.

$$r_1 = \pm 0,99 \quad r_2 = \pm 0,6 \quad r_3 = \pm 0,8$$

a) Se ajustará mejor a una recta la nube de puntos del apartado b.

b) A:  $r = 0,8$                       B:  $r = 0,99$                       C:  $r = -0,6$

13.25. (PAU) En las gráficas siguientes se muestran las rectas de regresión obtenidas en tres estudios estadísticos.



a) ¿En cuál de las gráficas el coeficiente de correlación lineal será mayor?

b) Indica en qué gráficas el coeficiente de correlación lineal sería negativo.

Justifica las respuestas.

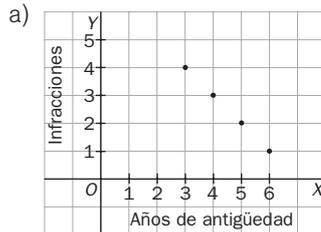
a) El de la gráfica B, ya que los puntos están más agrupados.

b) El de la gráfica C, ya que los puntos se agrupan en torno a una recta de pendiente negativa.

13.26. (PAU) En una empresa de transportes trabajan 4 conductores. Los años de antigüedad de sus permisos de conducir y las infracciones cometidas en el último año por cada uno son los siguientes:

X: años de antigüedad	3	4	5	6
Y: infracciones	4	3	2	1

- a) Representa gráficamente los datos anteriores. Razona si estos muestran correlación positiva o negativa.  
 b) Calcula el coeficiente de correlación e interprétalo en términos de la situación real.



Relación positiva

b)  $\bar{x} = \frac{3 + 4 + 5 + 6}{4} = \frac{18}{4} = 4,5$      $\bar{y} = \frac{4 + 3 + 2 + 1}{4} = \frac{10}{4} = 2,5$

$$s_x^2 = \frac{9 + 16 + 25 + 36}{4} - 4,5^2 = 1,25 \quad s_x = 1,12$$

$$s_y^2 = \frac{16 + 9 + 4 + 1}{4} - 2,5^2 = 1,25 \quad s_y = 1,12$$

$$s_{xy} = \frac{12 + 12 + 10 + 6}{4} - 4,5 \cdot 2,5 = 10 - 11,25 = -1,25$$

$$r = \frac{s_{xy}}{s_x s_y} = \frac{-1,25}{1,12 \cdot 1,12} = -0,996. \text{ Existe dependencia funcional negativa.}$$

### Modelo lineal de regresión

13.27. Cinco niñas de 2, 3, 5, 7 y 8 años de edad pesan, respectivamente, 14, 20, 32, 42 y 44 kilos.

- a) Halla la ecuación de la recta de regresión de la edad sobre el peso.  
 b) ¿Cuál sería el peso aproximado de una niña de 6 años?  
 c) ¿Tendría sentido usar la recta de regresión hallada para estimar el peso de una adolescente de 15 años?

a) Formamos la tabla:

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$
2	14	4	196	28
3	20	9	400	60
5	32	25	1024	160
7	42	49	1764	294
8	44	64	1936	352
25	152	151	5320	894

$$\bar{x} = \frac{25}{5} = 5$$

$$\bar{y} = \frac{152}{5} = 30,4$$

$$s_x^2 = \frac{151}{5} - 5^2 = 5,2 \quad s_y^2 = \frac{5320}{5} - 30,4^2 = 139,84$$

$$s_{xy} = \frac{894}{5} - 5 \cdot 30,4 = 26,8$$

Recta de regresión de X sobre Y:

$$x - 5 = \frac{26,8}{139,84} (y - 30,4) \quad x = 0,19y - 0,78$$

b) Recta de regresión de Y sobre X:  $y - 30,4 = \frac{26,8}{5,2} (x - 5) \Rightarrow y = 5,15x + 4,65$

A una niña de 6 años le corresponde un peso de:  $y = 5,15 \cdot 6 + 4,65 = 35,55$  kg.

- c) No tendría sentido porque el desarrollo físico en la adolescencia difiere notablemente del que se produce en la etapa de 2 a 8 años.

13.28. (PAU) a) Calcula la recta de regresión de Y sobre X en la distribución siguiente.

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

b) Según esta recta, ¿qué valor corresponde a  $x = 7$ ?

a) Formamos la tabla:

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$
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
25	30	183	220	102

$$\bar{x} = \frac{25}{5} = 5$$

$$\bar{y} = \frac{30}{5} = 6$$

$$s_x^2 = \frac{183}{5} - 5^2 = 11,6$$

$$s_y^2 = \frac{220}{5} - 6^2 = 8$$

$$s_{xy} = \frac{102}{5} - 5 \cdot 6 = -9,6$$

Recta de regresión de Y sobre X:

$$(y - 6) = \frac{-9,6}{11,6}(x - 5) \Rightarrow y = -0,83x + 10,14$$

b) Para  $x = 7$ , la recta estima un valor de  $y = 4,34$ .

13.29. (PAU) Al medir la pérdida de actividad de un preparado hormonal en el curso del tiempo, se obtuvo el resultado registrado en esta tabla:

Tiempo (meses)	1	2	3	4	5
% actividad restante	90	75	42	30	21

a) ¿Qué tanto por ciento de actividad restante quedará a los seis meses?

b) ¿Cuánto tiempo habrá de transcurrir para que quede el 50% de actividad restante?

Formamos la siguiente tabla:

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$
1	90	1	8100	90
2	75	4	5625	150
3	42	9	1764	126
4	30	16	900	120
5	21	25	441	105
15	258	55	16830	591

$$\bar{x} = \frac{15}{5} = 3$$

$$\bar{y} = \frac{258}{5} = 51,6$$

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

$$s_y^2 = \frac{16830}{5} - 51,6^2 = 703,44$$

$$s_{xy} = \frac{591}{5} - 3 \cdot 51,6 = -36,6$$

a) Recta de regresión de Y sobre X:  $(y - 51,6) = \frac{-36,6}{2}(x - 3) \Rightarrow y = -18,3x + 106,5$ .

A los seis meses quedará una actividad de  $y = -18,3 \cdot 6 + 106,5 = -3,3$ , según la recta de regresión calculada. Sin embargo, no tiene sentido que el tanto por ciento de actividad restante sea negativo, luego asumimos que a los seis meses el preparado hormonal del enunciado no tiene actividad.

b) Recta de regresión de X sobre Y:  $(x - 3) = \frac{-36,6}{703,44}(y - 51,6) \Rightarrow x = -0,052x + 5,68$ .

Para que quede un 50% de actividad deberá transcurrir un tiempo de  $x = -0,052 \cdot 50 + 5,68 = 3,08$  meses.

13.30. (PAU) Una persona rellena semanalmente una quiniela y un boleto de lotería primitiva, anotando el número de aciertos que tiene. Durante las 4 semanas del mes de febrero los aciertos fueron:

Semana	1. <sup>a</sup>	2. <sup>a</sup>	3. <sup>a</sup>	4. <sup>a</sup>
Aciertos en Quiniela	6	8	6	8
Aciertos en Primitiva	1	2	2	1

Obtén el coeficiente de correlación lineal e interprétalo. ¿Ofrecerían confianza las predicciones hechas con la recta de regresión?

Formamos la siguiente tabla, con la variable  $X$  el número de aciertos en la Quiniela, e  $Y$  el número de aciertos en la Primitiva:

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$
6	1	36	1	6
8	2	64	4	16
6	2	36	4	12
8	1	64	1	8
28	6	200	10	42

$$\bar{x} = \frac{28}{4} = 7$$

$$\bar{y} = \frac{6}{4} = 1,5$$

$$s_x^2 = \frac{200}{4} - 7^2 = 1$$

$$s_y^2 = \frac{10}{4} - 1,5^2 = 0,25$$

$$s_{xy} = \frac{42}{4} - 7 \cdot 1,5 = 0$$

$$r = \frac{s_{xy}}{s_x s_y} = \frac{0}{1 \cdot 0,5} = 0$$

Como  $r = 0$ , no existe correlación entre el número de aciertos en la Quiniela y el número de aciertos en la Primitiva. Es decir, son variables aleatoriamente independientes, y no podemos construir la recta de regresión para hacer predicciones.

13.31. (PAU) La información estadística obtenida de una muestra de tamaño 12 sobre la relación existente entre la inversión realizada,  $X$ , y el rendimiento obtenido,  $Y$ , en miles de euros para explotaciones agropecuarias se muestra en la siguiente tabla.

$X$	11	14	16	15	16	18	20	21	14	20	19	11
$Y$	2	3	5	6	5	3	7	10	6	10	5	6

a) Halla la recta de regresión de  $Y$  sobre  $X$ .

b) Determina la previsión de inversión que se obtendrá con un rendimiento de 7500 euros.

a) Formamos la tabla:

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$
11	2	121	4	22
14	3	196	9	42
16	5	256	25	80
15	6	225	36	90
16	5	256	25	80
18	3	324	9	54
20	7	400	49	140
21	10	441	100	210
14	6	196	36	84
20	10	400	100	200
19	5	361	25	95
11	6	121	36	66
195	68	3297	454	1163

$$\bar{x} = \frac{195}{12} = 16,25 \quad \bar{y} = \frac{68}{12} = 5,67$$

$$s_x^2 = \frac{3297}{12} - 16,25^2 = 274,75 - 264,06 = 10,68$$

$$s_y^2 = \frac{454}{12} - 5,67^2 = 37,83 - 32,15 = 5,68$$

$$s_{xy} = \frac{1163}{12} - 16,25 \cdot 5,67 = 96,92 - 92,14 = 4,78$$

Recta de regresión de  $Y$  sobre  $X$

$$y - 5,67 = \frac{4,78}{10,68} (x - 16,25) \Rightarrow y = 0,45x - 1,64$$

b) La recta de regresión de  $X$  sobre  $Y$  es:  $x - 16,25 = \frac{4,78}{5,68} (y - 5,67) \Rightarrow x = 0,84y + 11,49$

Para  $y = 7,5$ , sustituimos este valor en la ecuación obtenida:  $x = 0,84 \cdot 7,5 + 11,49 = 17,79$ . Por tanto, para un rendimiento de 7500 euros se prevé una inversión de 17 790.

13.32. Dada la distribución bidimensional:

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

- Calcula el coeficiente de correlación lineal, interpretando el resultado.
- Determina la recta de regresión de Y sobre X.
- Determina la recta de regresión de X sobre Y.
- Halla el punto en que se cortan las dos rectas.

a) Formamos la tabla:

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$
5	4,5	25	20,25	22,5
6,5	7	42,25	49	45,5
8	4,5	64	56,25	60
4	5	16	25	20
3	3,5	9	12,25	10,5
26,5	27,5	156,25	162,75	158,5

$$\bar{x} = \frac{26,5}{5} = 5,3$$

$$\bar{y} = \frac{27,5}{5} = 5,5$$

$$s_x^2 = \frac{156,25}{5} - 5,3^2 = 3,16$$

$$s_y^2 = \frac{162,75}{5} - 5,5^2 = 2,3$$

$$s_{xy} = \frac{158,5}{5} - 5,3 \cdot 5,5 = 2,66$$

$$r = \frac{s_{xy}}{s_x s_y} = \frac{2,55}{\sqrt{3,16 \cdot 2,3}} = 0,95$$

El coeficiente de correlación lineal es muy alto, por lo que las variables de la distribución del enunciado son aleatoriamente dependientes.

b) La recta de regresión de Y sobre X es  $(y - 5,5) = \frac{2,55}{3,16}(x - 5,3) \Rightarrow y = 0,807x + 1,22$ .

c) La recta de regresión de X sobre Y es  $(x - 5,3) = \frac{2,55}{2,3}(y - 5,5) \Rightarrow x = 1,109x - 0,798$ .

d) El punto en el que se cortan las dos rectas es  $(X, Y) = (5,3, 5,5)$ , cuyas coordenadas son las medias de las variables correspondientes.

13.33. Las rectas de regresión de cuatro distribuciones bidimensionales son las siguientes:

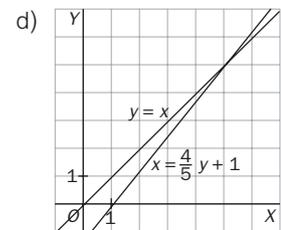
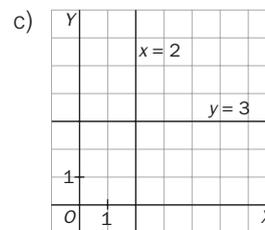
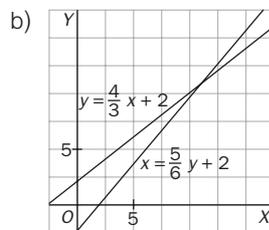
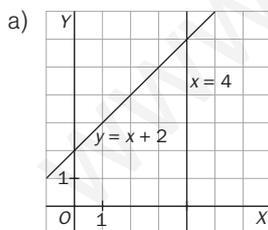
a)  $y = x + 2$ ;  $x = 4$

b)  $y = \frac{4}{5}x + 2$ ;  $x = \frac{5}{6}y + 2$

c)  $y = 3$ ;  $x = 2$

d)  $y = x$ ;  $x = \frac{4}{5}y + 1$

Indica en qué casos es significativa la correlación lineal.



El ángulo formado por las rectas es más pequeño en d y b. Por tanto, en esos casos es más significativa la correlación.

13.34. La temperatura en grados y la presión en milímetros de Hg en una ciudad a la misma hora del día durante una semana vienen reflejadas en la siguiente tabla:

Temperatura (°C)	16	18	18	17	16	12	13
Presión (mm)	716	720	719	710	705	690	680

- ¿Existe algún tipo de dependencia entre estas variables?
- ¿Qué presión se estima para un día en el que a esa hora la temperatura era igual a 15 °C?
- Halla el coeficiente de determinación e interpreta el resultado obtenido.

a) Formamos la siguiente tabla:

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$
16	716	256	512 656	11 456
18	720	324	518 400	12 960
18	719	324	516 961	12 942
17	710	289	504 100	12 070
16	705	256	497 025	11 280
12	690	144	476 100	8280
13	680	169	462 400	8840
110	4940	1762	3 487 642	77 828

$$\bar{x} = \frac{110}{7} = 15,72 \quad \bar{y} = \frac{4940}{7} = 705,71$$

$$s_x^2 = \frac{1762}{7} - 15,71^2 = 4,78$$

$$s_y^2 = \frac{3\,487\,642}{7} - 705,71^2 = 201,92$$

$$s_{xy} = \frac{77\,828}{7} - 15,71 \cdot 705,71 = 28,49$$

El coeficiente de correlación es  $r = \frac{s_{xy}}{s_x s_y} = \frac{28,49}{\sqrt{4,78 \cdot 201,92}} = 0,917$ , luego las variables Temperatura y Presión del enunciado presentan una dependencia aleatoria positiva fuerte.

- La recta de regresión de Presión sobre Temperatura es  $(y - 705,71) = \frac{28,49}{4,78}(x - 15,71) \Rightarrow y = 5,96x + 612,07$ .
- El coeficiente de determinación es  $r^2 = 0,84$ . Es decir, el 84% de la variación de Y se explica o es debido a la variación de X según el modelo de regresión lineal. Únicamente queda un 16% de la variación de Y no explicado y que puede ser debido al azar.

13.35. (PAU) Se considera la siguiente tabla de valores de dos variables:

X	1	6	9	3	2
Y	2	3	9	6	1

- Encuentra la recta de regresión de X sobre Y y de Y sobre X.
- Con los resultados obtenidos en el apartado anterior, determina el coeficiente de correlación de la variable (X, Y).
- Halla el coeficiente de determinación. ¿Será bueno el ajuste lineal mediante la recta de regresión?

a) Formamos la tabla:

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$
1	2	1	4	2
6	3	36	9	18
9	9	81	81	81
3	6	9	36	18
2	1	4	1	2
21	21	131	131	121

$$\bar{x} = \frac{21}{5} = 4,2 \quad \bar{y} = \frac{21}{5} = 4,2$$

$$s_x^2 = \frac{131}{5} - 4,2^2 = 8,56 \quad s_y^2 = \frac{131}{5} - 4,2^2 = 8,56$$

$$s_{xy} = \frac{121}{5} - 4,2 \cdot 4,2 = 6,56$$

La recta de regresión de Y sobre X es  $(y - 4,2) = \frac{6,56}{8,56}(x - 4,2) \Rightarrow y = 0,77x + 0,98$ .

La recta de regresión de X sobre Y es  $(x - 4,2) = \frac{6,56}{8,56}(y - 4,2) \Rightarrow x = 0,77y + 0,98$ .

b) El coeficiente de correlación lineal es  $r = \frac{s_{xy}}{s_x s_y} = \frac{6,56}{\sqrt{8,56 \cdot 8,56}} = 0,766$ .

c) El coeficiente de determinación es  $r^2 = 0,59$ . Como  $r^2 < 0,7$ , no es muy bueno el ajuste mediante una recta de regresión.

13.36. (PAU) Se midieron los valores de concentración en microgramos por centímetro cúbico 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 6 embarazadas al final de la gestación.

Concentración suero madre (X)	8	4	12	2	7	9
Concentración suero 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.

c) Halla el coeficiente de determinación e interprétalo para estudiar la bondad del ajuste.

a) Formamos la tabla siguiente:

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$
8	6	64	36	48
4	4	16	16	16
12	8	144	64	96
2	1	4	1	2
7	4	49	16	28
9	5	81	25	45
42	28	358	158	235

$$\bar{x} = \frac{42}{6} = 7$$

$$\bar{y} = \frac{28}{6} = 4,67$$

$$s_x^2 = \frac{358}{6} - 7^2 = 10,67$$

$$s_x = \sqrt{10,67} = 3,27$$

$$s_y^2 = \frac{158}{6} - 4,67^2 = 4,52$$

$$s_y = \sqrt{4,52} = 2,13$$

$$s_{xy} = \frac{235}{6} - 7 \cdot 4,67 = 6,48$$

$$r = \frac{6,48}{3,27 \cdot 2,13} = 0,93$$

b) Recta de regresión de Y sobre X:  $y - 4,67 = \frac{6,48}{10,67}(x - 7) \Rightarrow y = 0,607x + 3,41$

c) El coeficiente de determinación es  $r^2 = 0,865$ . Es decir, el 86,5% de la variación de Y puede ser explicado por la variación de X a partir de la recta de regresión y únicamente resta un 13,5% de la variación de Y que no puede ser explicado a partir de la recta de regresión. Por tanto, el ajuste lineal es bueno.

Recta de Tukey

13.37. Sea la variable bidimensional dada por la tabla siguiente:

x	1	2	3	4	5	6	7	8	9	10
y	3	14	5	6	7	8	9	10	11	12

- a) Halla la recta de Tukey.  
 b) Calcula la recta de regresión de Y sobre X.

a) Formamos con los datos ordenados tres grupos:

$$G_1 = \{(1, 3), (2, 14), (3, 5)\} \quad G_2 = \{(4, 6), (5, 7), (6, 8), (7, 9)\} \quad G_3 = \{(8, 10), (9, 11), (10, 12)\}$$

Para cada grupo  $G_i$  hallamos el punto  $P_i(x_i, y_i)$ :

$$P_1(2, 5) \quad P_2(5,5; 7,5) \quad P_3(9, 11)$$

El baricentro del triángulo de vértices  $P_1P_2P_3$  tiene por coordenadas:

$$x_G = \frac{2 + 5,5 + 9}{3} = 5,5 \quad y_G = \frac{5 + 7,5 + 11}{3} = 7,83$$

La pendiente  $P_1P_3$  es:  $m = \frac{11 - 5}{9 - 2} = \frac{6}{7} = 0,857$ .

La recta de Tukey es:  $y - 7,83 = 0,857(x - 5,5) \Rightarrow y = 0,857x + 3,12$ .

b) Formamos la siguiente tabla:

$x_i$	$y_i$	$x_i^2$	$x_i y_i$
1	3	1	3
2	14	4	28
3	5	9	15
4	6	16	24
5	7	25	35
6	8	36	48
7	9	49	63
8	10	64	80
9	11	81	99
10	12	100	120
55	85	385	515

$$\bar{x} = \frac{55}{10} = 5,5 \quad \bar{y} = \frac{85}{10} = 8,5$$

$$s_x^2 = \frac{385}{10} - 5,5^2 = 8,25 \quad s_{xy} = \frac{515}{10} - 5,5 \cdot 8,25 = 6,125$$

La recta de regresión de Y sobre X es:

$$y - 8,5 = \frac{6,125}{8,25}(x - 5,5) \Rightarrow y = 0,74x + 4,42$$

13.38. Dada la variable bidimensional cuyos datos se recogen en la siguiente tabla:

X	2	4	6	8	10	12	14	16	18	20
Y	23	25	26	20	11	23	24	27	28	24

- a) Calcula la recta de Tukey.  
 b) Halla la recta de regresión de Y sobre X.

a) Formamos con los datos ordenados tres grupos:

$$G_1 = \{(2, 23), (4, 25), (6, 26)\} \quad G_2 = \{(8, 20), (10, 11), (12, 23), (14, 24)\} \quad G_3 = \{(16, 27), (18, 28), (20, 24)\}$$

Para cada grupo  $G_i$  hallamos el punto  $P_i(x_i, y_i)$ :

$$P_1(4, 25) \quad P_2(11; 21,5) \quad P_3(18, 27)$$

El baricentro del triángulo de vértices  $P_1P_2P_3$  tiene por coordenadas:

$$x_G = \frac{4 + 11 + 18}{3} = 7,67 \quad y_G = \frac{25 + 21,5 + 27}{3} = 24,5$$

La pendiente  $P_1P_3$  es:  $m = \frac{27 - 25}{18 - 4} = \frac{2}{14} = 0,14$ .

La recta de Tukey es:  $y - 24,5 = 0,14(x - 7,67) \Rightarrow y = 0,14x + 23,43$

b) Formamos la tabla:

$x_i$	$y_i$	$x_i^2$	$x_i y_i$
2	23	4	46
4	25	16	100
6	26	36	156
8	29	64	160
10	11	100	110
12	23	144	276
14	24	196	336
16	27	256	432
18	28	324	504
20	24	400	480
110	231	1540	2600

$$\bar{x} = \frac{110}{10} = 11$$

$$\bar{y} = \frac{231}{10} = 23,1$$

$$s_x^2 = \frac{1540}{10} - 11^2 = 33$$

$$s_{xy} = \frac{2600}{10} - 11 \cdot 23,1 = 5,9$$

La recta de regresión de Y sobre X es:

$$y - 23,1 = \frac{5,9}{3,3}(x - 11) \Rightarrow y = 0,18x + 21,13$$

## PROBLEMAS

13.39. (PAU) En un estudio sobre la estatura ( $E$ ) y el peso ( $P$ ) realizado en un grupo de 50 estudiantes se ha obtenido una estatura media de 165 cm y un peso medio de 61 kg.

Sabiendo que al aumentar la estatura aumenta también el peso, identifica, entre las siguientes, cuál podría ser la recta de regresión del peso en función de la estatura obtenida a través de los datos recogidos en ese grupo de estudiantes.

a)  $P = 226 - E$       b)  $P = -104 + E$       c)  $P = 5 + \frac{1}{3}E$       d)  $P = 171 - \frac{2}{3}E$

a) La recta de regresión debe ser de pendiente positiva, ya que al aumentar la estatura, aumenta el peso. Por tanto, estudiaremos si las rectas b y c pasan por (165, 61).

b)  $P = -104 + 165 = 61$ . Cumple la condición.      c)  $P = 5 + \frac{165}{3} = 60$ . No cumple la condición.

La recta pedida podría ser  $P = -104 + E$ .

13.40. (PAU) Dos variables  $X$  e  $Y$  son medidas en 10 animales de laboratorio. A partir de los datos registrados se obtiene:  $\sum_{j=1}^{10} x_j = 200$ ;  $\sum_{j=1}^{10} y_j = 50$  y  $r = -0,7$ , siendo  $r$  el coeficiente de correlación lineal. ¿Cuál entre las siguientes podría ser la recta de regresión de la variable  $Y$  sobre la variable  $X$ ? Justifica la respuesta.

a)  $y = -45 + 2,5x$       c)  $y = 9 - 0,7x$   
 b)  $y = 35 - 1,5x$       d)  $y = -200 + 50x$

La recta de regresión debe cumplir dos propiedades:

1. Su pendiente debe ser positiva si la correlación es positiva (si  $r > 0$ ); en caso contrario debe tener pendiente negativa.

2. Debe pasar por el punto medio de la distribución  $(\bar{x}, \bar{y})$ .

Por la propiedad 1, hay que rechazar las rectas dadas en a y d, pues  $r = -0,7$ .

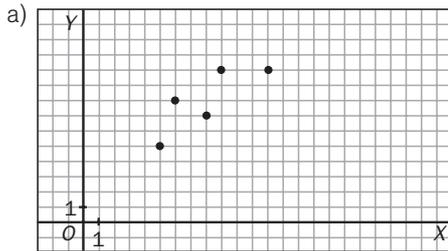
Como  $\bar{x} = \frac{200}{10} = 20$  e  $\bar{y} = \frac{50}{10} = 5$ , según la propiedad 2, la recta debe pasar por (20, 5).

La recta que cumple esta condición es la dada en b. Por tanto, la recta de regresión pedida es  $y = 35 - 1,5x$ .

13.41. (PAU) En la tabla se recogen las puntuaciones en dos pruebas (X, Y) de cinco alumnos.

X	6	5	12	8	9
Y	8	5	10	7	10

- Haz el diagrama de dispersión (nube de puntos).
- Si la recta de regresión de Y sobre X es  $y = 0,63x + 2,96$ , ¿qué puntuación espera en la prueba Y un sujeto que ha obtenido una puntuación de 10 en la prueba X?
- Calcula la covarianza de las puntuaciones.



b) Para  $X = 10$ , según la recta de regresión del enunciado  $Y = 0,63 \cdot 10 + 2,96 = 9,26$ .

c) Formamos la tabla:

$x_i$	$y_i$	$x_i y_i$
6	8	48
5	5	25
12	10	120
8	7	56
9	10	90
40	40	339

$$\bar{x} = \frac{40}{5} = 8 \qquad \bar{y} = \frac{40}{5} = 8$$

$$s_{xy} = \frac{339}{5} - 8 \cdot 8 = 3,8$$

13.42. (PAU) Las calificaciones obtenidas por un grupo de 10 alumnos en Filosofía y Matemáticas son las que se recogen en la tabla siguiente.

Filosofía	5	8	7	3	5	4	8	5	5	8
Matemáticas	7	9	10	4	7	4	10	5	7	9

- Halla la recta de regresión de la calificación obtenida en Filosofía (Y) respecto de la calificación obtenida en Matemáticas (X).
- Calcula la calificación estimada en Filosofía para un alumno que en Matemáticas ha obtenido un 6.

a) Formamos la siguiente tabla:

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$
7	5	49	25	35
9	8	81	64	72
10	7	100	49	70
4	3	16	9	12
7	5	49	25	35
4	4	16	16	16
10	8	100	64	80
5	5	25	25	25
7	5	49	25	35
9	8	81	64	72
72	58	566	366	452

$$\bar{x} = \frac{72}{10} = 7,2 \qquad \bar{y} = \frac{58}{10} = 5,8$$

$$s_x^2 = \frac{566}{10} - 7,2^2 = 4,76 \qquad s_y^2 = \frac{366}{10} - 5,8^2 = 2,96$$

$$s_{xy} = \frac{452}{10} - 5,8 \cdot 7,2 = 3,44$$

La recta de regresión de Y sobre X es:

$$(y - 5,8) = \frac{3,44}{4,76}(x - 7,2) \Rightarrow y = 0,45x + 2,54$$

b) Utilizando la recta de regresión anterior se obtiene para  $x = 6$  un valor estimado de  $y = 5,24$ , que sería la calificación estimada en Filosofía para un alumno que obtuviera un 6 en matemáticas.

13.43. El número de horas dedicadas al estudio de una prueba y las respuestas correctas obtenidas en un test de 100 preguntas vienen en la siguiente tabla.

X: horas de estudio	20	16	34	23	27	32	18	22
Y: aciertos	65	60	85	70	90	95	75	80

- a) Halla la recta de regresión de Y sobre X.  
 b) Calcula la calificación estimada para una persona que hubiese estudiado 28 horas

Formamos la tabla:

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$
20	65	400	4225	1300
16	60	256	3600	960
34	85	1156	7225	2890
23	70	529	4900	1610
27	90	729	8100	2430
32	95	1024	9025	3040
18	75	324	5625	1350
22	80	484	6400	1760
192	620	4902	49100	15340

$$a) \bar{x} = \frac{192}{8} = 24 \qquad \bar{y} = \frac{620}{8} = 77,5$$

$$s_x^2 = \frac{4902}{8} - 24^2 = 36,75 \qquad s_{xy} = \frac{15340}{8} - 24 \cdot 77,5 = 57,5$$

La recta de regresión de Y sobre X es

$$y - 77,5 = \frac{57,5}{36,75} (x - 24) \Rightarrow y = 1,56x + 39,95$$

- b) Si  $x = 28$ ,  $y = 1,56 \cdot 28 + 39,95 = 83,63$ . Por tanto, si un alumno dedica al estudio 28 horas, se espera que responda correctamente a 84 preguntas.

13.44. (PAU) El número de licencias de caza, en miles, y el número de votantes de un determinado partido político en seis comunidades autónomas, en decenas de millar, está expresado en la siguiente tabla:

N.º de votantes (Y)	206	26	27	14	24	12
N.º de licencias (X)	103	26	3	7	26	5

Calcula:

- a) La media y la varianza de las variables X e Y.  
 b) El coeficiente de correlación, interpretando su valor.  
 c) En el caso de que exista correlación: si en una determinada comunidad el partido político tiene 50 decenas de millar de votantes, ¿cuántas licencias de caza, en miles, se puede estimar que existen?

a) Formamos la tabla:

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$
103	206	10609	42436	21218
26	26	676	676	676
3	27	9	729	81
7	14	49	196	98
26	24	676	576	624
5	12	25	144	60
170	309	12044	44757	22755

$$\bar{x} = \frac{170}{6} = 28,3 \qquad \bar{y} = \frac{309}{6} = 51,5$$

$$s_x^2 = \frac{12044}{6} - 28,3^2 = 1204,56 \qquad s_x = 34,71$$

$$s_y^2 = \frac{44757}{6} - 51,5^2 = 4807,25 \qquad s_y = 69,33$$

$$s_{xy} = \frac{22755}{6} - 5,15 \cdot 28,3 = 2333,33$$

- b) El coeficiente de correlación es  $r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{2333,33}{34,71 \cdot 69,33} = 0,97$ , luego la correlación lineal entre las variables X e Y es positiva y fuerte.

- c) La recta de regresión de licencias sobre votantes es  $(x - 28,3) = \frac{2333,33}{4807,25} (y - 51,5) \Rightarrow x = 0,485y + 3,3$ , por lo que si en una comunidad autónoma tenemos 50 decenas de millar de votantes,  $y = 50$ , el número de licencias, en miles, será  $x = 0,485 \cdot 50 + 3,3 = 27,55$ .

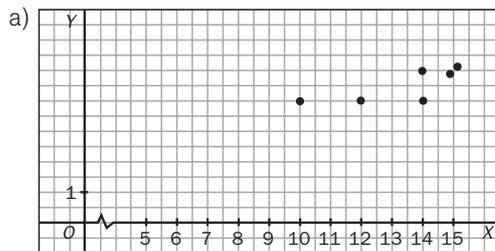
13.45. (PAU) La siguiente tabla relaciona la inversión, en millones, y la rentabilidad obtenida, en tanto por ciento, de seis inversores.

Inversión	10	12	14	14	15	15
Rentabilidad %	4	4	5	4	5	5

Determina:

- La nube de puntos.
- La media y la desviación típica de las variables inversión y rentabilidad.
- El coeficiente de correlación, e interprétalo.
- Si un inversionista invierte 13,5 millones, ¿qué rentabilidad puede esperar?
- Si un inversionista ha obtenido una rentabilidad del 5,5%, ¿qué capital se puede esperar que haya invertido?

Consideramos la inversión como variable X, y la rentabilidad, como variable Y.



Formamos la siguiente tabla:

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$
10	4	100	16	40
12	4	144	16	48
14	5	196	25	70
14	4	196	16	56
15	5	225	25	75
15	5	225	25	75
80	27	1086	123	364

b)  $\bar{x} = \frac{80}{6} = 13,33$     $s_x^2 = \frac{1086}{6} - 13,33^2 = 3,31$     $s_x = \sqrt{3,31} = 1,82$

$\bar{y} = \frac{27}{6} = 4,5$     $s_y^2 = \frac{123}{6} - 4,5^2 = 0,25$     $s_y = \sqrt{0,25} = 0,5$

c)  $s_{xy} = \frac{364}{6} - 13,33 \cdot 4,5 = 0,68$     $r = \frac{0,68}{1,82 \cdot 0,5} = 0,74$

Como el valor de  $r$  es próximo a 1, la correlación es directa y moderadamente fuerte. Por tanto, las variables están en dependencia aleatoria.

d) Hallamos la recta de regresión de Y sobre X:  $y - 4,5 = \frac{0,68}{3,31}(x - 13,33)$     $y = 0,21x + 1,76$

Por tanto, para  $x = 13,5$  se obtiene:  $y = 0,21 \cdot 13,5 + 1,76 = 4,59$ . Así pues, si un inversionista invierte 13,5 millones, se espera que obtenga una rentabilidad del 4,57%.

e) Hallamos la recta de regresión de X sobre Y:  $x - 13,33 = \frac{0,68}{0,25}(y - 4,5)$     $y = 2,72x + 1,09$

Por tanto, para  $y = 5,5$  se obtiene:  $x = 2,72 \cdot 5,5 + 1,09 = 16,05$ . Así pues, si un inversionista obtiene una rentabilidad del 5,5%, se supone que había invertido 16,05 millones.

13.46. (PAU) A partir de los datos recogidos sobre facturación anual y beneficios anuales en un determinado año sobre un conjunto de 50 grandes empresas europeas, se ha calculado una facturación media de 80 millones de euros y unos beneficios medios de 65 millones de euros.

- a) Teniendo en cuenta esa información, determina la recta de regresión que permite obtener los beneficios en función de la facturación, sabiendo que a partir de ella se han calculado unos beneficios de 59 millones de euros para una empresa que ha facturado 75 millones de euros en 1998.  
 b) ¿Qué signo tendría el coeficiente de correlación lineal entre ambas variables?

Consideramos  $X$  como la variable facturación e  $Y$  como la variable beneficio.

- a) La recta de regresión de  $Y$  sobre  $X$  es de la forma  $y = mx + n$ . Como la recta pasa por los puntos (80, 65) y (75, 59), se tiene:

$$65 = 80m + n$$

$$59 = 75m + n$$

Resolviendo el sistema, resulta:  $m = \frac{6}{5}$  y  $n = -31$ . La recta de regresión de  $Y$  sobre  $X$  es:  $y = \frac{6}{5}x - 31$ .

- b) El coeficiente de correlación tiene el mismo signo que la pendiente de la recta de regresión; por tanto, es positivo.

13.47. (PAU) La recta de regresión de una variable  $Y$  respecto de la variable  $X$  es  $y = 0,3x + 1$ . Los valores que ha tomado la variable  $x$  han sido {3, 4, 5, 6, 7}.

- a) Determina el valor esperado de  $Y$  para el valor particular de  $x = 3,5$ .  
 b) Si los valores de la variable  $Y$  utilizados para la regresión se multiplican por 10 y se dejan los mismos valores para la variable  $X$ , determina razonadamente la nueva recta de regresión.

- a) Para  $x = 3,5$ ,  $y = 0,3 \cdot 3,5 + 1 = 2,05$

- b) Si los valores de la variable  $Y$  se multiplican por 10, se tendrá:

$$\bar{y}' = 10\bar{y}, \text{ siendo } \bar{y} \text{ la media inicial. } s'_{xy} = \frac{\sum 10x_i y_i}{N} - \bar{x}10\bar{y} = 10s_{xy}$$

Con esto, la nueva recta será:

$$y - 10 = \frac{10s_{xy}}{s_x^2}(x - \bar{x}) \Rightarrow y = \frac{10s_{xy}}{s_x^2}x + 10\left(\bar{y} - \frac{s_{xy}}{s_x^2}\bar{x}\right) \Rightarrow y = 3x + 10$$

13.48. (PAU) Cien alumnos prepararon un examen de Matemáticas. Se representa por  $X$  el número de problemas hechos por cada alumno en la preparación y por  $Y$  la calificación obtenida. Sabiendo que las medias aritméticas de esas variables fueron  $\bar{x} = 9,2$  e  $\bar{y} = 9,5$ , que el coeficiente de correlación entre esas variables fue 0,7 y que la desviación típica de la variable  $Y$  fue el doble que la de la variable  $X$ , se pide obtener, razonadamente:

- a) Las ecuaciones de las rectas de regresión de  $Y$  sobre  $X$  y de  $X$  sobre  $Y$ .  
 b) La calificación que la recta de regresión adecuada predice para un alumno que sólo hizo 6 problemas durante la preparación del examen.

- a) Como la desviación típica de la variable  $Y$  fue el doble que la de la variable  $X$ , se tiene:

$$r = \frac{s_{XY}}{s_x s_y} = \frac{s_{XY}}{s_x \cdot 2s_x} = \frac{s_{XY}}{2s_x^2} = 0,7 \Rightarrow \frac{s_{XY}}{s_x^2} = 0,7 \cdot 2 = 1,4$$

La recta de regresión de  $Y$  sobre  $X$  es:  $(y - 9,5) = 1,4(x - 9,2) \Rightarrow y = 1,4x - 3,38$

$$\frac{s_{XY}}{s_y^2} = \frac{s_{XY}}{(2s_x)^2} = \frac{s_{XY}}{4s_x^2} = \frac{1}{4} \cdot 1,4 = 0,35$$

La recta de regresión de  $X$  sobre  $Y$  es:  $(x - 9,2) = 0,35(y - 9,5) \Rightarrow x = 0,35y + 5,875$

- b) Para  $x = 6$  problemas realizados durante la preparación, se estima que el alumno obtendrá una calificación de  $y = 1,4 \cdot 6 - 3,38 = 5,02$ .

13.49. (PAU) En un determinado grupo de Bachillerato de un centro de Educación Secundaria, las calificaciones de Matemáticas de 8 alumnos en las evaluaciones 1.ª y 2.ª están representadas en la siguiente tabla.

1.ª evaluación (X)	4	8	3	6	4	9	8	6
2.ª evaluación (Y)	3	7	3	5	4	7	6	5

- a) Calcula el coeficiente de correlación lineal.  
 b) Calcula la recta de regresión de Y sobre X.

Si un alumno del grupo ha obtenido una calificación de 7,5 en la primera evaluación, ¿qué calificación se supone que obtendrá en la 2.ª?

Construimos la tabla:

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$
4	3	16	9	12
8	7	64	49	56
3	3	9	9	9
6	5	36	25	30
4	4	16	16	16
9	7	81	49	63
8	6	64	36	48
6	5	36	25	30
48	40	322	218	264

$$a) \bar{x} = \frac{48}{8} = 6 \quad s_x^2 = \frac{322}{8} - 6^2 = 4,25 \quad s_x = \sqrt{4,25} = 2,06$$

$$\bar{y} = \frac{40}{8} = 5 \quad s_y^2 = \frac{218}{8} - 5^2 = 2,25 \quad s_y = \sqrt{2,25} = 1,5$$

$$s_{xy} = \frac{264}{8} - 6 \cdot 5 = 3 \quad r = \frac{3}{2,06 \cdot 1,5} = 0,97$$

- b) La recta de regresión de Y sobre X es

$$(y - 5) = \frac{3}{4,25}(x - 6) \Rightarrow y = 0,706x + 0,765$$

Para un alumno que ha obtenido una calificación de 7,5 en la 1.ª evaluación,  $x = 7,5$ , se estima que en la 2.ª evaluación obtenga una calificación de  $y = 6,06$ .

### PROFUNDIZACIÓN

13.50. (PAU) La tabla siguiente muestra los valores observados de dos variables X e Y en 5 individuos.

X	1	-1	x	2	3
Y	-2	-3	2	1	0

- a) Halla el valor x para que el coeficiente de correlación se anule.  
 b) Suponiendo que  $x = 4$ , halla la recta de regresión de Y sobre X y estudia el valor de Y cuando X toma el valor -2.

$$a) \bar{x} = \frac{5+x}{5}, \bar{y} = -\frac{2}{5} \Rightarrow s_{xy} = \frac{-2+3+2x+2}{5} - \frac{5+x}{5} \cdot \left(-\frac{2}{5}\right) = \frac{12x+25}{25}$$

$$\text{Como } r = \frac{s_{xy}}{s_x s_y} = 0, s_{xy} = 0; \frac{12x+25}{25} = 0, \text{ y, por tanto, } x = -\frac{25}{12}$$

$$b) \text{ Si } x = 4, \bar{x} = \frac{9}{5}; \bar{y} = -\frac{2}{5}; s_{xy} = \frac{73}{25}$$

$$s_x^2 = \frac{1+1+16+4+9}{5} - \frac{81}{25} = \frac{74}{25}$$

$$\text{Recta de regresión de Y sobre X: } y - \frac{2}{5} = \frac{73}{74} \left(x - \frac{9}{5}\right) \Rightarrow y = 0,986x - 2,176. \text{ Si } x = -2, y = -4,148$$

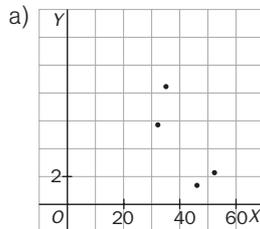
13.51. (PAU) Los siguientes pares de datos corresponden a las variables X (producto interior bruto en miles de millones de euros) e Y (tasa de inflación):

X	34	46	52	32
Y	8,3	1,5	2,1	5,8

- a) Dibuja el diagrama de dispersión de los datos.  
 b) Decide razonadamente cuál de las siguientes rectas es la de regresión de Y sobre X:

$$y = 16,26 + 2,88x \qquad y = 16,26 - 2,88x$$

- c) Calcula el valor esperado de la tasa de inflación que corresponde a un producto interior bruto de 43 mil millones de euros.



- b) De ser alguna de esas dos rectas, será la de la pendiente negativa,  $y = 16,26 - 2,88x$ , pues así lo sugiere el diagrama de dispersión.  
 c) Si  $x = 4,3$ , sustituyendo en la recta dada se obtiene:  $y = -2,88 \cdot 4,3 + 16,26 = 3,88$ .  
 Es decir, para un producto interior bruto de 4,3 decenas de millones de euros se espera una tasa de inflación del 3,88.

13.52. (PAU) Se ha observado una variable estadística bidimensional y se ha obtenido la siguiente tabla:

X \ Y	100	50	25
14	1	1	—
18	2	3	2
22	2	1	2

- a) Calcula la covarianza.  
 b) Obtén e interpreta el coeficiente de correlación lineal.  
 c) Determina la ecuación de la recta de regresión de Y sobre X.

Formamos la siguiente tabla:

$x_i$	$y_i$	$f_i$	$x_i f_i$	$x_i^2 f_i$	$y_i f_i$	$y_i^2 f_i$	$x_i y_i f_i$
100	14	1	100	10 000	14	196	1400
100	18	2	200	20 000	36	648	3600
50	14	1	50	2500	14	196	700
50	18	3	150	7500	54	972	2700
50	22	1	50	2500	22	484	1100
25	22	2	50	1250	44	968	1100
375	108	10	600	43 750	184	3464	10 600

a)  $\bar{x} = \frac{600}{10} = 60 \qquad \bar{y} = \frac{184}{10} = 18,4$   
 $s_x^2 = \frac{47\,750}{10} - 60^2 = 775 \qquad s_x = \sqrt{775} = 27,8$   
 $s_y^2 = \frac{3464}{10} - 18,4^2 = 7,84 \qquad s_y = \sqrt{7,84} = 2,8$   
 $s_{xy} = \frac{10\,600}{10} - 60 \cdot 18,4 = -44$   
 b)  $r = \frac{-44}{27,84 \cdot 2,8} = -0,56$

Se trata de una correlación negativa débil, ya que el valor absoluto de  $r$  se encuentra más cercano a 0 que a 1.

- c) Ecuación de la recta de regresión de Y sobre X:  $y - 18,4 = -0,06(x - 60)$ ;  $y = -0,06x + 22$ .

13.53. (PAU) Se ha solicitado a un grupo de 50 individuos información sobre el número de horas que dedican diariamente a dormir (X) y a ver la televisión (Y). Los resultados vienen dados por la siguiente tabla.

X	6	7	8	9	10
Y	4	3	3	2	1
f <sub>i</sub>	3	16	20	10	1

- Calcula el coeficiente de correlación entre X e Y, e interprétalo en los términos del enunciado.
- Calcula la ecuación de la recta de regresión de Y sobre X.
- Si una persona duerme 8 horas y media, ¿cuántas horas cabe esperar que vea la televisión?
- Si no se calcula la recta de regresión de X sobre Y, ¿en qué punto se cortará esta recta con la calculada en el apartado b)?
- Si una persona ve la televisión 2 horas, ¿cuánto tiempo cabe esperar que duerma?

Formamos la siguiente tabla:

x <sub>i</sub>	y <sub>i</sub>	f <sub>i</sub>	x <sub>i</sub> <sup>2</sup>	y <sub>i</sub> <sup>2</sup>	f <sub>i</sub> x <sub>i</sub>	f <sub>i</sub> y <sub>i</sub>	f <sub>i</sub> x <sub>i</sub> <sup>2</sup>	f <sub>i</sub> y <sub>i</sub> <sup>2</sup>	f <sub>i</sub> x <sub>i</sub> y <sub>i</sub>
6	4	3	36	16	18	12	108	48	72
7	3	16	49	9	112	48	784	144	336
8	3	20	64	9	160	60	1280	180	480
9	2	10	81	4	90	20	810	40	180
10	1	1	100	1	10	1	100	1	10
		50			390	141	3082	413	1078

$$\begin{aligned}
 \bar{x} &= \frac{390}{50} = 7,8 & s_x^2 &= \frac{3082}{50} - (7,8)^2 = 0,80 & s_x &= \sqrt{0,8} = 0,89 \\
 \bar{y} &= \frac{141}{50} = 2,82 & s_y^2 &= \frac{413}{50} - (2,82)^2 = 0,3076 & s_y &= \sqrt{0,3076} = 0,55 \\
 s_{xy} &= \frac{1078}{50} - (7,8)(2,82) = -0,436 & r &= \frac{s_{xy}}{s_x s_y} = \frac{-0,436}{0,89 \cdot 0,55} = -0,88
 \end{aligned}$$

La correlación lineal entre ambas variables es grande e inversa.

$$b) y - 2,82 = -\frac{0,436}{0,80}(x - 7,8) \Rightarrow y = -0,545x + 7,071$$

$$c) \text{ Si } x = 8,5: y = -0,545 \cdot 8,5 + 7,071 = 2,44$$

Si una persona duerme 8 horas y media, verá la televisión durante 2 horas 26,4 minutos.

d) La recta de regresión de Y sobre X y la recta de regresión de X sobre Y se cortan en el centro de gravedad de la nube de puntos, es decir, en el punto  $(\bar{x}, \bar{y}) = (7,8; 2,82)$ .

$$e) \text{ La recta de regresión de X sobre Y es: } x - 7,8 = \frac{-0,436}{0,3076}(y - 2,82); x = 11,787 - 1,417y.$$

Si  $y = 2$ ,  $x = 11,787 - 1,41 \cdot 2 = 8,953$  horas, es decir, que si una persona ve la televisión durante 2 horas, se espera que duerma aproximadamente 9 horas.

13.54. Los valores de dos variables X e Y se distribuyen según la tabla siguiente.

	X	0	2	4
Y				
1		2	1	3
2		1	4	2
3		2	5	0

- Determina el coeficiente de correlación y la recta de regresión de Y sobre X.
- Comenta lo fiables que son las predicciones basadas en esta recta.

a) Convertimos la tabla de doble entrada en tabla simple y efectuamos los siguientes cálculos:

$x_i$	$y_i$	$f_i$	$x_i f_i$	$x_i^2 f_i$	$y_i f_i$	$y_i^2 f_i$	$x_i y_i f_i$
0	1	2	0	0	2	2	0
0	2	1	0	0	2	4	0
0	3	2	0	0	6	18	0
2	1	1	2	4	1	1	2
2	2	4	8	16	8	16	16
2	3	5	10	20	15	45	30
4	1	3	12	48	3	3	12
4	2	2	8	32	4	8	16
		20	40	120	41	97	76

$$\bar{x} = \frac{40}{20} = 2 \quad s_x^2 = \frac{120}{20} - 2^2 = 2 \quad s_x = \sqrt{2} = 1,41$$

$$\bar{y} = \frac{41}{20} = 2,05 \quad s_y^2 = \frac{97}{20} - 2,05^2 = 0,65 \quad s_y = \sqrt{0,65} = 0,81$$

$$s_{xy} = \frac{76}{20} - 2 \cdot 2,05 = -0,3 \quad r = \frac{-0,3}{1,41 \cdot 0,81} = -0,26$$

Recta de regresión de Y sobre X:

$$y - 2,05 = -0,15(x - 2); y = -0,15x + 2,35$$

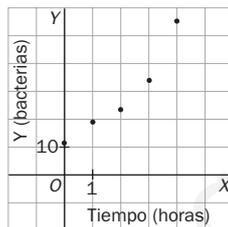
b) Como el coeficiente de correlación lineal en valor absoluto está alejado de la unidad, las predicciones que se puedan hacer sobre la recta anteriormente obtenida no serán fiables.

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

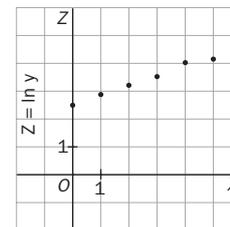
X: horas	0	1	2	3	4
Y: bacterias	1	3	8	21	53

¿Cuántas bacterias habrá al cabo de cinco horas? Ayuda: realiza el cambio de variable  $Z = \ln Y$ .

Dibujamos el diagrama de dispersión y observamos que existe una relación curvilínea.



X	Z = ln Y
0	0
1	1,099
2	2,079
3	3,045
4	3,970



Al realizar el cambio de variable  $Z = \ln Y$ , la nube de puntos se ajusta a una recta. Podemos calcular ahora la recta de regresión de Z sobre X.

Formamos la siguiente tabla:

$x_i$	$z_i$	$x_i^2$	$z_i^2$	$x_i z_i$
0	0	0	0	0
1	1,099	1	1,208	1,099
2	2,079	4	4,322	4,158
3	3,045	9	9,272	9,135
4	3,970	16	15,761	15,88
10	10,193	30	30,563	30,272

$$\bar{x} = \frac{10}{5} = 2$$

$$\bar{z} = \frac{10,193}{5} = 2,04$$

$$s_x^2 = \frac{30}{5} - 2^2 = 2$$

$$s_x = \sqrt{2} = 1,41$$

$$s_z^2 = \frac{30,563}{5} - 2,04^2 = 1,951$$

$$s_z = \sqrt{1,951} = 1,4$$

$$s_{xz} = \frac{30,272}{5} - 2 \cdot 2,04 = 1,97$$

Como  $r = \frac{1,97}{\sqrt{2} \cdot 2,04} = 0,98$ , nos indica que la relación existente entre X e Y es de tipo exponencial.

La recta de regresión de Z sobre X es:  $z - 2,04 = \frac{1,97}{2}(x - 2) \Rightarrow z = 0,985x + 0,07$

Como  $Z = \ln Y$ , se cumple que:  $\ln y = 0,985x + 0,07$ . Tomando exponenciales en los dos miembros queda:

$$y = e^{0,985x + 0,07}$$

El número de bacterias esperado al cabo de cinco horas es  $y = e^{0,985 \cdot 5 + 0,07} = e^{4,995} = 147,7$ .